Universidade de Lisboa
Instituto Superior de Economia e Gestão

Essays in Spatial Econometrics

Author:
Luís Filipe Ávila da Silveira dos Santos

Advisor:
Isabel Maria Dias Proença

A thesis submitted in fulfillment of the requirements
for the degree of Ph.D.

in

Applied Mathematics for Economics and Management

July 2020

Universidade de Lisboa
Instituto Superior de Economia e Gestão

Essays in Spatial Econometrics

Author:
Luís Filipe Ávila da Silveira dos Santos

Advisor:
Isabel Maria Dias Proença

A thesis submitted in fulfillment of the requirements
for the degree of Ph.D.
in
Applied Mathematics for Economics and Management

Committee Chair:

Nuno João de Oliveira Valério
Full Professor and President of the Scientific Board
ISEG, University of Lisbon.

Committee Members:

Joaquim José dos Santos Ramalho
Full Professor
ISCTE, University Institute of Lisbon.

Isabel Maria Dias Proença (advisor)
Senior Associate Professor
ISEG, University of Lisbon.

Paulo de Freitas Guimarães
Visiting Associate Professor
FEP, University of Porto.

Felipa Dias de Mello Sampayo
Assistant Professor
ISCTE, University Institute of Lisbon.

Paulo Miguel Dias Costa Parente
Assistant Professor
ISEG, University of Lisbon.

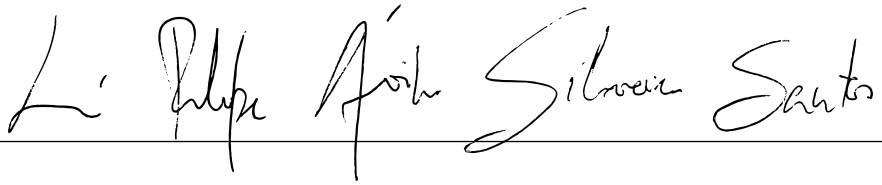
Financial support by FCT/MCTES through national funds
Ph.D. scholarship SFRH/BD/92277/2013

July 2020

Declaration of Authorship

I, Luís Filipe Ávila da Silveira dos Santos, declare that this thesis titled, "Essays in Spatial Econometrics" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: 

Date: July 15, 2020

This Ph.D. thesis is dedicated to the loving memory of my grandparents, João and Maria Victória

Abstract

This thesis addresses the specification and estimation of Spatial Lag Models for dichotomous or fractional responses. Three essays are presented. The first essay suggests a new method to approximate the inverse of the spatial lag operator, used in the estimation of Spatial Lag Models for binary outcomes. Related matrix operations are approximated, as well. Closed formulas for the elements of the approximated matrices are deduced. Computational time and complexity is greatly reduced. The second essay focus on the specification of Spatial Lag Models for fractional responses. Observations at the corners, zero and one, are allowed. Two specifications are proposed. The Fractional Response Spatial Lag Model (FRSLM), extends the seminal approach of Papke and Wooldridge (1996) to spatial frameworks. The approximate Fractional Response Spatial Lag Model (aFRSLM), allows to write the FRSLM as an approximate reduced form. Of particular relevance is the interpretation of policy effects. The third essay extends the second essay to the panel data setting. The individual unobserved effects are allowed to be correlated with the explanatory variables. The treatment of the unobserved heterogeneity is addressed as a central issue. Estimation is based on an iterative Generalized Method of Moments (iGMM) procedure, with well-known instruments. Inference is robust to spatial heteroskedasticity and spatial autocorrelation. The performance of the iGMM estimator is evaluated through detailed simulation studies. Results show that the iGMM estimator tends to perform well in terms of computational time, accuracy and precision. The adequacy of the proposed approaches is also assessed through empirical applications on the U.S. Metropolitan Statistical Areas. The first essay analyzes environmental effects over regional competitiveness and the degree of competitiveness spillovers. A new definition for binary competitiveness is introduced. Results show that competitiveness is significantly affected by air quality. Also, being competitive plays an important role in the competitiveness of neighboring areas. The third essay discusses regional knowledge and innovation spillovers, based on the proportion of high-tech patents. Results show that human capital plays a major role in regional innovative processes. However, due to regional aggregation, the degree of knowledge spillovers is significantly low.

Acknowledgements

First, I would like to express my deepest gratitude to my Ph.D. advisor, Isabel Proença, for all the availability, close guidance, profound knowledge and extreme patience, over the last 6 years.

Second, I would like to thank Zhenlin Yang and to all the participants of the XI and XII World Conference of the Spatial Econometrics Association (SEA), for the insightful comments and suggestions.

Third, I would like to thank my wonderful wife, Joana, a lifelong companion, for always supporting me in the pursuit of my personal goals; for always cheering me up and motivating me, even in hard times; for always believing in my capabilities; for always being by my side.

Fourth, I would like to thank my parents, Filomena and Adelino, for their emotional and financial support throughout the years, that allowed me to achieve my personal and academic goals; for acknowledging that academic research can, indeed, make a difference and be recognized; for always trying to take the best out of me.

Fifth, I would like to thank my sister, Patrícia, and to my brother-in-law, David, for being supportive and for raising two of the most pure and wonderful children I have ever met, my nephews João and Pedro.

Sixth, I would like to thank my uncle-in-law, Jorge, my aunt-in-law, Deolinda, and my cousin-in-law, Eduardo, for also being supportive; for the emotional talks; for the great dinners and the amazing wines.

Seventh, I would like to thank to my friends, Vando, Ricardo and Vítor, brothers from other mothers, for always making an effort to understand my research; for always caring about my mental health; for teaching me what true friendship feels like.

Eighth, I would like to thank the team of *Gabinete de Fisioterapia no Desporto (GFD)*, in particular to Ernesto, Vanessa, Cristina, Mafalda, Nuno, Hilário and Tiago, for always putting me up to any task and for continuously helping me relieving my back (and, recently, shoulder and arm) pain.

Ninth, I would like to thank Alexandre for always pushing me through the "finish line". Pressure, focus and a sandwich!

Tenth, I would like to thank my dog, Leia, for the unconditional love, affection and companionship, that is worth learning from. She is one of a kind.
Last but not the least, I would like to acknowledge the Portuguese Foundation for Science and Technology (*Fundação para a Ciência e a Tecnologia*) for the financial support through national funds (Ph.D. Scholarship SFRH/BD/92277/2013). I also would like to acknowledge the *Centro de Matemática Aplicada à Previsão e Decisão Económica (CEMAPRE)* for the financial support through the program CEMAPRE-UID/Multi/00491/2013.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the words of Anselin (1988), Spatial Econometrics consists "of those methods and techniques that, based on a formal representation of the structure of spatial dependence and spatial heterogeneity, provide the means to carry out the proper specification, estimation, hypothesis testing, and prediction for models in regional science". Spatial heterogeneity is a particular case of coefficient instability. This form of variability is driven by factors such as distance, location and/or regional size. The treatment of spatial heterogeneity in econometric modeling can be addressed in three ways. One, as a special case of unobserved heterogeneity, commonly used in panel data settings (Baltagi, 2013). Two, as a small number of regimes, with heterogeneous regression coefficients and/or explanatory variables (see section 9.4.3. of Anselin, 1988, – "Spatially Switching Regressions"). Three, as a large number of regimes, with heterogeneous regression coefficients and/or explanatory variables as well (Gamerman et al., 2003; Gelfand et al., 2003, – Random Coefficients Models). Spatial dependence is a special case of cross-sectional dependence. It is related to the way that different units interact in space. Spatial interactions are usually expressed by the means of a spatial weighting matrix, where each element represents the relative importance of a given spatial unit on its neighbors. The definition of the spatial weights typically follows geostatistical concepts, such as contiguity (Cressie, 2015) and nearest neighbor distances (Cliff and Ord, 1981), or uses proxies for economic "distances" (Case et al., 1993). Under an econometric framework, spatial dependence can be approached and interpreted in two ways. First, by introducing spatial correlation through the dependent variable – the Spatial Lag Model (SLM).

This is an equilibrium model. It specifies that the values of the outcomes, for every unit in the spatial system, are jointly determined. Second, by introducing spatial correlation through the error term – the Spatial Error Model (SEM). This is a special case of non-spherical error and does not require a formal specification of the spatial interactions. Combinations of the two previous specifications are possible. Without loss of generality, this thesis will focus on the specification and estimation of the former. Due to the simultaneous nature of the Spatial Lag Model, computational and/or interpretative issues generally occur. Of particular relevance are the ways that such problems can be addressed under nonlinear frameworks.

In the last 30 years, the access to regional-level and firm-level data, as well as the development of geostatistical software, has substantially increased. At the same time, new challenges were posed to both theoretical and applied researchers in Spatial Econometrics. One of the growing areas of research in this field concentrates on the specification, estimation and inference for spatial models with discrete, limited and fractional outcomes. Unlike the continuous unbounded dependent variable case, considering a spatial linear approach to model discrete, limited and fractional dependent variables can produce misleading predictions, implausible estimated magnitudes for the effects on the predicted outcomes and incorrect statistical inference (see Case, 1992; McMillen, 1992). Alternatively, several examples can be found in the literature, that properly address the previous issues. Naming a few, Pinkse and Slade (1998), Holloway et al. (2002), Beron et al. (2003) and Smith and LeSage (2004), for dichotomous dependent variables; Wang and Kockelman (2009) and Roorda et al. (2010) for ordered dependent variables; Bolduc et al. (1997), Garrido and Mahmassani (2000) and Chakir and Parent (2009) for multinomial dependent variables; LeSage (2000), Qu and Lee (2012), Qu and Lee (2013) and Xu and Lee (2015b) for limited dependent variables; Agarwal et al. (2002) and Lambert et al. (2010) for count dependent variables; Lin and Lee (2010) and Xu and Lee (2015a) for fractional dependent variables.

With regard to the specification of Spatial Lag Models for discrete, limited and fractional outcomes, Qu and Lee (2012) define a taxonomy that allows to distinguish two ways to address the specification of such spatial models. They differ in

the modeling approach and in the way the spatial interactions are interpreted. The first specification considers a latent variable approach. A spatial lag of the latent dependent variable is introduced in the specification of a linear latent variable model. This can be viewed as a linear structural model. In addition, it can be written as a reduced form, assuming that standard stability conditions hold, such that a particular matrix – the spatial lag operator – is invertible. Under this approach, it follows that the stance of a given agent, towards a particular decision, is determined by the stance of neighboring agents, towards the same decision. In this way, the spatial lag latent variable model is an equilibrium model for decision. The second specification follows a nonlinear simultaneous framework. A spatial lag of the dependent variable enters the specification in a nonlinear way, inside a nonlinear function. Similarly, this can be viewed as a structural model. However, in this case, obtaining a reduced form relies in complex numerical methods. In general, no analytically tractable formula is available. Additionally, this approach is no longer appropriate to address problems involving the decision of agents. The nonlinear spatial lag model is a simultaneous model for behavior (eventually in an equilibrium state). Here, the effective behavior of a given agent is determined by the behavior of neighboring agents.

In this thesis, particular attention is given to the specifications discussed by Qu and Lee (2012). The latent variable approach is used to derive a spatial lag model for dichotomous outcomes. The nonlinear simultaneous framework is used to define a spatial lag model for fractional responses. Of interest are the issues related to estimation and interpretation of the previous Spatial Lag Models. For the latent-based binary choice model, estimation is known to be computationally burdensome. Maximum Likelihood based methods and/or Bayesian based methods require N dimensional integrals to be computed, with N the sample size. Generalized Method of Moments (GMM) estimators require the computation of N dimensional matrix operations, most of them involving the inversion of the spatial lag operator, a full $N \times N$ matrix. In Chapter 2 of this thesis, a computationally simple method is developed to approximate the inverse of the spatial lag operator and the corresponding matrix operations. Under a GMM framework, estimation time and complexity are significantly reduced. For the fractional response model, computational complexity

is no longer a problem. However, specification and interpretation issues come to light. In addition, the literature on this subject is scarce. Lin and Lee (2010) consider a linear approach to model spatially lagged fractional responses. While being a simple and popular starting point, it has well known pitfalls. Xu and Lee (2015a) consider a nonlinear framework to model spatially lagged fractional responses with nonadditive errors. They apply a nonlinear transformation to the responses, such that the transformed model becomes linear. As a result, their approach is not appropriate for the case where the responses are defined at the boundaries, zero and one. In Chapter 3 of this thesis, a nonlinear spatial lag model is proposed. No transformations are used and observations at the boundaries are allowed. Moreover, an approximation for the proposed nonlinear spatial lag model is also developed. In this way, the model can be written as an approximate reduced form, with a tractable analytic expression. This facilitates the interpretation of the reduced form parameters and the analytic determination of (approximated) policy effects. In Chapter 4 of this thesis, the specifications presented in Chapter 3 are extended to the spatial panel framework.

Estimation of the proposed approaches is based on the iterative GMM procedure of Klier and McMillen (2008). The spatial heteroskedasticity and spatial autocorrelation (spatial HAC) robust estimator of Kelejian and Prucha (2007) is considered, to produce valid inference for the asymptotic covariance estimator of the GMM estimator for the unknown parameter vector.

Extensive Monte Carlo simulation studies are conducted in Chapter 2 and Chapter 3. The performance of the iterative GMM procedure is assessed in detail. Empirical applications are presented in Chapter 2 and Chapter 4. The adequacy of the methods and specifications developed in this thesis is illustrated using real spatial data.

The remainder of this thesis is organized as follows. In Chapter 2, which has the title *The inversion of the spatial lag operator in binary choice models: fast computation and a closed formula approximation*, a new method to approximate the inverse of the spatial lag operator, used in the estimation of spatial lag models for binary outcomes is presented. Related matrix operations are approximated as well. Closed

formulas for the elements of the approximated matrices are deduced. A GMM estimator is also presented. This estimator is a variant of Klier and McMillen (2008) iterative GMM estimator. The approximated matrices are used in the gradients of the new iterative GMM procedure. Monte Carlo experiments suggest that the proposed approximation is accurate and allows to significantly reduce the computational complexity, and consequently the computational time, associated with the estimation of spatial binary choice models, especially for the case where the spatial weighting matrix is large and dense. Also, the simulation experiments suggest that the proposed iterative GMM estimator performs well in terms of bias and root mean square error and exhibits a minimum trade-off between computational time and unbiasedness within a class of spatial GMM estimators. Finally, the new iterative GMM estimator is applied to the analysis of competitiveness in the U.S. Metropolitan Statistical Areas. A new definition for binary competitiveness is introduced. The estimation of spatial and environmental effects are addressed as central issues.

In Chapter 3, which has the title *Fractional responses with spatial dependence*, two specifications to estimate models for spatially lagged fractional responses with additive errors are introduced. No transformations are applied to the responses and the suggested specifications can handle observations at the boundaries, zero or one. Derivation and computation of the partial effects are addressed as central issues. The first specification, the Fractional Response Spatial Lag Model (FRSLM), extends the seminal approach of Papke and Wooldridge (1996) to spatial frameworks. A spatial lag of the dependent variable is introduced in the specification. The resulting nonlinear simultaneous model has no analytically tractable expression for the reduced form. Policy effects become difficult to interpret. The second specification, the approximate Fractional Response Spatial Lag Model (aFRSLM), allows to write the FRSLM as an analytically tractable approximate reduced form. The true partial effects are approximated by sums of nonlinear functions of the exogenous explanatory variables and their spatially lagged values. An extensive Monte Carlo simulation study is presented. The finite and large sample properties of the GMM estimator for the two proposed specifications and the corresponding partial effects are investigated. Experiments show that both the FRSLM and aFRSLM perform

well in terms of bias and root mean square error for a variety of sampling designs. The aFRSLM also proves to be accurate in terms of the estimation of the partial effects.

In Chapter 4, which has the title *Unobserved heterogeneity in spatial panel data models for fractional responses: an application to the proportion of high-tech patents in the U.S. Metropolitan Statistical Areas*, the knowledge spillovers in the U.S. Metropolitan Statistical Areas, are studied. The proportion of U.S. origin high-tech patents, between 2010 and 2015, is used as a proxy to measure the innovation output. Its spatial lag is used to capture the degree of regional concentration. The R&D expenditures at Colleges and Universities by source of funds and Human Capital proxies (wages and employment, by educational attainment) are used as measures for the inputs of the innovative process. The approach presented in Chapter 3 is extended to the panel data setting. Unobserved effects are allowed to be correlated with the explanatory variables. They are modeled according to the spatial approach of Debarsy (2012), a generalization of the classic Chamberlain-Mundlak approach. The partial effects are derived. Average direct effects and average indirect effects are of particular interest. Estimates are compared with those from panel data estimators for linear spatial lag models. The adequacy of the proposed spatial panel specifications is demonstrated. Results show that human capital plays a major role in regional innovative processes. However, the degree of high-tech patents concentration among the U.S. Metropolitan Statistical Areas is significantly low.

Chapter 5 presents the concluding remarks and discusses further research on the subjects addressed in this thesis.

Chapter 2

The inversion of the spatial lag operator in binary choice models: fast computation and a closed formula approximation

This chapter has been published in co-authorship with Isabel Proença

Silveira Santos, L. and I. Proença (2019): "The inversion of the spatial lag operator in binary choice models: Fast computation and a closed formula approximation", *Regional Science and Urban Economics*, 76, 74–102.

2.1. Introduction

Modeling binary choice outcomes with spatial dependence has become increasingly popular in recent years. Many applications can be found in the literature, that cover, for example, the choice on the participation in environmental policies (Beron et al., 2003; Murdoch et al., 2003), the adoption of new technologies in agriculture (Case, 1992; Holloway et al., 2002; Wollni and Andersson, 2014), the implementation of state income taxes (Beron and Vijverberg, 2004; Fiva and Rattsø, 2007), the location choice (Klier and McMillen, 2008; Miyamoto et al., 2004), the decision to (re)open a business (Holloway and Lapar, 2007; LeSage et al., 2011) or the existence of high crime rates in a given neighborhood (McMillen, 1992). However, the introduction of spatial dependence in models with dichotomous dependent variables raises several complications. Considering a latent variable approach to derive a model for

binary choice outcomes, Anselin (2007) shows that spatial dependence implies the presence of spatial heteroskedasticity and spatial autocorrelation, which leads to specification issues and to analytically intractable expressions for the quantities of interest. As a result, estimation becomes complex and computationally demanding.

To address the issues related to the estimation of spatial binary choice models, several approaches have been proposed. These approaches can be categorized into three major groups, according to the estimation method that they address: Maximum Likelihood (ML) methods, Bayesian methods and Generalized Method of Moments (GMM) estimators. Examples of ML based approaches are the EM algorithm (McMillen, 1992), the RIS simulator (Beron and Vijverberg, 2004), partial ML estimation based on pairwise correlations (Bhat, 2011; Wang et al., 2013), the GHK simulator (Pace and LeSage, 2016) and the Mendel-Elston approximation (Martinetti and Geniaux, 2017). The Gibbs sampler (LeSage, 2000) and Markov Chains Monte Carlo (Smith and LeSage, 2004) are examples of Bayesian based approaches. Finally, the estimator of Pinkse and Slade (1998) and the estimator of Klier and McMillen (2008) consider the GMM framework. Nevertheless, most of these approaches become computationally burdensome in large samples.

The computational issues associated with Maximum Likelihood methods and Bayesian methods are related to the computation or simulation of high-dimensional integrals. This is a consequence of requiring the specification of the joint distribution (or, at least, some structure of the distribution) of the spatial data. Even if the high-dimensional integrals are approximated by one-dimensional integrals (Martinetti and Geniaux, 2017) or obtained by simulation algorithms (Beron and Vijverberg, 2004; Pace and LeSage, 2016), estimation can still become computationally infeasible, especially if the spatial units are influenced by many neighbors and N (the sample size) is large.

Another possibility is to consider a GMM estimation approach. Under the GMM framework, the distributional assumptions can be relaxed in such a way that no high-dimensional integration is involved. In fact, GMM only requires that a set of moment conditions is correctly specified. But, even so, estimation becomes computationally impracticable in large samples, due to N-dimensional matrix operations

that have to be computed on each iteration. Nevertheless, these computational issues can be tackled through matrix approximation methods. The Taylor series approximation, the Chebyshev approximation (Pace and LeSage, 2004), the eigen-decomposition of a matrix, the Cholesky decomposition (Pace and Barry, 1997a,b), the LU decomposition or the conjugate gradient method (Smirnov, 2005, 2010) are examples of approximation methods that are commonly used in spatial frameworks. However, some of these approximation methods can be computationally demanding, especially when N is large, and their accuracy depends on the nature of the approximated matrices.

The purpose of this work is twofold. Firstly, it suggests a new approximation method to deal with the computational issues related to the N-dimensional matrix operations required in the GMM estimation of spatially lagged models for binary choice outcomes. The new approximation method focuses on the approximation of the spatial lag operator inverse, since every matrix operation required in the GMM estimation procedure involves the computation of this inverse. The setup for the proposed approximation method relies on non-restrictive assumptions about the spatial weighting matrix and allows to accommodate scenarios where the spatial weighting matrix can be symmetric and non-symmetric. Considering the series expansion of the inverse and the limiting properties of the eigenstructure of normalized spatial weighting matrices, it is shown that the spatial lag operator inverse can be approximated by a sum of known matrices and a simple matrix-vector product. As a result, other related N-dimensional matrix operations can be straightforwardly approximated, as well. Also, closed formulas for the elements of the approximated matrices are available and are deduced. They are especially useful to determine the partial effects.

Secondly, it proposes a computationally simple iterative GMM estimator. This estimator is based on the iterative GMM procedure of Klier and McMillen (2008) together with the approximated matrices deduced in the first part of this paper. This approach has two important advantages. One, the moment conditions of the suggested iterative GMM estimator correspond to orthogonality conditions that use

only the information in marginal distributions. In this way, no higher order integration is required. Two, the approximated matrices are used in the gradients of the iterative procedure. This allows to significantly reduce the computational complexity and computational time of the suggested GMM estimator, when compared to the traditional GMM estimator. In addition, the spatial heteroskedasticity and spatial autocorrelation robust estimator of Kelejian and Prucha (2007) is used to overcome potential biases in the estimated asymptotic covariance matrix of the GMM estimator for the unknown parameter vector. Note that, the computational simplicity associated with the GMM estimation comes at a cost of larger bias, in comparison to full information methods, where the joint distribution of the spatial data is used in the estimation.

It will be shown through a detailed simulation study that the proposed approximation method fairly approximates the matrices of interest, especially when N is large and the spatial weighting matrix is dense. In addition, the proposed iterative GMM estimator proves to be accurate, especially at low and moderate levels of spatial dependence. At high levels of spatial dependence, the spatial lag parameter tends to be overestimated, which is also a feature shared by other spatial GMM procedures. Moreover, using the approximated matrices in the GMM estimation, not only allows to reduce the associated computational complexity and the overall computational time, especially when N is large and the spatial weighting matrix is dense, but also it allows to increase the precision of the proposed iterative GMM estimator in comparison to other spatial GMM estimators.

The new estimation procedure is used to assess how environmental indicators contribute to influence regional competitiveness in the U.S. Metropolitan Statistical Areas, from 2001 to 2016. A new Binary Competitiveness Indicator (BCI) is introduced. The new competitiveness indicator is based on three dimensions: labor efficiency, capital efficiency and economic growth of the corresponding area. Results show a moderately high degree of spatial dependence between the U.S. Metropolitan Statistical Areas and evidence of an "U" shaped effect of the environmental indicators on regional competitiveness. Also, the suggested estimator exhibited a good performance, in terms of computational time and goodness-of-fit.

The remainder of this paper is organized as follows. Section 2.2 reviews the literature on the specification and estimation of spatial lag models for binary dependent variables. Section 2.3 reviews the literature on methods to approximate the inverse of the spatial lag operator and related matrix functions. Section 2.4 introduces the new method to approximate the inverse of the spatial lag operator. Section 2.5 derives the new iterative GMM estimation procedure. Section 2.6 conducts a set of Monte Carlo experiments to assess: firstly, the accuracy and the computational time of the proposed approximation method, compared with the other existing approximation methods; secondly, the statistical properties and the computational performance of the new iterative GMM estimator, compared with the traditional GMM estimator for spatial binary choice models and the GMM estimator of the linearized spatial lag model for binary dependent variables. Section 2.7 presents an empirical application on the environmental impacts over a spatially lagged Binary Competitiveness Indicator (BCI), in the U.S. Metropolitan Statistical Areas. Finally, section 2.8 concludes. The results of the Monte Carlo experiments are summarized in Appendix A2 and the estimation results of the empirical application are shown in Appendix B2.

2.2. Spatially lagged latent dependent variable model for binary outcomes

A spatial binary choice model can be derived based on the following spatially lagged latent variable specification:

$$Y_i^* = \alpha \sum_{i \neq j} w_{ij} Y_j^* + \mathbf{X}_i \boldsymbol{\beta} + \xi_i, \qquad i = 1, 2, \ldots, N \tag{1}$$

where Y_i^* is a general dependent variable (possibly not observable) for the unit i and N denotes the total number of spatial units. The coefficients w_{ij} are known non-negative scalars that refer to the spatial weight of unit j on unit i, with $j \neq i$ and $j = 1, 2, \ldots, N$. By convention, $w_{ii} = 0$, for all i. The scalar parameter α is the spatial lag parameter. The $1 \times K$ vector $\mathbf{X}_i$ includes the observations for a set of K

exogenous explanatory variables and a constant, for the unit i. The $K \times 1$ vector $\boldsymbol{\beta}$ is the corresponding vector of regression parameters. The disturbance term, ξ_i, is a random error for the unit i, with zero mean and is independent of $\mathbf{X}_i$, for all $i = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, K$.

Stacking over the cross-sectional units, the spatial lag model can be written as a reduced form for the dependent variable:

$$\mathbf{Y}^* = \alpha \mathbf{W} \mathbf{Y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi} = (\mathbf{I} - \alpha \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where $\mathbf{Y}^* = [Y_1^*, Y_2^*, \ldots, Y_N^*]^\intercal$ and $\mathbf{X} = [\mathbf{X}_1^\intercal, \mathbf{X}_2^\intercal, \ldots, \mathbf{X}_N^\intercal]^\intercal$. The error is now $\boldsymbol{\varepsilon} = (\mathbf{I} - \alpha \mathbf{W})^{-1} \boldsymbol{\xi}$, where $(\mathbf{I} - \alpha \mathbf{W})^{-1}$ is the spatial lag operator inverse and $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_N]^\intercal$. The $N \times N$ identity matrix is denoted by $\mathbf{I}$ and the $N \times N$ spatial weighting matrix is denoted by $\mathbf{W}$, with generic element w_{ij}.

If Y_i^* is observable, the conditional expectation is given by $\mathrm{E}(Y_i^* \,|\, \mathbf{X}, \mathbf{W}) = \mathbf{X}_i^{\#} \boldsymbol{\beta}$, where $\mathbf{X}_i^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha \mathbf{W})^{-1} \mathbf{X}$, and equation (2) defines a linear spatial lag model. Here, however, Y_i^* is not observable. The observed dependent variable is Y_i, a binary dependent variable, which is a function of particular characteristics of Y_i^* and defined as $Y_i = 1$ if $Y_i^* \geq 0$ and $Y_i = 0$ if $Y_i^* < 0$. The conditional expectation of a spatial lag model when Y_i^* is not observable and Y_i is a binary dependent variable follows as:

$$\begin{aligned} &\mathrm{E}(Y_i \,|\, \mathbf{X}, \mathbf{W}) = \mathrm{P}(Y_i = 1 \,|\, \mathbf{X}, \mathbf{W}) = \mathrm{P}(Y_i^* > 0 \,|\, \mathbf{X}, \mathbf{W}) \\ &= \mathrm{P}\left(\mathbf{X}_i^{\#}\boldsymbol{\beta} + \varepsilon_i > 0 \,\middle|\, \mathbf{X}, \mathbf{W}\right) = \mathrm{P}\left(\varepsilon_i > -\mathbf{X}_i^{\#}\boldsymbol{\beta} \,\middle|\, \mathbf{X}, \mathbf{W}\right) \\ &= 1 - \mathrm{P}\left(\varepsilon_i \leq -\mathbf{X}_i^{\#}\boldsymbol{\beta} \,\middle|\, \mathbf{X}, \mathbf{W}\right) = G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right), \qquad i = 1, 2, \ldots, N \end{aligned} \tag{3}$$

where $G(\eta)$ is a function that takes on values in the interval $0 < G(\eta) < 1$, for all $\eta \in \mathbb{R}$, and it is twice continuously differentiable, for all $\eta \in \mathbb{R}$, as well. Usually $G(\eta)$ is called the link function and η is called the index. It is further assumed that $G(\eta)$ is known[1] and given by the cumulative distribution function (CDF) of ξ_i conditional on $(\mathbf{X}, \mathbf{W})$. The parameter σ_i is the square root of the conditional

[1]Generally the link function, $G(\eta)$, is unknown and can be estimated using nonparametric and semiparametric methods. See Härdle et al. (2004) and Horowitz (2009) for details.

variance of ε_i, for each i, obtained from the diagonal elements of the conditional covariance matrix of $\boldsymbol{\varepsilon}$:

$$\operatorname{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}, \mathbf{W}) = \operatorname{Var}(\xi_i \mid \mathbf{X}, \mathbf{W}) \left[(\mathbf{I} - \alpha \mathbf{W})^{\intercal} (\mathbf{I} - \alpha \mathbf{W}) \right]^{-1} = \boldsymbol{\Sigma} \tag{4}$$

where $\operatorname{Var}(\xi_i \mid \mathbf{X}, \mathbf{W})$ is constant and fixed, to ensure identification. The scalar σ_i is strictly positive and finite, for all i, assuming that the rows and columns of the matrix $(\mathbf{I} - \alpha \mathbf{W})^{-1}$ are uniformly bounded in absolute value.

In most applications using binary response models, the conditional distribution of ξ_i is assumed to be a standard Normal distribution or a standard Logistic distribution. This implies that $\operatorname{Var}(\xi_i \mid \mathbf{X}, \mathbf{W}) = 1$ and $\operatorname{Var}(\xi_i \mid \mathbf{X}, \mathbf{W}) = \pi^2/3$, respectively, for all i. Under these two specifications, the probability distribution functions (PDFs) of the link functions are symmetric about zero, but this is generally not the case for other possible links[2].

Note that the specification in (3) is similar to the specification of McMillen (1992) and LeSage (2000). Under this approach, only the information in the marginal distributions of ε_i conditional on $(\mathbf{X}, \mathbf{W})$ is used. The implications of this approach, regarding estimation, are discussed in section 2.6.

Considering a generic link function, the spatial lag model for binary dependent variables follows as:

$$Y_i = G\left(\frac{\mathbf{X}_i^{\#} \boldsymbol{\beta}}{\sigma_i} \right) + u_i, \qquad i = 1, 2, \ldots, N \tag{5}$$

where u_i has zero mean and is independent of $\mathbf{X}_i^{\#}$ (thus, it is independent of $\mathbf{X}_i$), for all $i = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, K$. Note that u_i differs from ε_i because $u_i = Y_i - \mathrm{E}(Y_i \mid \mathbf{X}, \mathbf{W})$ and $\varepsilon_i = Y_i^* - \mathrm{E}(Y_i^* \mid \mathbf{X}, \mathbf{W})$. Hence, u_i is a discrete random variable assuming only two values, $1 - G(\cdot)$ and $-G(\cdot)$.

To estimate the model (5), a GMM approach is considered, based on the works of Pinkse and Slade (1998) and Klier and McMillen (2008). It is assumed that the

[2]See, for example, the complementary log-log link or the Weibull link.

unknown parameters $\boldsymbol{\beta}$ and α satisfy the following moment condition:

$$\mathrm{E}\left(\mathbf{Z}^{\intercal}\mathbf{u}_{*}\right)=\mathbf{0} \tag{6}$$

where $\mathbf{Z}$ is the $N\times(K+p)$ matrix of instruments, with p the number of additional instruments that are usually given by the product between the powers of $\mathbf{W}$ and the matrix of explanatory variables. The $N\times 1$ vector $\mathbf{u}_{*}$ correspond to the "generalized errors" (Gourieroux et al., 1987):

$$u_{*,i}=\frac{\left[Y_i-G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)\right]g\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)}{G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)\left[1-G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)\right]},\qquad i=1,2,\ldots,N \tag{7}$$

where the function $g\left(\cdot\right)$ is the first derivative of $G\left(\cdot\right)$ w.r.t. the index. The GMM estimates of the parameter vector, $\boldsymbol{\Theta}=\left(\boldsymbol{\beta},\alpha\right)^{\intercal}$, are obtained by minimizing the objective function:

$$\mathcal{Q}\left(\boldsymbol{\beta},\alpha\right)=\mathbf{u}_{*}^{\intercal}\mathbf{Z}\boldsymbol{\Xi}\mathbf{Z}^{\intercal}\mathbf{u}_{*} \tag{8}$$

where $\boldsymbol{\Xi}$ is a $(K+p)\times(K+p)$ symmetric positive definite matrix. Klier and McMillen (2008) sets $\boldsymbol{\Xi}=\left(\mathbf{Z}^{\intercal}\mathbf{Z}\right)^{-1}$ and the GMM estimator reduces to nonlinear two stages least squares (N2SLS). However, because the minimization problem in (8) does not have a closed formula, an iterative procedure is used to obtain a solution for the unknown parameters. The following steps are considered:

1. Assume initial values for the parameter vector $\boldsymbol{\Theta}=\left(\boldsymbol{\beta},\alpha\right)^{\intercal}$, $\boldsymbol{\Theta}^{(0)}$, and compute the gradients evaluated at the initial values, $\boldsymbol{\Gamma}_i^{(0)}=\left(\partial u_{*,i}/\partial\boldsymbol{\Theta}\right)|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(0)}}$, $i=1,2,\ldots,N$.

2. Regress $\boldsymbol{\Gamma}^{(0)}$ on $\mathbf{Z}$, in a similar fashion to (linear) 2SLS. Obtain $\hat{\boldsymbol{\Gamma}}^{(0)}$.

3. Construct new estimates as $\boldsymbol{\Theta}^{(1)}=\boldsymbol{\Theta}^{(0)}+\left[\left(\hat{\boldsymbol{\Gamma}}^{(0)}\right)^{\intercal}\left(\hat{\boldsymbol{\Gamma}}^{(0)}\right)\right]^{-1}\left(\hat{\boldsymbol{\Gamma}}^{(0)}\right)^{\intercal}\mathbf{u}_{*}^{(0)}$, where $\mathbf{u}_{*}^{(0)}$ are the generalized residuals evaluated at the estimates of step 0.

4. Repeat steps 1. to 3., using the estimates from the last iteration, until the algorithm converges.

The spatial heteroskedasticity and spatial autocorrelation robust covariance estimator of the (iterative) GMM estimator follows as in Kelejian and Prucha (2007):

$$\widehat{Avar\left(\hat{\boldsymbol{\Theta}}\right)} = \left(\sum_{i=1}^{N}\hat{\boldsymbol{\Gamma}}_i^{\intercal}\hat{\boldsymbol{\Gamma}}_i\right)^{-1}\left\{\sum_{i=1}^{N}\hat{u}_i^2\hat{\boldsymbol{\Gamma}}_i^{\intercal}\hat{\boldsymbol{\Gamma}}_i + \sum_{j=1}^{n-1}\left[K\left(\frac{j}{d^*}\right)\sum_{i=1}^{n-j}\hat{u}_i\hat{u}_{i+j}\left(\hat{\boldsymbol{\Gamma}}_i^{\intercal}\hat{\boldsymbol{\Gamma}}_{i+j} + \hat{\boldsymbol{\Gamma}}_{i+j}^{\intercal}\hat{\boldsymbol{\Gamma}}_i\right)\right]\right\}\left(\sum_{i=1}^{N}\hat{\boldsymbol{\Gamma}}_i^{\intercal}\hat{\boldsymbol{\Gamma}}_i\right)^{-1} \tag{9}$$

where $K\left(j/d^*\right)$ is a Kernel function with $K:\mathbb{R}\rightarrow[-1,1]$, $K\left(0\right)=1$, $K\left(j/d^*\right)=K\left(-j/d^*\right)$ and $K\left(j/d^*\right)=0$, for $|j/d^*|>1$, that satisfies $|K\left(j/d^*\right)-1|\leq c_K|j/d^*|^{\rho_K}$, for $|j/d^*|\leq 1$, for some $\rho_K\geq 1$ and a finite positive c_K. The scalar d^* is a distance threshold.

The individual gradients for each parameter are:

$$\left(\boldsymbol{\Gamma}_{\boldsymbol{\beta}}\right)_i = \frac{\partial u_{*,i}}{\partial\boldsymbol{\beta}^{\intercal}} = -u_{*,i}\left(\frac{g'\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)}{g\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)} - u_{*,i}\right)\frac{\mathbf{X}_i^{\#}}{\sigma_i},\ i=1,2,\ldots,N \tag{10}$$

and

$$\begin{aligned}\left(\boldsymbol{\Gamma}_{\alpha}\right)_i &= \frac{\partial u_{*,i}}{\partial\alpha}\\ &= -u_{*,i}\left(\frac{g'\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)}{g\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)} - u_{*,i}\right)\left[\frac{1}{\sigma_i}\left(\mathbf{H}_i\boldsymbol{\beta} - \frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{2\sigma_i^2}\Upsilon_{ii}\right)\right],\ i=1,2,\ldots,N\end{aligned} \tag{11}$$

where $g'\left(\cdot\right)$ is the first derivative of the function $g\left(\cdot\right)$ w.r.t. the index, $\mathbf{H}_i$ is the ith row of the matrix product $\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\mathbf{W}\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\mathbf{X}$ and Υ_{ii} is the ith element of the diagonal of the matrix:

$$\begin{aligned}\boldsymbol{\Upsilon} &= \left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\left\{\mathbf{W}\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1} + \left[\mathbf{W}\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\right]^{\intercal}\right\}\left[\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\right]^{\intercal}\\ &= \left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\mathbf{W}\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\left[\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\right]^{\intercal} +\\ &\quad + \left\{\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\mathbf{W}\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\left[\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\right]^{\intercal}\right\}^{\intercal}\end{aligned} \tag{12}$$

with the diagonal of $\boldsymbol{\Upsilon}$ equal to

$$\operatorname{diag}(\boldsymbol{\Upsilon}) = 2 \times \operatorname{diag}\left((\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1}\left[(\mathbf{I} - \alpha\mathbf{W})^{-1}\right]^{\intercal}\right) \tag{13}$$

Having closed formulas for the gradients help to accelerate the numerical optimization process. However, because they depend on the spatial lag operator inverse, which has to be computed on each iteration, the estimation procedure becomes computationally burdensome, especially for large samples and/or dense spatial weighting matrices. To solve this issue, Klier and McMillen (2008) suggest a first order Taylor series approximation of model (5) around $\alpha = 0$. In this way, the previous iterative GMM procedure is simplified, because no large matrices need to be inverted. The major drawback of this approach is related to the poor accuracy of the estimates for the spatial lag parameter when $\alpha > 0.5$.

Another possibility is to address the previously mentioned computational issues through the approximation of the spatial lag operator inverse. Under this approach, the nonlinearity of model (5) is preserved and the estimates for the spatial lag parameter yield reasonable results for all admissible values of α. The methods that are commonly used in the literature are presented in the section below.

2.3. Approximation methods for the spatial lag operator inverse

To deal with the computational issues related to the inversion of the spatial lag operator, several methods have been proposed in the literature. These methods approach the matrix inversion explicitly or implicitly. For the explicit methods, the $N \times N$ spatial lag operator inverse is explicitly obtained; examples are the Taylor series approximation, the Chebyshev approximation (Pace and LeSage, 2004) and the eigendecomposition of the spatial weighting matrix. For the implicit methods, a system that involves the spatial lag operator (usually a matrix-vector product) is solved and a $N \times 1$ vector is obtained rather than a $N \times N$ matrix; examples are the Cholesky decomposition (Pace and Barry, 1997a,b), the LU decomposition and the conjugate gradient (Smirnov, 2005, 2010). Before presenting the details of the

previous methods, a set of assumptions are stated. Note that these assumptions are already commonly used in the literature of spatial binary choice models (see e.g., Baltagi et al., 2016; Billé, 2013).

2.3.1. Assumptions

Focusing on the properties of an initial spatial weighting matrix, $\mathbf{W}_0$, it is assumed that:

Assumption 2.1. The matrix $\mathbf{W}_0$ is non-stochastic and diagonalizable.

Assumption 2.2. All of the diagonal elements of $\mathbf{W}_0$ are equal to zero.

Assumption 2.3. The matrix $(\mathbf{I} - \alpha\mathbf{W}_0)$ is non-singular for all $\alpha \in \left(-1/|\lambda|_{0,\max}, 1/|\lambda|_{0,\max}\right)$, where $|\lambda|_{0,\max}$ is the largest absolute eigenvalue of $\mathbf{W}_0$. Additionally, $|\lambda|_{0,\max}$ is assumed to be bounded away from zero by some fixed constant $c_{|\lambda|_{0,\max}}$.

Assumption 2.4. Both row and column sums of $\mathbf{W}_0$ and $(\mathbf{I} - \alpha\mathbf{W}_0)^{-1}$ are uniformly bounded in absolute value by some constant $c_{\mathbf{W}_0}$, with $0 < c_{\mathbf{W}_0} < \infty$.

Assumption 2.5. The matrix $\mathbf{W}$ is row-normalized and equal to $\mathbf{D}_R^{-1}\mathbf{W}_0$, where $\mathbf{D}_R$ is a $N \times N$ diagonal matrix whose diagonal elements are the row sums of $\mathbf{W}_0$.

Remark 2.1. Under Assumption 2.5 the following properties are verified:

(a) $\mathbf{W}$ is non-symmetric;

(b) The eigenvalues of $\mathbf{W}$ are, in absolute value, less than or equal to one;

(c) The largest absolute eigenvalue of $\mathbf{W}$, $|\lambda|_{\max}$, is equal to one;

(d) The eigenvector of $\mathbf{W}$ associated with the largest absolute eigenvalue is the vector of ones, $\boldsymbol{\iota}$;

(e) The matrix $(\mathbf{I} - \alpha\mathbf{W})$ is non-singular for all $\alpha \in (-1, 1)$ – (see also Kelejian and Robinson, 1995).

Alternatively, $\mathbf{W}_0$ can be normalized through the transformation $\mathbf{W}_{sim} = \mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}$ (Ord, 1975), where $\mathbf{W}_{sim}$ is a $N \times N$ matrix that is similar to $\mathbf{W}$. In fact, $\mathbf{W}_{sim}$ can be written as $\mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2}$. By definition, the eigenvalues of $\mathbf{W}_{sim}$

and $\mathbf{W}$ are equal, which imply that their eigenvectors are directly related. Also, in general, $\mathbf{W}_{sim}$ is non-symmetric. However, if $\mathbf{W}_0$ is symmetric, then $\mathbf{W}_{sim}$ is symmetric as well[3]. A later discussion on the eigendecomposition of $\mathbf{W}$ will recover this result.

2.3.2. Explicit methods

Consider the Taylor series expansion of the inverse:

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} = \mathbf{I} + \alpha\mathbf{W} + \alpha^2\mathbf{W}^2 + \alpha^3\mathbf{W}^3 + \ldots = \sum_{h=0}^{\infty} \alpha^h \mathbf{W}^h \tag{14}$$

which converges absolutely for all $\alpha \in (-1, 1)$. Following LeSage and Pace (2009), the series (14) is partitioned into a finite lower-order and an infinite higher-order series:

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} = \sum_{h=0}^{q} \alpha^h \mathbf{W}^h + \sum_{h=q+1}^{\infty} \alpha^h \mathbf{W}^h \tag{15}$$

As suggested by several authors (Arbia, 2014; Elhorst, 2014; LeSage and Pace, 2009, to name a few), for the case where α quickly converges to zero, $(\mathbf{I} - \alpha\mathbf{W})^{-1}$ can be accurately approximated through the finite lower-order series:

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} \approx \sum_{h=0}^{q} \alpha^h \mathbf{W}^h \tag{16}$$

where q is small. The expression (16) corresponds to the Taylor series approximation of the spatial lag operator inverse.

The Chebyshev approximation (Pace and LeSage, 2004) for the spatial lag operator inverse is:

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} \approx \left[\sum_{h=0}^{q} c_l(\alpha)\, T_h(\mathbf{W})\right] - \frac{1}{2} c_0(\alpha)\, \mathbf{I} \tag{17}$$

[3] If $\mathbf{W}_0$ is symmetric, then $\mathbf{W}_0 = \mathbf{W}_0^\intercal$. Replacing $\mathbf{W}_0$ by $\mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}$ yields $\mathbf{W}_{sim} = \mathbf{W}_{sim}^\intercal$.

where

$$c_l(\alpha) = \frac{2}{q+1} \sum_{m=1}^{q+1} f(x_m) \cos\left(\frac{\pi\, l\left(m-\frac{1}{2}\right)}{q+1}\right) \tag{18}$$

$$x_m = \cos\left(\frac{\pi\left(m-\frac{1}{2}\right)}{q+1}\right) \tag{19}$$

$$f(x) = (1-\alpha x)^{-1} \tag{20}$$

and $T_{k+1}(\mathbf{W}) = 2\mathbf{W}T_k(\mathbf{W}) - T_{k-1}(\mathbf{W})$, for $k \geq 1$, with $T_0(\mathbf{W}) = \mathbf{I}$ and $T_1(\mathbf{W}) = \mathbf{W}$. The scalars $c_l(\alpha)$, with $l = 0, 1, 2, \ldots, q$ are the Chebyshev coefficients and depend on the spatial lag parameter. The functions $T_k(\mathbf{W})$, with $k = 0, 1, 2, \ldots, q$ are the Chebyshev polynomials and depend only on the spatial weighting matrix.

Finally, consider the eigendecomposition of the spatial lag operator inverse:

$$\begin{aligned}
(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{I} + \alpha\mathbf{W} + \alpha^2\mathbf{W}^2 + \alpha^3\mathbf{W}^3 + \ldots \\
&= \mathbf{I} + \alpha\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^2\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^2 + \alpha^3\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^3 + \ldots \\
&= \mathbf{I} + \alpha\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^2\mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^{-1} + \alpha^3\mathbf{V}\boldsymbol{\Lambda}^3\mathbf{V}^{-1} + \ldots \\
&= \mathbf{V}\left(\mathbf{I} + \alpha\boldsymbol{\Lambda} + \alpha^2\boldsymbol{\Lambda}^2 + \alpha^3\boldsymbol{\Lambda}^3 + \ldots\right)\mathbf{V}^{-1} \\
&= \mathbf{V}\left(\mathbf{I} - \alpha\boldsymbol{\Lambda}\right)^{-1}\mathbf{V}^{-1}
\end{aligned} \tag{21}$$

where the $N \times N$ diagonal matrix $\boldsymbol{\Lambda}$ contains the corresponding eigenvalues of $\mathbf{W}$ and the $N \times N$ matrix $\mathbf{V}$ contains, on each column, the ith eigenvector associated with the ith eigenvalue of $\mathbf{W}$. Contrary to the previous approximation methods, the expression in (21) is exact. Also, the inverse of the eigenvector matrix is only required to be computed once. Nevertheless, for the case where $\mathbf{W}_{sim}$ is symmetric, the eigenvectors in (21) can be expressed as orthogonal eigenvectors. In fact, replacing $\mathbf{W}_{sim}$ in (21) yields:

$$\begin{aligned}
(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{D}_R^{-1/2}\left(\mathbf{I} - \alpha\mathbf{W}_{sim}\right)^{-1}\mathbf{D}_R^{1/2} \\
&= \mathbf{D}_R^{-1/2}\mathbf{V}_{sim}\left(\mathbf{I} - \alpha\boldsymbol{\Lambda}\right)^{-1}\mathbf{V}_{sim}^{\intercal}\mathbf{D}_R^{1/2}
\end{aligned} \tag{22}$$

where the only matrix that is required to be computed is the $N \times N$ matrix $\mathbf{V}_{sim}$, that correspond to the orthogonal eigenvectors of $\mathbf{W}_{sim}$.

It is important to note that these methods can also be applied to approximate other matrix operations. In particular, they are useful to derive approximate or exact expressions for the matrix $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$, the diagonal elements of $\boldsymbol{\Upsilon}$ and the diagonal elements of $\boldsymbol{\Sigma}$, that are required in the computation of the gradients (10) and (11), on each iteration.

Focusing on the term $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$, the series expansion is given by:

$$\begin{aligned}(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{W}+2\alpha\mathbf{W}^2+3\alpha^2\mathbf{W}^3+\ldots \\ &= \sum_{h=0}^{\infty}(h+1)\,\alpha^h\mathbf{W}^{h+1}\end{aligned} \tag{23}$$

and the lower-order Taylor series approximation is:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} \approx \sum_{h=0}^{q}(h+1)\,\alpha^h\mathbf{W}^{h+1} \tag{24}$$

Therefore, the diagonal of $\boldsymbol{\Upsilon}$ is approximately equal to:

$$\operatorname{diag}(\boldsymbol{\Upsilon}) \approx 2\times\operatorname{diag}\left(\sum_{j=1}^{N}\left(\left[\sum_{h=0}^{q}(h+1)\,\alpha^h\mathbf{W}^{h+1}\right]\circ\left[\sum_{h=0}^{q}\alpha^h\mathbf{W}^h\right]\right)_{ij}\right) \tag{25}$$

where "$\circ$" is the Hadamard product operator. The previous expression implies that the diagonal elements of $\boldsymbol{\Upsilon}$ are approximately given by the row sums of the Hadamard product between the approximation of the matrix $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$ and the approximation of the matrix $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ – the spatial lag operator inverse.

Analogously, the Chebyshev approximation of $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$ is given by (17), where the Chebyshev coefficients are replaced by the function $f(x) = x/(1-\alpha x)^2$, and the eigendecomposition of $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$ is:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} = \mathbf{V}\boldsymbol{\Lambda}(\mathbf{I}-\alpha\boldsymbol{\Lambda})^{-2}\mathbf{V}^{-1} \tag{26}$$

For the case where $\mathbf{W}_{sim}$ is symmetric, (26) can be written as:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} = \mathbf{D}_R^{-1/2}\mathbf{V}_{sim}\boldsymbol{\Lambda}(\mathbf{I}-\alpha\boldsymbol{\Lambda})^{-2}\mathbf{V}_{sim}^{\intercal}\mathbf{D}_R^{1/2} \tag{27}$$

Again, note that (26) and (27) are exact expressions.

With regard to the diagonal elements of $\mathbf{\Sigma}$, they can be obtained as the row sums of the Hadamard square of the approximated or exact expression for the spatial lag operator inverse.

Nevertheless, these approaches can still be extremely demanding if the sample size is large and/or the spatial weighting matrix is dense. This is because, for both Taylor series and Chebyshev approximation, there are as many matrix operations as the number of lower-order powers of $\mathbf{W}$, and, for the eigendecomposition, the full eigensystem is required. Also, the approximate functional form for the elements of the approximated matrices is complicated, especially for the elements of the spatial lag operator inverse.

2.3.3. Implicit methods

The implicit methods used to compute the inverse of the spatial lag operator are based in the solution of the following equation:

$$(\mathbf{I} - \alpha\mathbf{W})\,\boldsymbol{\varepsilon} = \boldsymbol{\xi} \tag{28}$$

for $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\xi}$ are $N \times 1$ vectors.

Consider the LU decomposition, which generalizes the Cholesky decomposition to non-symmetric matrices. Following LeSage and Pace (2009), suppose that $(\mathbf{I} - \alpha\mathbf{W}) = \mathbf{LU}$. The solution for the system $\mathbf{LU}\boldsymbol{\varepsilon} = \boldsymbol{\xi}$ is identical to the solution for $\mathbf{L}\boldsymbol{\nu} = \boldsymbol{\xi}$, where $\boldsymbol{\nu} = \mathbf{U}\boldsymbol{\varepsilon}$.

The conjugate gradient method (Smirnov, 2005, 2010) is a numerical method that minimizes the quadratic form:

$$\mathfrak{f}\left(\boldsymbol{\varepsilon}\right) = \frac{1}{2}\boldsymbol{\varepsilon}^{\intercal}\left(\mathbf{I} - \alpha\mathbf{W}\right)\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\intercal}\boldsymbol{\xi} \tag{29}$$

based on orthogonal descent directions.

In contrast to the explicit methods, the implicit methods have the advantage that the $N \times N$ inverse is not explicitly computed. However, for the LU decomposition the spatial lag operator has to be decomposed into a lower triangular and

upper triangular matrix, which can be computationally demanding if the sample size is large and/or the spatial weighting matrix is dense. As for the conjugate gradient method, it may not converge for certain designs of $\mathbf{W}$ because it is not accurate for matrices that are not symmetric and positive definite. Also, no functional form for the elements of the spatial lag operator inverse is available.

2.4. The new explicit approximation method based on known matrices

In this section a new explicit method to obtain the inverse of the spatial lag operator is proposed. Consider the series expansion of the inverse in (14). The idea is to approximate the powers $h \geq 2$ of $\mathbf{W}$ by a "long run" matrix, $\mathbf{W}^{\infty}$, which is obtained from the limiting properties of the eigenstructure of $\mathbf{W}$ and it is equal to a simple matrix-vector product. In this way, no additional matrix operations are required and a closed formula for the elements of the spatial lag operator inverse are available and deduced. The details of this new procedure are presented and discussed below.

Consider that the assumptions of Section 2.3.1 hold. In addition, consider the following assumption on the eigenstructure of $\mathbf{W}$:

Assumption 2.6. The algebraic multiplicity of $|\lambda|_{\max}$, amult $(|\lambda|_{\max})$, is equal to one. For a block diagonal $\mathbf{W}$, the largest absolute eigenvalue of each block has algebraic multiplicity equal to one.

Note that, in general, the cases where the algebraic multiplicity of $|\lambda|_{\max}$ is greater than one are those where there are only one or two neighbors for every spatial unit. In practice, this assumption is not too restrictive, because, in most of the applications, there are more than two neighbors for every spatial unit or there are few spatial units with less than two neighbors. Nevertheless, this assumption can be relaxed, but at a cost of computational accuracy, as it will be shown in a Monte Carlo simulation study.

Now, under Assumption 2.1 to Assumption 2.6, the approximation of the spatial lag operator inverse is given by:

$$\begin{aligned}(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{I}+\alpha\mathbf{W}+\alpha^2\mathbf{W}^2+\alpha^3\mathbf{W}^3+\ldots+\alpha^q\mathbf{W}^q+\ldots \\ &\approx \mathbf{I}+\alpha\mathbf{W}+\alpha^2\mathbf{W}^\infty+\alpha^3\mathbf{W}^\infty+\ldots+\alpha^q\mathbf{W}^\infty+\ldots \\ &= \mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^\infty\end{aligned} \tag{30}$$

which converges absolutely for all α in the parameter space (see Assumption 2.5 and Kelejian and Robinson, 1995). The $N \times N$ matrix $\mathbf{W}^\infty$ is the "long run" matrix and equal to $\lim_{h\to\infty}\mathbf{W}^h$. Since the eigendecomposition of $\mathbf{W}$ is available, $\mathbf{W}^\infty$ can also be written as:

$$\begin{aligned}\mathbf{W}^\infty &= \lim_{h\to\infty}\mathbf{W}^h = \mathbf{V}\left(\lim_{h\to\infty}\mathbf{\Lambda}^h\right)\mathbf{V}^{-1} = \mathbf{V}\left[\lim_{h\to\infty}\operatorname{diag}\left(1^h,\lambda_2^h,\lambda_3^h\ldots,\lambda_N^h\right)\right]\mathbf{V}^{-1} \\ &= \mathbf{V}\begin{bmatrix}1 & 0 & \ldots & 0\\ 0 & 0 & \ldots & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \ldots & 0\end{bmatrix}\mathbf{V}^{-1} = \operatorname{col}(\mathbf{V})_1 \operatorname{row}\left(\mathbf{V}^{-1}\right)_1\end{aligned} \tag{31}$$

because the eigenvalues $\lambda_2, \lambda_3 \ldots, \lambda_N$ are, in absolute value, less than one. Also, $\lambda_1 = |\lambda|_{\max} = 1$ (see Assumption 2.5 and Assumption 2.6). The $N \times 1$ vector $\operatorname{col}(\mathbf{V})_1$ is the first column of $\mathbf{V}$ and the $1 \times N$ vector $\operatorname{row}(\mathbf{V}^{-1})_1$ is the first row of $\mathbf{V}^{-1}$. It is important to note that obtaining these vectors entail drastically different implications. On one hand, the expression for $\operatorname{col}(\mathbf{V})_1$ is exact and equal to $N \times 1$ vector of ones, $\boldsymbol{\iota}$ (see Assumption 2.5). On the other hand, to obtain $\operatorname{row}(\mathbf{V}^{-1})_1$ the entire linear system has to be solved, which becomes computationally infeasible in large samples.

Here, the issue related to the computation of $\operatorname{row}(\mathbf{V}^{-1})_1$ is addressed through the orthogonalization of the eigenvectors of $\mathbf{W}$, analogous to the approach presented in Section 2.3.2 for the eigendecomposition problem. The similar matrix, $\mathbf{W}_{sim}$, is used, because the eigenvectors are related to those of $\mathbf{W}$. However, as previously mentioned, $\mathbf{W}_{sim}$ is not necessarily symmetric, as it depends on the properties of the initial spatial weighting matrix, $\mathbf{W}_0$. Therefore, for the case where $\mathbf{W}_{sim}$ is

symmetric, row $(\mathbf{V}^{-1})_1$ can be straightforwardly written as a function of an orthogonal eigenvector. For the case where $\mathbf{W}_{sim}$ is not symmetric, a "symmetrization" procedure is suggested, such that row $(\mathbf{V}^{-1})_1$ can be approximated by a function of an orthogonal eigenvector.

In the next subsections, the exact and approximated expressions for row $(\mathbf{V}^{-1})_1$ are derived, according to the symmetric and non-symmetric scenarios of $\mathbf{W}_0$. Also, it will be shown that, the expressions for row $(\mathbf{V}^{-1})_1$ are based on known quantities. In this way, the approximated expressions for the elements of the spatial lag operation inverse will be derived, as well.

2.4.1. Case 1: symmetric $\mathbf{W}_0$

Consider that $\mathbf{W}_0$ is symmetric. Therefore, $\mathbf{W}_{sim}$ is also symmetric. Because $\mathbf{W}_{sim}$ can be written as a function of $\mathbf{W}$, consider the eigendecomposition for both matrices:

$$\mathbf{W}_{sim} = \mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} \Leftrightarrow \mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{\intercal} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \tag{32}$$

where $\mathbf{\Lambda}$ is equal in both sides of the equation due to matrix similarity. The equation above implies that the eigenvectors of $\mathbf{W}_{sim}$ and the eigenvectors of $\mathbf{W}$ are related in the following way:

$$\mathbf{V}_{sim} = \mathbf{D}_R^{1/2}\mathbf{V} \text{ and } \mathbf{V}_{sim}^{\intercal} = \mathbf{V}^{-1}\mathbf{D}_R^{-1/2}, \text{ but also } \mathbf{V}_{sim}^{\intercal} = \mathbf{V}^{\intercal}\mathbf{D}_R^{1/2} \tag{33}$$

Consider the hth power of (32):

$$\begin{aligned} \mathbf{V}_{sim}\mathbf{\Lambda}^h\mathbf{V}_{sim}^{\intercal} &= \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}^h\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \Leftrightarrow \\ \Leftrightarrow \mathbf{D}_R^{-1/2}\mathbf{V}_{sim}\mathbf{\Lambda}^h\mathbf{V}_{sim}^{\intercal}\mathbf{D}_R^{1/2} &= \mathbf{V}\mathbf{\Lambda}^h\mathbf{V}^{-1} \end{aligned} \tag{34}$$

Using the eigenvector relationship in (33) yields:

$$\mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\mathbf{V}\right)\mathbf{\Lambda}^h\left(\mathbf{V}^{\intercal}\mathbf{D}_R^{1/2}\right)\mathbf{D}_R^{1/2} = \mathbf{V}\mathbf{\Lambda}^h\mathbf{V}^{-1} \tag{35}$$

Now, as $h \to \infty$:

$$\begin{aligned}
\mathbf{W}^{\infty} &= \frac{1}{\left\|\mathbf{D}_R^{1/2}\text{col}\,(\mathbf{V})_1\right\|_2 \left\|\text{col}\,(\mathbf{V})_1^{\intercal}\,\mathbf{D}_R^{1/2}\right\|_2} \times \mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\text{col}\,(\mathbf{V})_1\,\text{col}\,(\mathbf{V})_1^{\intercal}\,\mathbf{D}_R^{1/2}\right)\mathbf{D}_R^{1/2} \\
&= \frac{1}{\left[\sqrt{\left(d_{R,1}^{1/2}\right)^2 + \left(d_{R,2}^{1/2}\right)^2 + \ldots + \left(d_{R,N}^{1/2}\right)^2}\right]^2} \times \boldsymbol{\iota}\boldsymbol{\iota}^{\intercal}\mathbf{D}_R \qquad (36) \\
&= \left(\sum_{i=1}^{N} d_{R,i}\right)^{-1} \times \mathbf{J}\mathbf{D}_R
\end{aligned}$$

where $d_{R,i}$ is the sum of the ith row of $\mathbf{W}_0$ and $\mathbf{J}$ is the $N \times N$ matrix of ones. The "long run" matrix, $\mathbf{W}^{\infty}$, is rescaled by the sum of all rows of $\mathbf{W}_0$ because the first eigenvector of $\mathbf{W}$ is now orthogonal. There are two major advantages related to this expression. First, the matrix $\mathbf{W}^{\infty}$ is given by a simple matrix-vector product, since $\mathbf{D}_R$ is a diagonal matrix. Second, each element of the matrix $\mathbf{W}^{\infty}$ have an exact closed formula:

$$w_{ij}^{\infty} = \left(\sum_{i=1}^{N} d_{R,i}\right)^{-1} \times d_{R,j} \qquad (37)$$

which implies that the rows of $\mathbf{W}^{\infty}$ are all equal and given by the sum of the ith row of $\mathbf{W}_0$, that is $\text{row}\,(\mathbf{W}^{\infty})_i = (d_{R,1}, d_{R,2}, \ldots, d_{R,N})$, for all i.

Plugging (36) into (30), the approximation of the spatial lag operator inverse is given by:

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} \approx \mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\left(\sum_{i=1}^{N} d_{R,i}\right)^{-1}\mathbf{J}\mathbf{D}_R \qquad (38)$$

which still converges absolutely for all α in the parameter space, because the expression for $\mathbf{W}^{\infty}$ is exact. Also, an approximate closed formula is available for the elements of $(\mathbf{I} - \alpha\mathbf{W})^{-1}$:

$$\left((\mathbf{I} - \alpha\mathbf{W})^{-1}\right)_{ij} \approx \mathbb{1}_{i=j} + \alpha \times w_{ij} + \frac{\alpha^2}{1-\alpha} \times \left(\sum_{i=1}^{N} d_{R,i}\right)^{-1} \times d_{R,j} \qquad (39)$$

where $\mathbb{1}_{i=j}$ is the indicator function that is equal to one if $i = j$ and equal to zero if $i \neq j$, for all $i, j = 1, 2, \ldots, N$. The accuracy of this approximation depends on

how fast the powers of the eigenvalues $\lambda_2, \lambda_3, \ldots, \lambda_N$ converge to zero, for a given value of α. In fact, this is a special case of the approximation method proposed by Griffith (2000), for linear models.

The expressions for the approximation of the spatial lag operator inverse, in (38) and (39), allow for an interesting interpretation of the product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$ which is approximately equal to $\mathbf{X} + \alpha\mathbf{W}\mathbf{X} + \alpha^2/(1-\alpha)^{-1}\left(\sum_{i=1}^{N} d_{R,i}\right)^{-1}\mathbf{J}\mathbf{D}_R\mathbf{X}$. This means that the previous matrix product can be decomposed into the original matrix $\mathbf{X}$, a spatial lag of the matrix $\mathbf{X}$ and the "long run" spatial lag of the matrix $\mathbf{X}$, that incorporates the bilateral effects (the combination of the neighboring effects on a given unit and the effects of a given unit on its neighbors).

Note that the previous results are valid when $\mathbf{W}$ is column-normalized or when $\mathbf{W}$ is doubly stochastic (simultaneously row- and column-normalized). For the first case, the approximation method is applied to $\mathbf{W}^\intercal$, because it is row stochastic. For the second case, considering that $\mathbf{W}_0$ is symmetric, the doubly stochastic $\mathbf{W}$ is also symmetric, which implies that its eigenvectors are already orthogonal and the "long run" matrix simplifies to $\mathbf{W}^\infty = (1/n)\,\mathbf{J}$. [4]

2.4.2. Case 2: non-symmetric $\mathbf{W}_0$

Consider that $\mathbf{W}_0$ is non-symmetric. In this case, the previous result for $\mathbf{W}^\infty$ is not valid. To see this write $\mathbf{W}_{sim}$ as a function of $\mathbf{W}$ and consider the eigendecomposition for both matrices:

$$\mathbf{W}_{sim} = \mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} \Leftrightarrow \mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{-1} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \tag{40}$$

with $\mathbf{V}_{sim}^{-1} \neq \mathbf{V}_{sim}^\intercal$ because the eigenvectors of $\mathbf{W}_{sim}$ are no longer orthogonal. Therefore, to approximate the spatial lag operator inverse without additional computational burden, it is crucial to obtain an expression for $\mathbf{V}_{sim}^{-1}$ based on a symmetric matrix.

Let $\mathbf{W}_0^*$ be the "symmetrized" variant of $\mathbf{W}_0$, such that if unit j is a neighbor of unit i, then unit i is also a neighbor of unit j with equal weight, for all $i, j =$

[4] Note that if (38) is multiplied by a matrix or vector with zero mean, the proposed approximation gives the same result as the linear transformation $\mathbf{I} + \alpha\mathbf{W}$.

$1, 2, \ldots, N$ and $i \neq j$. This follows as:

$$\mathbf{W}_0^* = \mathbf{W}_0 - \frac{1}{2}\left\{\mathbf{W}_0 - \mathbf{W}_0^\intercal - \left[(\mathbf{W}_0 - \mathbf{W}_0^\intercal)^{\circ 2}\right]^{\circ \frac{1}{2}}\right\} = \mathbf{W}_0 + \mathbf{A} \tag{41}$$

where $\mathbf{A}$ is the $N \times N$ "symmetrization" matrix. The operators "$\circ 2$" and "$\circ \frac{1}{2}$" are element-wise operations and correspond to the Hadamard square and to the Hadamard square root, respectively. Also, as in the previous case, a row-normalized matrix and a similar matrix can be defined, based on $\mathbf{W}_0^*$. The row normalized matrix is equal to $\mathbf{W}^* = \mathbf{D}_{R^*}^{-1}\mathbf{W}_0^*$, where $\mathbf{D}_{R^*}$ is a $N \times N$ diagonal matrix whose diagonal elements are the row sums of $\mathbf{W}_0^*$, and the similar matrix is $\mathbf{W}_{sim}^* = \mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0^*\mathbf{D}_{R^*}^{-1/2}$. Note that, here, the previous assumptions (see Section 2.3.1) are also valid for $\mathbf{W}_0^*$, $\mathbf{W}^*$ and $\mathbf{W}_{sim}^*$.

For $\mathbf{A}$ close to the null matrix, $\mathbf{0}$, the matrix $\mathbf{W}_0$ is well approximated by $\mathbf{W}_0^*$. Then the eigenvectors of $\mathbf{W}_{sim}$ can be approximated by the orthogonal eigenvectors of $\mathbf{W}_{sim}^*$. To see this, write $\mathbf{W}_{sim}^*$ as a function of $\mathbf{W}_{sim}$ and consider the eigendecomposition for both matrices:

$$\begin{aligned}
\mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0^*\mathbf{D}_{R^*}^{-1/2} &\approx \mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \mathbf{W}_{sim}^* \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \mathbf{W}_{sim}^* \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{W}_{sim}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
\Leftrightarrow \mathbf{V}_{sim}^*\mathbf{\Lambda}^*(\mathbf{V}_{sim}^*)^\intercal &\approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}
\end{aligned} \tag{42}$$

where the $N \times N$ diagonal matrix $\mathbf{\Lambda}^*$ contains the corresponding eigenvalues of $\mathbf{W}_{sim}^*$ and the $N \times N$ matrix $\mathbf{V}_{sim}^*$ contains, on each column, the ith eigenvector associated with the ith eigenvalue of $\mathbf{W}_{sim}^*$. Note that $\mathbf{\Lambda}^*$ and $\mathbf{\Lambda}$ are not equal, but because $\mathbf{W}_{sim}^*$ and $\mathbf{W}_{sim}$ are similar to the corresponding row-normalized matrices, $\lim_{h\to\infty}(\mathbf{\Lambda}^*)^h = \lim_{h\to\infty}\mathbf{\Lambda}^h$. Also, the equation above implies that the eigenvectors of $\mathbf{W}_{sim}^*$ and the eigenvectors of $\mathbf{W}_{sim}$ are approximately related as:

$$\begin{aligned}
\mathbf{V}_{sim}^* \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{V}_{sim} \text{ and } (\mathbf{V}_{sim}^*)^\intercal \approx \mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}, \\
\text{but also } (\mathbf{V}_{sim}^*)^\intercal \approx \mathbf{V}_{sim}^\intercal\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}
\end{aligned} \tag{43}$$

where $\mathbf{V}_{sim}^{-1}$ can be straightforwardly approximated by $(\mathbf{V}_{sim}^{*})^{\intercal}\,\mathbf{D}_{R^*}^{1/2}\mathbf{D}_{R}^{-1/2}$. Analogously to the results in (32) and (33), the eigenvectors of $\mathbf{W}_{sim}^{*}$ are related to the eigenvectors of $\mathbf{W}^{*}$ as $\mathbf{V}_{sim}^{*} = \mathbf{D}_{R^*}^{1/2}\mathbf{V}^{*}$. This implies that:

$$\mathbf{V}_{sim}^{-1} \approx (\mathbf{V}^{*})^{\intercal}\,\mathbf{D}_{R^*}^{1/2}\mathbf{D}_{R^*}^{1/2}\mathbf{D}_{R}^{-1/2} \tag{44}$$

Consider the hth power of (40):

$$\begin{aligned} \mathbf{V}_{sim}\boldsymbol{\Lambda}^{h}\mathbf{V}_{sim}^{-1} &= \mathbf{D}_{R}^{1/2}\mathbf{V}\boldsymbol{\Lambda}^{h}\mathbf{V}^{-1}\mathbf{D}_{R}^{-1/2} \Leftrightarrow \\ \Leftrightarrow \mathbf{D}_{R}^{-1/2}\mathbf{V}_{sim}\boldsymbol{\Lambda}^{h}\mathbf{V}_{sim}^{-1}\mathbf{D}_{R}^{1/2} &= \mathbf{V}\boldsymbol{\Lambda}^{h}\mathbf{V}^{-1} \end{aligned} \tag{45}$$

Using the eigenvector relationships in (33) and (43) yields:

$$\mathbf{D}_{R}^{-1/2}\left(\mathbf{D}_{R}^{1/2}\mathbf{V}\right)\boldsymbol{\Lambda}^{h}\left\{\left[(\mathbf{V}^{*})^{\intercal}\,\mathbf{D}_{R^*}^{1/2}\right]\mathbf{D}_{R^*}^{1/2}\mathbf{D}_{R}^{-1/2}\right\}\mathbf{D}_{R}^{1/2} = \mathbf{V}\boldsymbol{\Lambda}^{h}\mathbf{V}^{-1} \tag{46}$$

Note that $\mathbf{V}_{sim} = \mathbf{D}_{R}^{1/2}\mathbf{V}$ because no approximation is required, then the result in (33) holds. Now, as $h \to \infty$:

$$\begin{aligned} \mathbf{W}^{\infty} &\approx \frac{1}{\left\|\mathbf{D}_{R}^{1/2}\mathrm{col}\,(\mathbf{V})_1\right\|_2\left\|\mathrm{col}\,(\mathbf{V}^{*})_1^{\intercal}\,\mathbf{D}_{R^*}^{1/2}\right\|_2} \times \mathbf{D}_{R}^{-1/2}\left(\mathbf{D}_{R}^{1/2}\mathrm{col}\,(\mathbf{V})_1\,\mathrm{col}\,(\mathbf{V}^{*})_1^{\intercal}\,\mathbf{D}_{R^*}^{1/2}\right)\mathbf{D}_{R^*}^{1/2} \\ &= \left(\sum_{i=1}^{N} d_{R,i}\right)^{-1/2}\left(\sum_{i=1}^{N} d_{R,i}^{*}\right)^{-1/2} \times \mathbf{J}\mathbf{D}_{R^*} \end{aligned} \tag{47}$$

where $d_{R,i}^{*}$ is the sum of the ith row of $\mathbf{W}_0^{*}$ and $\mathrm{col}\,(\mathbf{V}^{*})_1 = \boldsymbol{\iota}$ because $\mathbf{W}^{*}$ is row-normalized. As before, the "long run" matrix, $\mathbf{W}^{\infty}$, is rescaled due to the orthogonalization of the first eigenvector of $\mathbf{W}$. In this case, the geometric mean between the sum of all rows of $\mathbf{W}_0$ and $\mathbf{W}_0^{*}$ is used. The remaining results are straightforward.

2.5. GMM estimation with approximated gradients

The estimation of model (5) is addressed through a variant of the iterative GMM estimator of Klier and McMillen (2008). The iterative procedure deduced in

Section 2.2 is used and the N-dimensional matrix operations from the individual gradients (10) and (11) are approximated, considering the new method presented in Section 2.4. Under this approach, it is no longer required to compute the inverse of the spatial lag operator and related matrix operations on each iteration. Also, it is possible to deduce approximate closed formulas for the elements of the approximated matrices. In this way, the overall computational complexity and the computational time of the estimation is significantly reduced.

As in Section 2.3.2, consider the matrix $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$. Also, consider the matrices $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Sigma}$, to derive the approximation for their diagonal elements.

Focusing on $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$, consider the corresponding series expansion and replace the powers $h \geq 2$ of $\mathbf{W}$ by $\mathbf{W}^{\infty}$. This yields:

$$\begin{aligned}(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{W}+2\alpha\mathbf{W}^{2}+3\alpha^{2}\mathbf{W}^{3}+\ldots+(q+1)\,\alpha^{q}\mathbf{W}^{q+1}+\ldots \\ &\approx \mathbf{W}+2\alpha\mathbf{W}^{\infty}+3\alpha^{2}\mathbf{W}^{\infty}+\ldots+(q+1)\,\alpha^{q}\mathbf{W}^{\infty}+\ldots \\ &= \mathbf{W}+\left(\frac{1}{(1-\alpha)^{2}}-1\right)\mathbf{W}^{\infty} \end{aligned} \tag{48}$$

In this way, the diagonal of $\boldsymbol{\Upsilon}$ is approximately equal to:

$$\begin{aligned}\operatorname{diag}(\boldsymbol{\Upsilon}) \approx 2\times\operatorname{diag}\Bigg(\sum_{j=1}^{N}\Bigg(&\left[\mathbf{W}+\left(\frac{1}{(1-\alpha)^{2}}-1\right)\mathbf{W}^{\infty}\right] \\ &\circ\left[\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^{2}}{1-\alpha}\mathbf{W}^{\infty}\right]\Bigg)_{ij}\Bigg)\end{aligned} \tag{49}$$

After some algebra (49) simplifies to:

$$\begin{aligned}\Upsilon_{ii} \approx{}& \frac{(2-\alpha)\,2\alpha}{(1-\alpha)^{2}}w_{ii}^{\infty}+2\alpha\sum_{j=1}^{N}w_{ij}^{2} \\ &+\frac{(3-2\alpha)\,2\alpha^{2}}{(1-\alpha)^{2}}\sum_{j=1}^{N}w_{ij}w_{1j}^{\infty}+\frac{(2-\alpha)\,2\alpha^{3}}{(1-\alpha)^{3}}\sum_{j=1}^{N}\left(w_{ij}^{\infty}\right)^{2}, \quad i=1,2,\ldots,N\end{aligned} \tag{50}$$

Note that, because the row vectors of $\mathbf{W}^{\infty}$ are all equal, the Hadamard product between $\mathbf{W}$ and $\mathbf{W}^{\infty}$ is simplified to the element-wise product between a matrix

and a row vector.

Lastly, the diagonal elements of $\boldsymbol{\Sigma}$ are equal to the row sums of the Hadamard square of the spatial lag operator inverse:

$$\operatorname{diag}(\boldsymbol{\Sigma}) \approx 2 \times \operatorname{Var}(\xi_i \,|\, \mathbf{X}, \mathbf{W}) \times \operatorname{diag}\left(\sum_{j=1}^{N}\left(\left[\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right]^{\circ 2}\right)_{ij}\right) \tag{51}$$

or simply:

$$\begin{aligned}\sigma_i^2 \approx \sigma_{\boldsymbol{\xi}}^2 \Bigg[2 + \frac{4\alpha^2}{1-\alpha}\left(w_{ii}^{\infty}\right)^2 + 2\alpha^2 \sum_{j=1}^{N} w_{ij}^2 \\ + \frac{4\alpha^3}{1-\alpha}\sum_{j=1}^{N} w_{ij} w_{1j}^{\infty} + \frac{2\alpha^4}{(1-\alpha)^2}\sum_{j=1}^{N}\left(w_{ij}^{\infty}\right)^2\Bigg], \quad i = 1, 2, \ldots, N\end{aligned} \tag{52}$$

and $\sigma_{\boldsymbol{\xi}}^2 = \operatorname{Var}(\xi_i \,|\, \mathbf{X}, \mathbf{W})$, for all $i = 1, 2, \ldots, N$, as well. In this way, the approximate expression for the non-constant variances and related quantities are obtained with minimal computational requirements. Also, these quantities can now be used and interpreted in a meaningful way.

2.6. Monte Carlo simulations

In this section, a set of Monte Carlo experiments are presented. The explicit approximation method based on known matrices (AMBKM) is compared with the methods presented in Section 2.3 (the Taylor series approximation, the Chebyshev approximation, the eigendecomposition of the spatial weighting matrix, the LU decomposition and the conjugate gradient method) in terms of the accuracy to approximate the inverse of the spatial lag operator, $(\mathbf{I} - \alpha\mathbf{W})^{-1}$, the diagonal elements of the matrix $\boldsymbol{\Upsilon}$ and the matrix-vector product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$. Also, the proposed iterative GMM estimator with approximated gradients (iGMMa) is compared to the estimators of Klier and McMillen (2008) – the iterative GMM estimator (iGMM) and the GMM estimator of the linearized spatial lag model for binary choice outcomes (LGMM) –, in terms of bias, root mean squared errors and computational

time. A variety of simulation designs are considered, with particular interest on the adequacy of the proposed procedures to large samples frameworks.

2.6.1. Simulation design

The binary dependent variable is constructed following the setting of Klier and McMillen (2008). Consider the simplified version of the model (5) with a single explanatory variable. The explanatory variable, $\mathbf{X}$, is randomly drawn, for each unit, from a $\mathcal{U}(-1,1)$ distribution. Under a Probit specification, the probability of success is given by:

$$P_i = \Phi\left(\frac{\beta_0 x_{i1}^{\#}}{\sigma_i} + \frac{\beta_1 x_{i2}^{\#}}{\sigma_i}\right), \qquad i = 1, 2, \ldots, N \tag{53}$$

where $\Phi(\cdot)$ is the standard Normal CDF, $x_{i1}^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\boldsymbol{\iota}$ and $x_{i2}^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$. The scalars σ_i are the square root of the diagonal elements of the matrix $[(\mathbf{I} - \alpha\mathbf{W})^{\intercal}(\mathbf{I} - \alpha\mathbf{W})]^{-1}$. The observed dependent variable, Y_i, is defined as $Y_i = 1$ if $e_i \leq P_i$ and $Y_i = 0$ otherwise, where e_i is randomly drawn, for each unit, from a $\mathcal{U}(0,1)$ distribution.

The working spatial weighting matrix, $\mathbf{W}$, is constructed according to a two stage setting. In the first stage, the N spatial units are randomly drawn points in the unit square. In the second stage, based on a distance criteria (radial distance or nearest neighbor), an initial spatial weighting matrix, $\mathbf{W}_0$, is constructed and row normalized afterwards. For the case where $\mathbf{W}_0$ is based on the radial distance criterion, the maximum distance to the closest neighbor is computed and a multiplicative factor, δ_R, is used to determine the maximum distance such that the unit j is considered to be a neighbor of unit i, for all $i, j = 1, 2, \ldots, N$. For the case where $\mathbf{W}_0$ is based on the nearest neighbor criterion, the number of nearest neighbors is given by $\delta_{NN} \times N$, where δ_{NN} is the matrix density (the complement of sparsity), the proportion of non-zero elements in $\mathbf{W}$. In this way, the large sample properties of the proposed procedures can be addressed according to the spatial statistics definitions of increasing-domain asymptotics and infill asymptotics (Cressie, 2015). The former corresponds to a sampling scenario where new spatial units are added to the

edges of the lattice and the number of neighbors, for each spatial unit, remains fixed, as $N \to \infty$. The latter corresponds to a scenario where new observations are added between the existing ones and a bounded area becomes dense (Anselin, 2007). Also, it is important to note that, under the radial distance criterion, $\mathbf{W}_0$ is symmetric, while under the nearest neighbor criterion, $\mathbf{W}_0$ is non-symmetric. Therefore, simulations are performed to assess the adequacy of the AMBKM when the assumption of symmetry is not valid.

The Monte Carlo experiments are conducted for each design of $\mathbf{W}$ and for each GMM estimator, as well. The number of spatial units, N, vary over the set $\{100, 1000, 2000\}$ and the spatial lag parameter takes on values $\alpha \in \{0, 0.2, 0.5, 0.8\}$. For the case where $\mathbf{W}$ is based on the radial distance criterion, δ_R vary over the restricted set $\{1, 2, 4\}$. For the case where $\mathbf{W}$ is based on the nearest neighbor criterion, δ_{NN} vary over the restricted set $\{0.01, 0.1, 0.2\}$. In this way, the number of neighbors is approximately equal for the two criteria. The regression parameters are held fixed at $\beta_0 = 0$ and $\beta_1 = 1$ and the matrix of instruments used in all estimation procedures is $\mathbf{Z} = [\mathbf{X}\ \mathbf{WX}\ \mathbf{W}^2\mathbf{X}\ \mathbf{W}^3\mathbf{X}]$. For each experiment, 1000 replications are used. The experiments were performed in a Linux based server, with 64 GB of RAM and composed by 24 AMD Opteron CPUs, ranging from 0.8 GHz to 2.1 GHz.

For each set of experiments per approximation method, the accuracy of the approximated spatial lag operator inverse is summarized in terms of the average relative norm difference w.r.t. the identity matrix:

$$\frac{1}{1000}\sum_{r=1}^{1000}\frac{\left\|\left(\mathbf{I}-\alpha\mathbf{W}\right)\left(\mathbf{I}-\alpha\mathbf{W}\right)_{approx}^{-1}-\mathbf{I}\right\|_2}{\left\|\mathbf{I}\right\|_2} \tag{54}$$

while the accuracy of the approximated diagonal elements of the matrix $\boldsymbol{\Upsilon}$ is summarized in terms of the average relative norm difference w.r.t. the true values:

$$\frac{1}{1000}\sum_{r=1}^{1000}\frac{\left\|\mathrm{diag}\left(\boldsymbol{\Upsilon}\right)_{approx}-\mathrm{diag}\left(\boldsymbol{\Upsilon}\right)\right\|_2}{\left\|\mathrm{diag}\left(\boldsymbol{\Upsilon}\right)\right\|_2} \tag{55}$$

The accuracy of the approximation of the matrix-vector product $\left(\mathbf{I}-\alpha\mathbf{W}\right)^{-1}\mathbf{X}$ is

summarized by the average correlation coefficient between the approximated and the true values of the resulting vector.

For each set of experiments per GMM estimator, the estimates of the regression parameters, $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\alpha}$ are reported, as well as three computational indicators: time per loop (in seconds), number of iterations and total time (in seconds). The parameter estimates are summarized by both the mean and the root mean squared error (RMSE), while the computational indicators are summarized only by the mean. Also, for the case where $\mathbf{W}$ is based on the radial distance criterion, the number of neighbors is reported and summarized by the mean, while for the case where $\mathbf{W}$ is based on the nearest neighbor criterion, the percentage of asymmetric neighbors is reported and summarized by the mean. The calculations were performed using R and the package McSpatial from McMillen (2013).

2.6.2. Results

The results of the Monte Carlo experiments are presented in Appendix A. The simulation results on the accuracy of the approximation methods are detailed in Appendix A.1 and the simulation results on the statistical and computational properties of the GMM estimators are detailed in Appendix A.2. Also, the simulation results are organized according to the criteria chosen to construct the spatial weighting matrix, $\mathbf{W}$, and according to the true values of α.

The accuracy of the approximations considerably relies on the true values of α. For $\alpha = 0$, the approximations are trivial. However, as α becomes close to unity, their accuracy worsens. In particular, the accuracy of AMBKM rapidly deteriorates for $\alpha \geq 0.5$. This highlights the fact that the weight of the infinite higher-order term that is neglected (or approximated), in the series expansion of the inverse – see (15) and (17) –, becomes larger as $\alpha \to 1$. In this way, the higher-order term is more informative to the approximations at moderate and high levels of spatial dependence.

In addition, there is a slight improvement in the accuracy of the approximations as $\mathbf{W}$ becomes dense (δ_R and δ_{NN} are increasing), for a fixed N. This happens

because the magnitude of each element of the spatial lag operator inverse and related matrices, including the matrix $\boldsymbol{\Upsilon}$, is smaller for a denser $\mathbf{W}$. Also, since $\left\|\alpha^h \mathbf{W}^h\right\| \leq |\alpha|^h \|\mathbf{W}\|^h = |\alpha|^h$, the elements of the hth term of the series expansion of the inverse are bounded by $|\alpha|^h$, for any $\mathbf{W}$ satisfying Assumption 2.5. Therefore, as $\mathbf{W}$ becomes dense, elements with progressively smaller magnitudes are added to the series expansion of the inverse. Hence, for the case where $\mathbf{W}$ is sparse ($\delta_R = 1$ and $\delta_{NN} = 0.01$), the average relative norm differences are w.r.t. large values, while for the case where $\mathbf{W}$ is dense ($\delta_R = 4$ and $\delta_{NN} = 0.2$), the average relative norm differences are w.r.t. small values.

It should be noted that, for the reasons discussed above when $\mathbf{W}$ is sparse, using the approximated matrices in the gradients of the GMM estimation procedure, may reduce the accuracy of the estimates of α, since the term involving Υ_{ii}, in the gradient of α (see equation 11), dominates the expression.

With regard to the approximation of the matrix-vector product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$, the simulations show that the average correlation coefficient between the approximated and the true matrix-vector product is, in general, approximately equal to 1. However, when $\alpha = 0.8$, the average correlation coefficient deteriorates as $\mathbf{W}$ becomes sparse (δ_R and δ_{NN} are decreasing). This is particularly obvious for the Taylor series approximation and the AMBKM, where the minimum average correlation coefficient is equal to 0.987 and 0.887, respectively, corresponding to the case where $N = 100$. For the case where $N \geq 1000$, the minimum average correlation coefficient becomes equal to 0.990 and 0.937, respectively. These results emphasize, once again, the issues related to the accuracy of the approximations under the scenarios where $\mathbf{W}$ is sparse and the degree of spatial dependence is high.

In terms of computational time, for $N = 100$, all the approximation methods are fairly quick. However, as $N \to \infty$, the computational time associated with the eigendecomposition, the LU decomposition and the conjugate gradient method clearly increases, in comparison to the remaining methods, since they involve matrix operations that become computationally burdensome for large N. Considering the eigendecomposition, the full eigensystem and an N-dimensional matrix product have to be computed. For the LU decomposition, $(\mathbf{I} - \alpha\mathbf{W})$ has to be factored. For the

conjugate gradient method, an N-dimensional matrix-vector has be computed on each iteration. Similarly, the computational demand associated with the Taylor series approximation and the Chebyshev approximation tends to increase, as $\mathbf{W}$ becomes dense. This is because the first four powers of $\mathbf{W}$ need to be computed. To the contrary, the computational time of the AMBKM is much less sensitive to the size and density of $\mathbf{W}$, since it involves a simple summation of known matrices.

Despite the simulation results showing that, under very specific scenarios, the AMBKM produces larger average relative norm differences and less correlated approximations w.r.t. the true operation, these effects are mitigated when regarding estimation (see, in particular, Table A2.2.6 to Table A2.2.8). Furthermore, the AMBKM is the approximation method that requires minimal computational time (less than a second) to recover the quantities of interest and allows to approximate the partial effects. For these reasons, the AMBKM is particularly useful when iterative procedures have to be used to estimate spatial binary choice models with large samples and dense spatial weighting matrices.

Now, focusing on the performance of the GMM estimators, the results are, in general, consistent with the previous findings in the literature (see Billé, 2013; Calabrese and Elkink, 2014; Klier and McMillen, 2008). The estimates of the regression parameters, $\hat{\beta}_0$ and $\hat{\beta}_1$, are extremely accurate, except for $\alpha = 0.8$. In that case, they exhibit a small bias (a downward bias for the iGMMa and the LGMM estimators and an upward bias for the iGMM estimator) that tends to vanish as $N \to \infty$ and $\mathbf{W}$ becomes dense.

The estimates of the spatial lag parameter, $\hat{\alpha}$, are far more open to discussion, since its accuracy is simultaneously affected by the true value of the parameter, the sample size and the density of $\mathbf{W}$.

For $\alpha \leq 0.5$ and a fixed N, as $\mathbf{W}$ becomes dense, the iGMM estimator exhibits a significant growing downward bias, whereas the LGMM and iGMMa estimators are much less biased. The only exceptions are for $\alpha = 0.5$ and $\alpha = 0$, where the LGMM and iGMMa estimators exhibit a growing upwards bias, respectively. The decreased accuracy of the LGMM estimator, at moderate and high levels of spatial dependence, is expected, considering the existing simulation studies. The spurious

spatial dependence estimated by the iGMMa estimator, for the case where $\alpha = 0$, and the biases displayed by the iGMM estimator evidence that, in general, the spatial GMM estimators can be severely distorted under infill asymptotics (fixed N, denser $\mathbf{W}$). This is especially obvious when $N = 100$. See Lahiri (1996) for a discussion on this matter.

For $\alpha \leq 0.5$ and a fixed density of $\mathbf{W}$, as $N \to \infty$, the iGMM estimator exhibits a downward bias that tends to decline more rapidly when $\mathbf{W}$ is sparse. The LGMM and the iGMMa estimators typically exhibit an upwards bias. For $0 \leq \alpha \leq 0.2$, this bias tends to vanish more rapidly when $\mathbf{W}$ is sparse, while for $\alpha = 0.5$, it tends to vanish more rapidly when $\mathbf{W}$ is dense. Note that, here, both infill and increasing domain asymptotics appear to operate. This implies that the rate of convergence for the various parameters can be different and possibly slower than $\sqrt{N}$, as argued by Lee (2004).

The case of $\alpha = 0.8$ is of particular relevance, since all the spatial GMM estimators exhibit a significant upward bias. Recall that, under the GMM framework, consistency relies on the validity of moment conditions, that use only the information in marginal distributions. However, other estimation methods that consider the joint dependence structure of the spatial data in the estimation, typically perform better at high levels of spatial dependence. In fact, this corroborates with the simulation results of Billé (2013) and Calabrese and Elkink (2014).

Nevertheless, there are two important results regarding the accuracy of the iGMMa estimator that should be emphasized. First, for $N \geq 1000$ and as $\mathbf{W}$ becomes dense, α is better estimated when using the iGMMa estimator, especially for the case where $\mathbf{W}$ is based on the nearest neighbor criterion[5]. This suggests that, for the iGMMa estimator, the number of neighbors for each spatial unit can diverge to infinity at a faster rate than that of the iGMM estimator, without compromising consistency (see Lee, 2004). Second, for $\alpha \geq 0.5$, the iGMMa estimator is typically less biased than the other spatial GMM estimators. The only exception is for $\alpha = 0.8$ and $N = 2000$, where α is better estimated when using the iGMM estimator, but

[5]Under the nearest neighbor criterion and using the AMBKM, $\mathbf{W}$ is based on a symmetrized version of an initial spatial weighting matrix, which implies that the number of neighbors for each spatial unit necessarily increases.

β_0 and β_1 are better estimated when using the iGMMa estimator, especially if $\mathbf{W}$ is not sparse ($\delta_R > 1$ or $\delta_{NN} > 0.01$).

With regard to the RMSEs of the estimated parameters, the simulation results show that the RMSEs of $\hat{\beta}_0$ and $\hat{\beta}_1$ are substantially smaller than the RMSEs of $\hat{\alpha}$.

For a fixed N, as $\mathbf{W}$ becomes dense, all the RMSEs increase. In particular, when $N = 100$, the RMSEs of $\hat{\alpha}$ largely increase. These facts evidence, once again, how the estimates can be severely distorted under infill asymptotics.

For a fixed density of $\mathbf{W}$, as $N \to \infty$, the RMSEs of $\hat{\beta}_0$ and $\hat{\beta}_1$ decrease, whereas the RMSEs of $\hat{\alpha}$ exhibit a fairly different behavior considering the criteria chosen for $\mathbf{W}$. For the case where $\mathbf{W}$ is based on the nearest neighbor criterion, the RMSEs of $\hat{\alpha}$ increase. For the case where $\mathbf{W}$ is based on the radial distance criterion, the RMSEs of $\hat{\alpha}$ decrease.

Additionally note that the RMSEs of the iGMMa estimator are typically smaller than the RMSEs of the remaining spatial GMM estimators. In particular, the RMSEs of $\hat{\alpha}$ in the iGMMa estimator are substantially smaller than in both the iGMM and LGMM estimators, even for the case where $\alpha = 0.8$.

In terms of the computational ability associated with the spatial GMM estimators, measured by the average computational time required to produce estimates for the parameters of interest, it strongly relies on the sample size and on the density of $\mathbf{W}$. As $N \to \infty$ and $\mathbf{W}$ becomes dense, the average computational time increases. In particular, for $\alpha = 0.8$, the average computational time is even larger, since the spatial GMM estimators require, on average, 1 to 2 additional iterations to converge. This is because the inverse of the spatial lag operator is approaching singularity and the computation of the gradients becomes troublesome.

When $N = 2000$ and $\mathbf{W}$ is dense, the average computational time of the iGMMa estimator is about 3 to 6 times less than that of the iGMM estimator, depending whether $\mathbf{W}$ is based on the radial distance criterion or on the nearest neighbor criterion, respectively. Also, the iGMMa estimator is typically less biased than the iGMM estimator, especially for $\alpha \leq 0.5$.

The average computational time of the LGMM estimator is clearly impossible to overcome. However, the iGMMa estimator proves its ability to estimate β_0, β_1

and α with more accuracy, more precision and in a reasonable amount of time, even when the true value of α is close to unity.

2.7. Empirical application

In this section, an empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas (U.S. MSAs) is presented. The adequacy of the previous GMM estimators to real data is assessed and compared.

The strategies to promote and/or to improve competitiveness at the regional and country level are currently centering the attention of policy makers. However, the definition of competitiveness is far from being consensual. In the words of Porter (1990), competitiveness is more than bilateral comparisons, it is related to the ability of the industries to innovate. Fagerberg (1988) defines competitiveness as the growth in relative unit labor costs (the cost of labour per units of output) and, eight years later, considers that competitiveness can be addressed by the growth of GDP per capita or the change in research and development as a percentage of GDP (Fagerberg, 1996). More recently, in a report from the World Economic Forum, Schwab and Sala-i-Martin (2010) defined 12 pillars for competitiveness, based on institutional background, physical infrastructures, macroeconomic environment, efficiency and innovation. Then, in a broad sense, competitiveness is considered a measure for economic performance. Moreover, while promoting competitiveness, the possible environmental impacts cannot be disregarded.

The relationship between environmental degradation and economic growth has been extensively studied in the literature and hypothesized as an "U"-shaped relationship, the so-called Environmental Kuznets Curve (EKC) hypothesis (Grossman and Krueger, 1991; Panayotou, 1993; Shafik and Bandyopadhyay, 1992). However, the EKC hypothesis is not free of criticism, mainly due to the shape of the relationship and the lack of empirical evidence. Also, as Porter et al. (2015) points out, the promotion of efficient energy infrastructures and a low-carbon transition may help to improve competitiveness. In fact, this consists in an inversion of the EKC hypothesis, yet to be tested empirically. Most of the applied works focus on

the analysis of competitiveness and environmental quality as separate subjects and only a few consider the analysis under a spatial framework – Rice et al. (2006) and Dudensing and Barkley (2010) on the spatial spillovers of regional competitiveness and Millimet et al. (2003) and Rupasingha et al. (2004) on the shape of the EKC and on the spatial spillovers associated with the emission of air pollutants. Furthermore, none of the previous works estimate a spatial model with binary dependent variables.

Here, the analysis of the environmental effects over the competitiveness in the U.S. MSAs is addressed. A combined dataset of socioeconomic data and environmental data from the U.S. Bureau of Economic Analysis (BEA) and the U.S. Environmental Protection Agency (EPA), respectively, is used. This dataset contains information about the GDP, labor costs, price indexes, dividends, total employment and population, as well as, information about the annualized Air Quality Index (AQI) and for five main air pollutants – ground-level ozone (O_3), particle pollutants ($PM_{2.5}$ and PM_{10}), carbon monoxide (CO), sulfur dioxide (SO_2) and nitrogen dioxide (NO_2). The U.S. MSAs that are included in this analysis correspond to the continental MSAs that continuously report information for the previous variables, between 2001 and 2016 ($N = 4,848$).

As previously mentioned, there are numerous ways to define competitiveness. Because it is difficult to provide a clear interpretation or to have precise units of measurement, competitiveness can be considered a latent variable. Therefore, the many existing proxies to measure competitiveness can be used to define a new indicator. In this way, a Binary Competitiveness Indicator (BCI) is proposed. A given Metropolitan Statistical Area (MSA) is defined as competitive if, simultaneously, (1) its employment-to-population ratio is greater than the employment-to-population ratio in the combined area of the excluded MSAs and the non-MSAs; (2) its GDP per capita is greater than the GDP per capita in the combined area of the excluded MSAs and the non-MSAs; (3) its Unit Labor Costs (the cost of labor per unit of output) are less than the Unit Labor Costs in the combined area of the excluded MSAs and the non-MSAs or the Unit Capital Costs (the cost of capital per unit of output) are less than the Unit Labor Costs in the combined area of the

excluded MSAs and the non-MSAs, depending on whether the labor intensity ratio (the cost of labor to the cost of capital) is greater than or less than 1, respectively.

In Table B2.1 the descriptive statistics for the variables included in this study are presented. Considering the BCI, about 15% of the U.S. MSAs are labeled as competitive. The variables AQI_{min} and AQI_{max} are, respectively, the minimum and maximum annual values for the AQI, and, as expected, AQI_{min} exhibits a low variability pattern, contrarily to AQI_{max}, that is influenced by the existence of severe outliers. The variables % days O_3, % days $PM_{2.5}$, % days PM_{10}, % days CO, % days SO_2 and % days NO_2 correspond to the percentage of days that the observed value of the AQI was determined by the concentration levels of each pollutant. On average, O_3 and $PM_{2.5}$, by a large amount, the most important contributors for the observed values of the AQI, in this sample. The variables % days Above Moderate and % days Exceptional Events correspond, respectively, to the percentage of days that the observed value of the AQI was above 0.51 and to the percentage of days that the observed value of the AQI was affected by "exceptional events" (wildfires or other natural disasters).

A spatial lag Probit is applied to the pooled sample of the U.S. MSAs to study the effects of the environmental quality indicators over the spatially lagged BCI. The spatial weighting matrix, $\mathbf{W}$, is block-diagonal and based on the radial distance criterion with a distance threshold equal to 1, according to the pattern of proximity displayed in Figure B2.1. Under this specification for $\mathbf{W}$, the spurious spatial interactions are controlled, because only the closest U.S. MSAs are considered to be neighbors. Two models are estimated: the unrestricted model and the restricted model. The first model considers a general specification, based on the available information on the pollutants and air quality and assuming that there is a quadratic relationship between the AQI and the BCI. The second model is a restricted version of the first model, focusing on statistically significant effects.

In Table B2.2 the estimation results for the previous models are presented, considering the three GMM estimators (iGMMa, iGMM and LGMM). The instruments $\mathbf{Z} = [\mathbf{X}\ \mathbf{WX}\ \mathbf{W}^2\mathbf{X}\ \mathbf{W}^3\mathbf{X}]$ were used in all estimation procedures. Also, time effects were added and the Mundlak (1978) approach was used to filter the eventual

dependence between the unobserved regional effects and the explanatory variables. The estimation routines[6] are based on the R package McSpatial of McMillen (2013).

In general, the estimates for the unrestricted model are quite poor in terms of statistical significance, except for the linear and quadratic effects of AQI_{min} and for the estimate for the spatial lag parameter. In fact, the estimated signs for the coefficients of AQI_{min} are of particular interest, due to the statistical evidence towards the existence of an "U"-shaped effect. Also, the estimates for the spatial lag parameter reveal that there may be a high degree of spatial dependence in the sample. However, because the Wald test rejects the null of overall significance, the robustness of the previous results to the exclusion of several statistically insignificant variables should be checked. In fact, the test for exclusion restrictions allowed to considerably simplify the initial specification to a more parsimonious one. In the new specification, only the linear and quadratic AQI_{min}, % days O_3 and % days $PM_{2.5}$ remained. Interestingly, these variables are also used in the applied literature (Millimet et al., 2003; Rupasingha et al., 2004).

The estimates for the restricted model are now individually and jointly statistically significant (except for the variable % days $PM_{2.5}$, in the iGMM estimation). Most importantly, the magnitude of the estimates do not change much, in comparison to the unrestricted model. Therefore, based on the previously noted "U"-shaped effect of AQI_{min}, there is evidence towards the idea that the implementation of environmental-friendly policies may initially involve substantial conversion costs, penalizing regional competitiveness, but, at some point, those costs can be transformed into development opportunities based on new services or products, with large benefits to the economy as a whole. This follows along the lines of Porter et al. (2015) and it is referred as a "win-win path". Nevertheless, some ambiguity may arise concerning the positive estimated signs for the variables % days O_3 and % days $PM_{2.5}$. However, note that, for the case where environmental-friendly policies are implemented and the air quality is actually improved, the observed values for the AQI can still be determined by the concentration levels of the previous pollutants. Recall that O_3 and $PM_{2.5}$ largely contribute to the observed values of the

[6] Available upon request.

AQI. Lastly, having estimates for the spatial lag parameter above 0.7 is evidence towards the importance of the spillover effects over regional competitiveness. This emphasizes the idea that regional policies do benefit the neighboring areas, regarding their economic efficiency.

From the estimation of the previous models, both the iGMMa and iGMM estimators exhibit a quite similar performance, based on Hansen tests and on three measures of goodness-of-fit: the McFadden R^2, the squared correlation coefficient between the observed and the predicted values – $\rho^2(Y, \hat{Y})$ – and the percentage of the correctly predicted observations – $\%(\hat{Y} = Y)$. The adequacy of the moment conditions is not rejected and the predictive power is quite noticeable. This contrasts with the performance of the LGMM estimator, where the Hansen tests reject the null of correct moment conditions and the McFadden R^2 is persistently negative, displaying a very poor fit to the data. In terms of computational time, the iGMMa estimator clearly outruns the iGMM estimator. In this way, the iGMMa estimator proves to be a feasible and an adequate alternative to estimate spatial binary choice models using real data.

2.8. Conclusions

In this paper a new approximation method based on known matrices (AMBKM) is proposed. It addresses the computational issues related to the GMM estimation of spatially lagged models for binary dependent variables. Focusing on the inversion of the spatial lag operator, a simple and intuitive approximation is deduced and applied to approximate other related N-dimensional matrix operations. It is demonstrated that, these matrices are approximated by known matrices and simple matrix-vector operations. Furthermore, it is demonstrated that closed formulas for the elements of the approximated matrices can be easily deduced.

The proposed AMBKM is based on a set of non-restrictive assumptions that allow to accommodate several frameworks for the spatial weighting matrix. This

method is computationally feasible in large samples, because the resulting approximations are based on known matrices, up to an estimated parameter. This is important to note, since it avoids the N-dimensional matrix operations required in the alternative approximation methods, which turns them computationally infeasible in large samples. Moreover, it allows to obtain a closed formula to approximate the partial effects, that can be decomposed into three separate effects (regardless a scale factor): the pure direct effects (from $\mathbf{I}$), the first order neighboring effects (from $\mathbf{W}$) and the "global" effects (from $\mathbf{W}^{\infty}$), which combines the "long run" direct and indirect effects.

This paper also proposes a new GMM estimator based on a modification of the iterative GMM estimator of Klier and McMillen (2008). Aiming at the reduction of the overall computational complexity and the computational time, the approximated matrices are used in the gradients of the new estimation procedure.

Simulations show that the proposed approximation method yields reasonably accurate approximations for the spatial lag operator inverse and related matrices, especially when the spatial weighting matrix is large and dense. Also, the computational time required to obtain these approximations is minimal, regardless the computational complexity of the true operation and the dimension of the spatial weighting matrix.

In addition, the Monte Carlo experiments show that the proposed estimator – the iterative GMM with approximated gradients (iGMMa) –, performs reasonably well in terms of the estimation of the parameters, except for the case where α is close to unity. Nevertheless, for $\alpha \leq 0.5$, the existing biases are attenuated as the spatial weighting matrix becomes large and dense. Also, the iGMMa estimator proved to be surprisingly accurate, for the case where the spatial weighting matrix was based on the nearest neighbor criterion, with a moderate to large number of neighbors. Furthermore, the iGMMa estimator outperformed the benchmark iterative GMM (iGMM) estimator in terms of computational time, accuracy and precision, and outperformed the GMM estimator of the linearized spatial lag model for binary choice outcomes (LGMM) in terms of accuracy and precision. In fact, the iGMMa estimator stood as most precise estimator, even for the case where α is close to unity.

The usefulness of the proposed iGMMa estimator is illustrated in an empirical application that measures the impact of environmental indicators over the competitiveness of the U.S. Metropolitan Statistical Areas. A new Binary Competitiveness Indicator (BCI) is introduced and a spatial lag Probit is estimated, addressing the level of spatial dependence in regional competitiveness. The iGMMa estimator proved to perform as well as the benchmark iGMM estimator, in terms of predictive power, and outperformed the LGMM estimator. Moreover, in this example, where a large data set is used and several explanatory variables are included in the estimation, the iGMMa estimator proved to be computationally superior to the other spatial GMM estimators.

The performance and attractiveness of the proposed iGMMa estimator in estimating models with spatially lagged binary dependent variables lead to obvious extensions, especially the estimation of models with spatially lagged errors and with higher order spatial lag models. The estimation of spatial models for other discrete and censored dependent variables can be addressed by GMM, using the approximated matrices, as well.

All the algorithms used in this paper, the proposed approximation method and the estimation procedures, can be easily implemented using the R package McSpatial from McMillen (2013).

APPENDIXES

A2. Simulation results

A2.1. Approximation methods for the spatial lag operator inverse

TABLE A2.1.1.1: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0$ and $\mathbf{W}$ based on the radial distance criterion.

	δ_R	1			2			4		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.059]	[0.003]	[0.001]	[0.013]	[0.002]	≈[0.000]	[0.016]	[0.004]	≈[0.000]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.063]	[0.056]	≈[0.000]	[0.036]	[0.008]	≈[0.000]	[0.038]	[0.008]	≈[0.000]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.072]	[0.088]	≈[0.000]	[0.029]	[0.033]	≈[0.000]	[0.040]	[0.044]	≈[0.000]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.106]	[0.085]	≈[0.000]	[0.053]	[0.060]	≈[0.000]	[0.068]	[0.075]	≈[0.000]
	Eigen	0.000	–	1.000	0.000	–	1.000	0.000	–	1.000
		[0.010]	[0.005]	≈[0.000]	[0.009]	[0.003]	≈[0.000]	[0.011]	[0.003]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.008]			[0.006]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.073]			[0.009]			[0.010]
1000	True	[0.023]	[0.004]	[0.001]	[0.077]	[0.008]	[0.001]	[0.848]	[0.022]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.309]	[0.151]	[0.001]	[0.216]	[0.156]	[0.001]	[0.412]	[0.165]	[0.002]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.068]	[0.133]	[0.001]	[0.309]	[0.537]	[0.003]	[1.738]	[2.889]	[0.007]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.184]	[0.152]	[0.001]	[0.402]	[0.553]	[0.003]	[2.033]	[2.291]	[0.007]
	Eigen	0.000	–	1.000	0.000	–	1.000	0.000	–	1.000
		[2.189]	[1.102]	[0.005]	[2.241]	[1.100]	[0.005]	[2.164]	[1.099]	[0.005]
	LU			1.000			1.000			1.000
				[0.264]			[0.261]			[0.329]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.428]			[0.428]			[0.429]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.1.2: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0$ and $\mathbf{W}$ based on the radial distance criterion (cont.)

δ_R		1			2			4		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
2000	True	[0.043]	[0.007]	[0.001]	[0.306]	[0.017]	[0.001]	[3.894]	[0.052]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.675]	[0.606]	[0.001]	[0.796]	[0.620]	[0.001]	[0.943]	[0.615]	[0.003]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.126]	[0.270]	[0.002]	[0.765]	[1.828]	[0.007]	[4.505]	[8.716]	[0.020]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.225]	[0.285]	[0.002]	[1.163]	[1.504]	[0.008]	[5.646]	[6.497]	[0.020]
	Eigen	0.000	–	1.000	0.000	–	1.000	0.000	–	1.000
		[15.669]	[8.442]	[0.018]	[15.534]	[8.442]	[0.018]	[15.797]	[8.430]	[0.018]
	LU			1.000			1.000			1.000
				[1.712]			[1.721]			[1.713]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.048]			[3.048]			[3.048]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.2: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.2$ and $\mathbf{W}$ based on the radial distance criterion.

δ_R		1			2			4		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.014]	[0.013]	[0.001]	[0.018]	[0.019]	[0.001]	[0.078]	[0.020]	[0.001]
	AMBKM	0.042	0.665	≈1.000	0.028	0.501	≈1.000	0.010	0.236	≈1.000
		[0.026]	[0.010]	[0.001]	[0.027]	[0.009]	≈[0.000]	[0.031]	[0.008]	≈[0.000]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000
		[0.021]	[0.020]	≈[0.000]	[0.032]	[0.033]	≈[0.000]	[0.039]	[0.046]	≈[0.000]
	Cheb4	≈0.000	0.001	≈1.000	≈0.000	0.003	≈1.000	≈0.000	0.002	≈1.000
		[0.045]	[0.047]	≈[0.000]	[0.115]	[0.064]	≈[0.000]	[0.133]	[0.076]	≈[0.000]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[0.011]	[0.003]	≈[0.000]	[0.010]	[0.003]	≈[0.000]	[0.010]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.004]			[0.005]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.009]			[0.009]			[0.010]
1000	True	[0.267]	[14.103]	[0.008]	[1.003]	[14.350]	[0.008]	[3.221]	[14.903]	[0.008]
	AMBKM	0.041	0.706	≈1.000	0.039	0.703	≈1.000	0.036	0.647	≈1.000
		[0.153]	[0.129]	[0.003]	[0.179]	[0.129]	[0.003]	[0.242]	[0.133]	[0.003]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.052]	[0.093]	[0.001]	[0.425]	[0.779]	[0.004]	[1.541]	[2.663]	[0.007]
	Cheb4	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000	≈0.000	0.003	≈1.000
		[0.152]	[0.118]	[0.001]	[0.671]	[0.686]	[0.004]	[1.957]	[2.125]	[0.007]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[2.226]	[1.098]	[0.005]	[2.226]	[1.152]	[0.005]	[2.150]	[1.097]	[0.005]
	LU			1.000			1.000			1.000
				[0.257]			[0.257]			[0.373]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.443]			[0.445]			[0.440]
2000	True	[1.390]	[114.155]	[0.031]	[4.264]	[115.969]	[0.031]	[21.177]	[118.638]	[0.031]
	AMBKM	0.041	0.712	≈1.000	0.040	0.724	≈1.000	0.038	0.685	≈1.000
		[0.547]	[0.528]	[0.012]	[0.601]	[0.584]	[0.013]	[0.785]	[0.532]	[0.013]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.133]	[0.291]	[0.002]	[0.602]	[1.404]	[0.006]	[4.295]	[8.630]	[0.020]
	Cheb4	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000	≈0.000	0.003	≈1.000
		[0.288]	[0.299]	[0.002]	[0.938]	[1.218]	[0.006]	[5.628]	[6.445]	[0.020]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[17.567]	[8.442]	[0.018]	[15.495]	[8.437]	[0.018]	[15.782]	[8.440]	[0.018]
	LU			1.000			1.000			1.000
				[1.704]			[1.730]			[1.710]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.116]			[3.117]			[3.100]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.3: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.5$ and $\mathbf{W}$ based on the radial distance criterion.

δ_R		1			2			4		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.014]	[0.014]	[0.001]	[0.017]	[0.017]	[0.001]	[0.020]	[0.020]	[0.001]
	AMBKM	0.258	0.786	0.995	0.208	0.630	0.999	0.108	0.260	≈1.000
		[0.026]	[0.009]	≈[0.000]	[0.027]	[0.008]	≈[0.000]	[0.086]	[0.008]	≈[0.000]
	Taylor4	0.038	0.110	≈1.000	0.033	0.089	≈1.000	0.032	0.118	≈1.000
		[0.023]	[0.023]	≈[0.000]	[0.028]	[0.034]	≈[0.000]	[0.037]	[0.041]	≈[0.000]
	Cheb4	0.006	0.018	≈1.000	0.004	0.032	≈1.000	0.003	0.014	≈1.000
		[0.046]	[0.049]	≈[0.000]	[0.054]	[0.059]	≈[0.000]	[0.066]	[0.072]	≈[0.000]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[0.010]	[0.003]	≈[0.000]	[0.011]	[0.060]	≈[0.000]	[0.009]	[0.003]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.005]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.009]			[0.009]			[0.009]
1000	True	[0.283]	[14.053]	[0.008]	[0.897]	[14.280]	[0.008]	[3.143]	[14.939]	[0.008]
	AMBKM	0.257	0.847	0.996	0.247	0.818	0.999	0.228	0.720	≈1.000
		[0.154]	[0.130]	[0.003]	[0.234]	[0.128]	[0.003]	[0.239]	[0.134]	[0.003]
	Taylor4	0.032	0.095	≈1.000	0.032	0.083	≈1.000	0.032	0.087	≈1.000
		[0.056]	[0.103]	[0.001]	[0.285]	[0.606]	[0.003]	[1.330]	[2.575]	[0.007]
	Cheb4	0.006	0.023	≈1.000	0.004	0.035	≈1.000	0.003	0.037	≈1.000
		[0.107]	[0.128]	[0.001]	[0.449]	[0.552]	[0.003]	[1.790]	[2.135]	[0.007]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[2.184]	[1.098]	[0.005]	[2.178]	[1.097]	[0.005]	[2.151]	[1.096]	[0.005]
	LU			1.000			1.000			1.000
				[0.257]			[0.260]			[0.259]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.461]			[0.451]			[0.453]
2000	True	[1.302]	[114.794]	[0.031]	[3.984]	[116.021]	[0.031]	[12.962]	[117.668]	[0.031]
	AMBKM	0.259	0.853	0.996	0.250	0.842	0.999	0.242	0.795	≈1.000
		[0.559]	[0.581]	[0.013]	[0.593]	[0.583]	[0.013]	[0.726]	[0.533]	[0.013]
	Taylor4	0.033	0.092	≈1.000	0.032	0.081	≈1.000	0.032	0.083	≈1.000
		[0.115]	[0.307]	[0.002]	[0.601]	[1.413]	[0.006]	[2.938]	[6.074]	[0.016]
	Cheb4	0.007	0.024	≈1.000	0.004	0.037	≈1.000	0.003	0.038	≈1.000
		[0.204]	[0.259]	[0.002]	[0.929]	[1.217]	[0.006]	[3.965]	[4.699]	[0.016]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[17.265]	[8.441]	[0.018]	[15.486]	[8.438]	[0.018]	[15.746]	[8.436]	[0.018]
	LU			1.000			1.000			1.000
				[1.709]			[1.719]			[1.722]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.191]			[3.153]			[3.154]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.4: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.8$ and $\mathbf{W}$ based on the radial distance criterion.

	δ_R	1			2			4		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.014]	[0.014]	≈[0.000]	[0.016]	[0.017]	[0.001]	[0.020]	[0.020]	[0.001]
	AMBKM	0.663	0.861	0.933	0.550	0.469	0.977	0.206	0.041	0.999
		[0.026]	[0.008]	≈[0.000]	[0.027]	[0.009]	≈[0.000]	[0.030]	[0.008]	≈[0.000]
	Taylor4	0.354	0.696	0.990	0.340	0.636	0.997	0.340	0.747	≈1.000
		[0.021]	[0.020]	≈[0.000]	[0.027]	[0.030]	≈[0.000]	[0.039]	[0.044]	≈[0.000]
	Cheb4	0.128	0.139	0.999	0.094	0.091	0.999	0.069	0.205	≈1.000
		[0.045]	[0.047]	≈[0.000]	[0.054]	[0.113]	≈[0.000]	[0.067]	[0.128]	≈[0.000]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[0.068]	[0.003]	≈[0.000]	[0.009]	[0.005]	≈[0.000]	[0.064]	[0.003]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.004]			[0.005]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.010]			[0.009]			[0.009]
1000	True	[0.290]	[14.017]	[0.008]	[0.790]	[14.391]	[0.008]	[2.992]	[14.871]	[0.008]
	AMBKM	0.659	0.952	0.937	0.636	0.873	0.970	0.579	0.639	0.992
		[0.152]	[0.131]	[0.003]	[0.168]	[0.129]	[0.003]	[0.233]	[0.128]	[0.003]
	Taylor4	0.340	0.645	0.990	0.334	0.602	0.996	0.335	0.613	0.999
		[0.059]	[0.110]	[0.001]	[0.225]	[0.488]	[0.003]	[1.335]	[2.489]	[0.007]
	Cheb4	0.144	0.121	0.999	0.086	0.118	≈1.000	0.081	0.109	≈1.000
		[0.110]	[0.136]	[0.001]	[0.366]	[0.510]	[0.003]	[1.844]	[1.997]	[0.007]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[2.180]	[1.097]	[0.005]	[2.178]	[1.099]	[0.005]	[2.146]	[1.151]	[0.005]
	LU			1.000			1.000			1.000
				[0.258]			[0.255]			[0.260]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.497]			[0.480]			[0.472]
2000	True	[1.188]	[115.265]	[0.031]	[4.058]	[115.905]	[0.030]	[23.791]	[118.997]	[0.031]
	AMBKM	0.661	0.962	0.938	0.641	0.912	0.972	0.605	0.739	0.991
		[0.544]	[0.529]	[0.012]	[0.654]	[0.524]	[0.013]	[0.900]	[0.533]	[0.013]
	Taylor4	0.343	0.633	0.990	0.333	0.593	0.996	0.334	0.602	0.999
		[0.108]	[0.226]	[0.002]	[0.584]	[1.356]	[0.006]	[4.739]	[9.435]	[0.020]
	Cheb4	0.142	0.113	0.999	0.084	0.126	≈1.000	0.077	0.112	≈1.000
		[0.191]	[0.242]	[0.002]	[0.908]	[1.177]	[0.006]	[5.975]	[6.938]	[0.020]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[18.701]	[8.439]	[0.018]	[15.560]	[8.438]	[0.018]	[15.782]	[8.438]	[0.018]
	LU			1.000			1.000			1.000
				[1.713]			[1.711]			[1.725]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.334]			[3.267]			[3.228]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.5: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.060]	[0.071]	[0.001]	[0.012]	[0.002]	≈[0.000]	[0.013]	[0.002]	≈[0.000]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.062]	[0.056]	≈[0.000]	[0.034]	[0.011]	≈[0.000]	[0.034]	[0.010]	≈[0.000]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.071]	[0.026]	≈[0.000]	[0.025]	[0.027]	≈[0.000]	[0.090]	[0.033]	≈[0.000]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.101]	[0.081]	≈[0.000]	[0.048]	[0.052]	≈[0.000]	[0.111]	[0.118]	≈[0.000]
	Eigen	1.000	–	0.971	0.743	–	0.987	0.725	–	0.991
		[0.014]	[0.002]	≈[0.000]	[0.099]	[0.004]	≈[0.000]	[0.043]	[0.005]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.009]			[0.004]			[0.006]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.074]			[0.067]			[0.010]
1000	True	[0.025]	[0.004]	[0.001]	[0.413]	[0.018]	[0.001]	[1.935]	[0.033]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.261]	[0.154]	[0.001]	[0.286]	[0.216]	[0.001]	[0.381]	[0.160]	[0.002]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.068]	[0.133]	[0.001]	[1.142]	[2.176]	[0.006]	[2.958]	[4.869]	[0.008]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.128]	[0.154]	[0.001]	[1.562]	[1.872]	[0.006]	[3.554]	[3.817]	[0.008]
	Eigen	2.175	–	0.995	1.163	–	0.998	1.206	–	0.999
		[22.666]	[2.106]	[0.012]	[22.468]	[2.104]	[0.012]	[22.258]	[2.167]	[0.012]
	LU			1.000			1.000			1.000
				[0.258]			[0.258]			[0.313]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.425]			[0.426]			[0.426]
2000	True	[0.125]	[0.010]	[0.001]	[11.273]	[0.076]	[0.001]	[17.259]	[0.151]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.743]	[0.599]	[0.001]	[1.142]	[0.622]	[0.004]	[1.537]	[0.648]	[0.007]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.334]	[0.653]	[0.004]	[7.081]	[13.673]	[0.025]	[21.942]	[36.299]	[0.031]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.470]	[0.610]	[0.004]	[9.009]	[9.972]	[0.025]	[23.981]	[25.574]	[0.031]
	Eigen	2.793	–	0.995	1.275	–	≈1.000	1.306	–	0.999
		[165.543]	[16.423]	[0.045]	[164.393]	[16.438]	[0.046]	[164.121]	[16.440]	[0.046]
	LU			1.000			1.000			1.000
				[1.700]			[1.726]			[1.732]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.038]			[3.042]			[3.051]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.6: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.2$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.012]	[0.002]	≈[0.000]	[0.015]	[0.017]	[0.001]	[0.017]	[0.017]	[0.001]
	AMBKM	0.070	0.674	≈1.000	0.041	0.612	≈1.000	0.035	0.513	≈1.000
		[0.028]	[0.009]	[0.001]	[0.026]	[0.008]	≈[0.000]	[0.028]	[0.009]	≈[0.000]
	Taylor4	0.001	0.004	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.019]	[0.016]	≈[0.000]	[0.024]	[0.025]	≈[0.000]	[0.027]	[0.088]	≈[0.000]
	Cheb4	≈0.000	0.001	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000
		[0.041]	[0.101]	≈[0.000]	[0.049]	[0.052]	≈[0.000]	[0.055]	[0.059]	≈[0.000]
	Eigen	1.031	0.147	0.961	0.454	0.045	0.995	0.949	0.094	0.991
		[0.011]	[0.002]	≈[0.000]	[0.104]	[0.005]	≈[0.000]	[0.042]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.006]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.009]			[0.009]			[0.009]
1000	True	[0.385]	[14.318]	[0.008]	[2.376]	[14.864]	[0.008]	[5.806]	[15.566]	[0.008]
	AMBKM	0.044	0.714	≈1.000	0.038	0.633	≈1.000	0.032	0.527	≈1.000
		[0.149]	[0.124]	[0.004]	[0.225]	[0.125]	[0.004]	[0.321]	[0.128]	[0.003]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.066]	[0.130]	[0.001]	[1.133]	[2.098]	[0.006]	[3.005]	[4.911]	[0.008]
	Cheb4	≈0.000	0.003	≈1.000	≈0.000	0.004	≈1.000	≈0.000	0.003	≈1.000
		[0.124]	[0.152]	[0.001]	[1.438]	[1.680]	[0.006]	[3.540]	[3.850]	[0.008]
	Eigen	2.111	0.074	0.993	1.389	0.036	0.997	1.169	0.029	0.999
		[22.391]	[2.103]	[0.012]	[22.338]	[2.158]	[0.012]	[22.256]	[2.161]	[0.012]
	LU			1.000			1.000			1.000
				[0.256]			[0.255]			[0.258]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.442]			[0.438]			[0.435]
2000	True	[2.986]	[116.255]	[0.031]	[39.494]	[121.062]	[0.031]	[50.895]	[127.362]	[0.031]
	AMBKM	0.042	0.721	≈1.000	0.037	0.634	≈1.000	0.032	0.526	≈1.000
		[0.567]	[0.512]	[0.013]	[0.905]	[0.528]	[0.012]	[1.322]	[0.581]	[0.012]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.332]	[0.651]	[0.004]	[7.047]	[13.587]	[0.025]	[21.848]	[36.129]	[0.031]
	Cheb4	≈0.000	0.003	≈1.000	≈0.000	0.004	≈1.000	≈0.000	0.004	≈1.000
		[0.465]	[0.604]	[0.004]	[8.925]	[9.940]	[0.025]	[24.132]	[25.575]	[0.031]
	Eigen	2.597	0.046	0.996	1.318	0.020	0.999	1.316	0.021	0.999
		[165.286]	[16.469]	[0.045]	[164.427]	[16.479]	[0.046]	[164.123]	[16.473]	[0.046]
	LU			1.000			1.000			1.000
				[1.695]			[1.781]			[1.708]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.118]			[3.151]			[3.096]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.7: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.5$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.012]	[0.002]	≈[0.000]	[0.015]	[0.017]	[0.001]	[0.016]	[0.017]	[0.001]
	AMBKM	0.498	0.831	0.981	0.252	0.690	0.997	0.212	0.525	0.999
		[0.025]	[0.010]	≈[0.000]	[0.026]	[0.009]	≈[0.000]	[0.028]	[0.008]	≈[0.000]
	Taylor4	0.059	0.154	≈1.000	0.035	0.101	≈1.000	0.034	0.101	≈1.000
		[0.019]	[0.016]	≈[0.000]	[0.022]	[0.025]	≈[0.000]	[0.027]	[0.031]	≈[0.000]
	Cheb4	0.021	0.022	≈1.000	0.005	0.026	≈1.000	0.004	0.030	≈1.000
		[0.041]	[0.043]	≈[0.000]	[0.047]	[0.053]	≈[0.000]	[0.054]	[0.062]	≈[0.000]
	Eigen	1.360	0.144	0.969	0.712	0.015	0.996	0.837	0.014	0.992
		[0.014]	[0.058]	≈[0.000]	[0.098]	[0.004]	≈[0.000]	[0.043]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.005]			[0.005]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.019]			[0.009]			[0.009]
1000	True	[0.374]	[14.235]	[0.008]	[2.472]	[14.851]	[0.008]	[5.844]	[15.611]	[0.008]
	AMBKM	0.273	0.851	0.996	0.236	0.704	≈1.000	0.201	0.541	≈1.000
		[0.150]	[0.127]	[0.004]	[0.228]	[0.184]	[0.003]	[0.321]	[0.181]	[0.004]
	Taylor4	0.036	0.097	≈1.000	0.033	0.090	≈1.000	0.034	0.097	≈1.000
		[0.067]	[0.130]	[0.001]	[1.076]	[2.146]	[0.006]	[3.000]	[4.861]	[0.008]
	Cheb4	0.005	0.028	≈1.000	0.004	0.038	≈1.000	0.004	0.036	≈1.000
		[0.178]	[0.152]	[0.001]	[1.492]	[1.682]	[0.006]	[3.542]	[3.860]	[0.008]
	Eigen	1.977	0.013	0.991	1.115	0.005	0.999	0.885	0.006	0.999
		[22.299]	[2.101]	[0.012]	[22.355]	[2.104]	[0.012]	[22.282]	[2.156]	[0.012]
	LU			1.000			1.000			1.000
				[0.253]			[0.255]			[0.253]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.461]			[0.447]			[0.447]
2000	True	[3.133]	[116.241]	[0.032]	[39.723]	[121.320]	[0.031]	[50.680]	[127.814]	[0.031]
	AMBKM	0.262	0.851	0.998	0.235	0.704	≈1.000	0.202	0.541	≈1.000
		[0.627]	[0.521]	[0.012]	[0.905]	[0.520]	[0.012]	[1.308]	[0.532]	[0.012]
	Taylor4	0.034	0.088	≈1.000	0.033	0.089	≈1.000	0.034	0.096	≈1.000
		[0.278]	[0.659]	[0.004]	[7.169]	[13.616]	[0.025]	[21.948]	[36.176]	[0.031]
	Cheb4	0.005	0.035	≈1.000	0.004	0.039	≈1.000	0.004	0.037	≈1.000
		[0.522]	[0.605]	[0.004]	[8.836]	[9.995]	[0.025]	[23.965]	[25.363]	[0.031]
	Eigen	2.631	0.011	0.994	1.269	0.003	0.999	1.002	0.003	0.999
		[165.227]	[16.518]	[0.045]	[164.365]	[16.474]	[0.045]	[164.075]	[16.475]	[0.046]
	LU			1.000			1.000			1.000
				[1.699]			[1.760]			[1.722]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.168]			[3.265]			[3.136]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

TABLE A2.1.8: Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.8$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag $(\boldsymbol{\Upsilon})$	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.012]	[0.002]	≈[0.000]	[0.015]	[0.017]	[0.001]	[0.016]	[0.017]	[0.001]
	AMBKM	2.551	0.955	0.887	0.873	0.661	0.936	0.748	0.376	0.984
		[0.027]	[0.008]	≈[0.000]	[0.026]	[0.008]	≈[0.000]	[0.027]	[0.008]	≈[0.000]
	Taylor4	0.665	0.781	0.987	0.373	0.677	0.989	0.358	0.675	0.999
		[0.019]	[0.016]	≈[0.000]	[0.022]	[0.026]	≈[0.000]	[0.027]	[0.033]	≈[0.000]
	Cheb4	0.863	0.241	0.999	0.119	0.074	0.999	0.088	0.069	≈1.000
		[0.042]	[0.043]	≈[0.000]	[0.050]	[0.052]	≈[0.000]	[0.054]	[0.060]	≈[0.000]
	Eigen	3.641	0.136	0.970	0.865	0.004	0.995	0.822	0.004	0.992
		[0.012]	[0.002]	≈[0.000]	[0.043]	[0.004]	≈[0.000]	[0.043]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.005]			[0.005]			[0.005]
	CGrad			0.603			≈1.000			≈1.000
				[0.016]			[0.010]			[0.010]
1000	True	[0.375]	[14.218]	[0.008]	[2.381]	[14.871]	[0.008]	[5.800]	[15.617]	[0.008]
	AMBKM	0.735	0.943	0.939	0.612	0.628	0.994	0.680	0.344	0.998
		[0.263]	[0.125]	[0.004]	[0.284]	[0.128]	[0.004]	[0.323]	[0.129]	[0.004]
	Taylor4	0.377	0.651	0.990	0.346	0.631	0.999	0.356	0.659	≈1.000
		[0.066]	[0.128]	[0.001]	[1.188]	[2.096]	[0.006]	[2.963]	[4.872]	[0.008]
	Cheb4	0.129	0.094	0.999	0.088	0.082	≈1.000	0.090	0.064	≈1.000
		[0.180]	[0.150]	[0.001]	[1.437]	[1.686]	[0.006]	[3.604]	[3.818]	[0.008]
	Eigen	1.840	0.003	0.995	0.982	0.001	0.999	1.056	0.001	0.999
		[22.331]	[2.102]	[0.012]	[22.302]	[2.102]	[0.012]	[22.282]	[2.161]	[0.012]
	LU			1.000			1.000			1.000
				[0.252]			[0.256]			[0.256]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.498]			[0.475]			[0.498]
2000	True	[3.090]	[116.212]	[0.031]	[39.872]	[121.396]	[0.031]	[50.982]	[127.827]	[0.031]
	AMBKM	0.671	0.937	0.957	0.605	0.621	0.997	0.682	0.343	0.999
		[0.560]	[0.580]	[0.012]	[0.976]	[0.524]	[0.012]	[1.429]	[0.594]	[0.013]
	Taylor4	0.350	0.616	0.993	0.345	0.627	≈1.000	0.354	0.658	≈1.000
		[0.277]	[0.652]	[0.004]	[7.097]	[13.523]	[0.025]	[21.872]	[36.160]	[0.031]
	Cheb4	0.099	0.110	0.999	0.088	0.084	≈1.000	0.086	0.064	≈1.000
		[0.470]	[0.610]	[0.004]	[8.781]	[9.921]	[0.025]	[24.093]	[25.557]	[0.031]
	Eigen	2.170	0.002	0.998	1.304	≈0.000	0.999	1.128	≈0.000	0.999
		[165.791]	[16.481]	[0.046]	[164.499]	[16.476]	[0.046]	[164.210]	[16.478]	[0.046]
	LU			1.000			1.000			1.000
				[1.716]			[1.721]			[1.705]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.299]			[3.244]			[3.261]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag $(\boldsymbol{\Upsilon})$ are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Averages are based on 1000 replications.

A2.2. GMM estimation

TABLE A2.2.1: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0$ and $\mathbf{W}$ based on the radial distance criterion.

δ_R		1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.074	−0.033	−0.006	0.295	−0.498	−0.020	0.672	−0.894	−0.098
		(0.377)	(0.429)	(0.576)	(0.720)	(1.295)	(1.319)	(2.168)	(3.360)	(5.309)
	$\hat{\beta}_0$	0.005	0.001	0.001	0.010	0.023	0.012	−0.015	0.072	−0.039
		(0.141)	(0.172)	(0.159)	(0.149)	(0.352)	(0.241)	(0.274)	(0.754)	(0.944)
	$\hat{\beta}_1$	1.084	1.097	1.024	1.094	1.145	1.019	1.092	1.143	1.020
		(0.263)	(0.268)	(0.256)	(0.272)	(0.367)	(0.250)	(0.247)	(0.316)	(0.267)
	Time:									
	Loop	0.062	0.034		0.066	0.035		0.069	0.042	
	# Iterations	4	4		5	5		5	5	
	Total	0.305	0.201	0.068	0.371	0.233	0.064	0.418	0.288	0.076
	# Neighbors	6	6	6	21	21	21	60	60	60
1000	$\hat{\alpha}$	0.005	−0.014	0.001	0.056	−0.057	−0.005	0.188	−0.336	−0.055
		(0.201)	(0.202)	(0.203)	(0.351)	(0.435)	(0.418)	(0.572)	(1.154)	(0.906)
	$\hat{\beta}_0$	0.001	0.001	0.001	0.003	0.001	0.001	0.000	0.000	0.001
		(0.044)	(0.046)	(0.045)	(0.044)	(0.049)	(0.047)	(0.041)	(0.073)	(0.057)
	$\hat{\beta}_1$	1.002	1.005	1.002	1.003	1.006	0.999	1.012	1.030	1.005
		(0.074)	(0.075)	(0.074)	(0.078)	(0.078)	(0.078)	(0.081)	(0.142)	(0.076)
	Time:									
	Loop	1.867	0.504		2.115	1.091		1.953	3.989	
	# Iterations	3	4		4	4		5	5	
	Total	6.819	2.517	0.824	9.517	5.697	0.881	10.572	21.991	1.447
	# Neighbors	9	9	9	33	33	33	118	118	118
2000	$\hat{\alpha}$	0.000	−0.010	0.000	0.017	−0.044	−0.007	0.129	−0.129	−0.015
		(0.148)	(0.147)	(0.148)	(0.285)	(0.326)	(0.306)	(0.462)	(0.698)	(0.601)
	$\hat{\beta}_0$	0.000	0.000	0.000	−0.002	−0.001	−0.001	0.000	−0.002	0.000
		(0.030)	(0.030)	(0.030)	(0.031)	(0.034)	(0.033)	(0.030)	(0.040)	(0.035)
	$\hat{\beta}_1$	1.003	1.005	1.003	1.003	1.005	1.002	1.001	1.007	0.999
		(0.054)	(0.054)	(0.054)	(0.053)	(0.054)	(0.054)	(0.056)	(0.083)	(0.055)
	Time:									
	Loop	6.935	2.651		7.304	7.372		6.839	20.889	
	# Iterations	3	3		4	4		4	5	
	Total	23.064	11.109	2.464	30.193	33.943	2.489	34.917	107.516	4.050
	# Neighbors	9	9	9	37	37	37	134	134	134

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

TABLE A2.2.2: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.2$ and $\mathbf{W}$ based on the radial distance criterion.

	δ_R	1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.189	0.132	0.228	0.279	−0.147	0.221	0.827	−0.099	0.143
		(0.388)	(0.389)	(0.589)	(0.655)	(1.184)	(1.370)	(1.866)	(3.875)	(5.414)
	$\hat{\beta}_0$	0.006	0.009	0.004	−0.007	−0.004	−0.004	0.011	−0.064	0.020
		(0.132)	(0.142)	(0.148)	(0.141)	(0.246)	(0.224)	(0.254)	(0.689)	(0.778)
	$\hat{\beta}_1$	1.083	1.109	1.022	1.105	1.104	1.017	1.135	1.144	1.020
		(0.275)	(0.278)	(0.254)	(0.272)	(0.349)	(0.259)	(0.301)	(0.343)	(0.263)
	Time:									
	Loop	0.063	0.033		0.065	0.036		0.070	0.041	
	# Iterations	4	5		5	5		5	6	
	Total	0.319	0.198	0.063	0.361	0.230	0.067	0.439	0.283	0.077
	# Neighbors	6	6	6	20	20	20	58	58	58
1000	$\hat{\alpha}$	0.225	0.186	0.223	0.238	0.130	0.237	0.284	−0.028	0.218
		(0.205)	(0.165)	(0.202)	(0.380)	(0.355)	(0.428)	(0.518)	(1.070)	(0.889)
	$\hat{\beta}_0$	−0.001	−0.001	−0.001	0.000	0.001	0.000	−0.003	−0.006	−0.004
		(0.033)	(0.035)	(0.033)	(0.036)	(0.039)	(0.037)	(0.036)	(0.072)	(0.053)
	$\hat{\beta}_1$	1.001	1.007	0.999	1.011	1.010	1.004	1.011	1.033	1.001
		(0.077)	(0.077)	(0.077)	(0.084)	(0.080)	(0.080)	(0.083)	(0.139)	(0.078)
	Time:									
	Loop	1.489	0.441		1.596	0.933		1.670	3.549	
	# Iterations	3	4		4	5		5	5	
	Total	5.708	2.185	0.629	7.547	4.965	0.684	9.273	19.786	1.311
	# Neighbors	9	9	9	33	33	33	119	119	119
2000	$\hat{\alpha}$	0.222	0.189	0.220	0.234	0.166	0.232	0.270	0.064	0.243
		(0.156)	(0.125)	(0.153)	(0.292)	(0.251)	(0.300)	(0.463)	(0.610)	(0.628)
	$\hat{\beta}_0$	0.000	0.000	0.000	−0.001	−0.001	−0.001	0.000	−0.001	0.000
		(0.023)	(0.024)	(0.023)	(0.026)	(0.027)	(0.026)	(0.027)	(0.037)	(0.031)
	$\hat{\beta}_1$	0.997	1.002	0.996	1.001	1.003	1.001	1.001	1.021	0.998
		(0.054)	(0.054)	(0.054)	(0.056)	(0.056)	(0.055)	(0.053)	(0.156)	(0.054)
	Time:									
	Loop	6.381	2.596		5.949	6.381		6.356	21.036	
	# Iterations	3	4		4	4		5	5	
	Total	22.786	11.184	2.015	25.671	29.407	2.021	33.666	108.958	3.964
	# Neighbors	9	9	9	37	37	37	133	133	133

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

TABLE A2.2.3: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.5$ and $\mathbf{W}$ based on the radial distance criterion.

	δ_R	1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.409	0.328	0.684	0.496	−0.110	0.627	0.837	−0.620	0.739
		(0.451)	(0.389)	(0.686)	(0.733)	(1.330)	(1.528)	(1.773)	(4.238)	(4.292)
	$\hat{\beta}_0$	−0.002	0.002	0.001	0.007	−0.010	0.005	−0.030	−0.031	−0.049
		(0.110)	(0.118)	(0.119)	(0.160)	(0.509)	(0.282)	(0.300)	(1.054)	(0.738)
	$\hat{\beta}_1$	1.051	1.082	0.947	1.121	1.109	1.003	1.066	1.139	1.007
		(0.265)	(0.268)	(0.250)	(0.318)	(0.326)	(0.259)	(0.268)	(0.371)	(0.264)
	Time:									
	Loop	0.063	0.033		0.066	0.035		0.069	0.042	
	# Iterations	5	5		5	5		5	6	
	Total	0.335	0.203	0.064	0.379	0.223	0.067	0.436	0.300	0.078
	# Neighbors	6	6	6	20	20	20	59	59	59
1000	$\hat{\alpha}$	0.669	0.477	0.678	0.615	0.376	0.718	0.474	0.274	0.696
		(0.267)	(0.106)	(0.279)	(0.414)	(0.247)	(0.483)	(0.554)	(1.010)	(0.944)
	$\hat{\beta}_0$	0.000	0.000	0.000	0.002	0.002	0.002	0.001	0.001	−0.001
		(0.019)	(0.025)	(0.019)	(0.027)	(0.029)	(0.027)	(0.041)	(0.078)	(0.054)
	$\hat{\beta}_1$	0.985	1.004	0.961	1.009	1.002	0.990	1.020	1.075	1.003
		(0.093)	(0.076)	(0.084)	(0.105)	(0.076)	(0.076)	(0.105)	(0.296)	(0.078)
	Time:									
	Loop	1.557	0.457		1.635	0.959		1.645	3.483	
	# Iterations	5	5		5	5		5	5	
	Total	8.076	2.621	0.561	8.765	5.337	0.674	9.473	19.264	1.302
	# Neighbors	9	9	9	34	34	34	120	120	120
2000	$\hat{\alpha}$	0.705	0.492	0.688	0.698	0.443	0.738	0.553	0.422	0.721
		(0.256)	(0.070)	(0.245)	(0.357)	(0.141)	(0.386)	(0.494)	(0.544)	(0.670)
	$\hat{\beta}_0$	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	−0.001
		(0.012)	(0.016)	(0.012)	(0.015)	(0.018)	(0.015)	(0.024)	(0.030)	(0.026)
	$\hat{\beta}_1$	0.979	1.004	0.966	1.008	1.005	0.993	1.007	1.050	0.997
		(0.065)	(0.056)	(0.064)	(0.062)	(0.052)	(0.053)	(0.058)	(0.209)	(0.056)
	Time:									
	Loop	8.178	3.004		7.848	7.641		6.861	21.501	
	# Iterations	5	5		5	5		5	5	
	Total	42.737	16.435	2.237	42.574	41.329	2.379	39.126	116.620	4.037
	# Neighbors	10	10	10	36	36	36	136	136	136

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

TABLE A2.2.4: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.8$ and $\mathbf{W}$ based on the radial distance criterion.

	δ_R	1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.797	0.575	1.582	0.852	0.397	1.732	1.392	0.004	1.160
		(0.589)	(0.509)	(1.418)	(0.985)	(1.460)	(2.137)	(2.318)	(4.172)	(5.572)
	$\hat{\beta}_0$	−0.002	−0.010	−0.007	0.016	0.024	0.010	−0.032	−0.124	0.113
		(0.123)	(0.140)	(0.305)	(0.285)	(0.484)	(0.604)	(0.696)	(1.847)	(2.455)
	$\hat{\beta}_1$	0.933	0.995	0.748	0.969	1.000	0.873	1.013	1.040	0.912
		(0.274)	(0.338)	(0.349)	(0.277)	(0.272)	(0.284)	(0.260)	(0.284)	(0.275)
	Time:									
	Loop	0.066	0.032		0.067	0.036		0.071	0.042	
	# Iterations	5	5		5	5		6	6	
	Total	0.389	0.211	0.062	0.410	0.242	0.068	0.467	0.291	0.076
	# Neighbors	6	6	6	21	21	21	60	60	60
1000	$\hat{\alpha}$	1.403	0.690	1.584	1.418	1.009	1.853	1.173	1.330	1.824
		(0.641)	(0.117)	(0.836)	(0.796)	(0.711)	(1.157)	(0.991)	(1.257)	(1.398)
	$\hat{\beta}_0$	0.002	−0.004	0.001	−0.007	−0.001	−0.003	0.010	−0.018	−0.005
		(0.044)	(0.014)	(0.063)	(0.060)	(0.096)	(0.100)	(0.113)	(0.163)	(0.131)
	$\hat{\beta}_1$	0.880	0.956	0.799	0.941	1.114	0.916	0.988	1.110	0.973
		(0.171)	(0.080)	(0.215)	(0.109)	(0.350)	(0.114)	(0.076)	(0.258)	(0.081)
	Time:									
	Loop	1.673	0.449		1.581	1.046		1.653	3.683	
	# Iterations	6	6		6	6		5	6	
	Total	11.297	3.034	0.521	10.185	6.592	0.569	10.449	22.218	1.276
	# Neighbors	8	8	8	39	39	39	125	125	125
2000	$\hat{\alpha}$	1.255	0.650	1.616	1.522	0.825	1.880	1.795	1.491	1.876
		(0.455)	(0.150)	(0.844)	(0.798)	(0.363)	(1.141)	(1.159)	(1.087)	(1.283)
	$\hat{\beta}_0$	−0.007	−0.019	0.002	−0.003	0.009	−0.001	0.001	0.007	0.001
		(0.007)	(0.019)	(0.044)	(0.060)	(0.040)	(0.071)	(0.079)	(0.115)	(0.086)
	$\hat{\beta}_1$	0.890	0.986	0.813	0.953	1.129	0.926	0.983	1.264	0.975
		(0.110)	(0.014)	(0.196)	(0.091)	(0.272)	(0.091)	(0.060)	(0.420)	(0.060)
	Time:									
	Loop	6.707	2.594		5.965	7.410		5.962	22.797	
	# Iterations	7	5		6	6		5	6	
	Total	48.554	14.623	1.728	37.195	47.109	1.916	35.994	135.114	3.671
	# Neighbors	8	8	8	44	44	44	143	143	143

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

TABLE A2.2.5: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	−0.005	−0.001	0.004	0.198	−0.215	0.017	0.249	−0.499	−0.024
		(0.225)	(0.215)	(0.250)	(0.570)	(0.847)	(0.953)	(0.776)	(1.401)	(1.445)
	$\hat{\beta}_0$	0.003	0.005	0.002	−0.004	0.019	−0.001	0.004	0.004	0.003
		(0.142)	(0.152)	(0.141)	(0.132)	(0.246)	(0.183)	(0.146)	(0.377)	(0.266)
	$\hat{\beta}_1$	1.050	1.083	1.024	1.106	1.097	1.018	1.133	1.140	1.033
		(0.256)	(0.267)	(0.257)	(0.295)	(0.279)	(0.257)	(0.289)	(0.305)	(0.260)
	Time:									
	Loop	0.061	0.032		0.063	0.035		0.066	0.037	
	# Iterations	4	4		5	5		5	5	
	Total	0.288	0.181	0.061	0.353	0.224	0.066	0.376	0.242	0.064
	% Asymmetry	0.274	0.274	0.274	0.151	0.151	0.151	0.146	0.146	0.146
1000	$\hat{\alpha}$	0.012	−0.015	0.004	0.184	−0.260	0.028	0.264	−0.654	−0.025
		(0.227)	(0.229)	(0.230)	(0.561)	(1.102)	(0.812)	(0.645)	(1.959)	(1.345)
	$\hat{\beta}_0$	0.001	0.002	0.002	0.000	0.002	0.003	0.001	0.005	0.001
		(0.044)	(0.045)	(0.044)	(0.040)	(0.076)	(0.053)	(0.041)	(0.127)	(0.074)
	$\hat{\beta}_1$	1.003	1.006	1.002	1.016	1.028	1.003	1.015	1.031	1.006
		(0.077)	(0.077)	(0.078)	(0.082)	(0.174)	(0.077)	(0.087)	(0.138)	(0.078)
	Time:									
	Loop	1.382	0.495		1.568	2.324		1.575	5.631	
	# Iterations	3	4		5	5		5	5	
	Total	5.420	2.430	0.549	8.526	12.931	0.969	9.796	31.838	1.970
	% Asymmetry	0.130	0.130	0.130	0.106	0.106	0.106	0.125	0.125	0.125
2000	$\hat{\alpha}$	0.006	−0.021	0.000	0.176	−0.309	−0.030	0.303	−0.426	0.032
		(0.232)	(0.242)	(0.232)	(0.583)	(1.104)	(0.838)	(0.708)	(1.903)	(1.258)
	$\hat{\beta}_0$	−0.001	−0.001	−0.001	−0.001	0.003	0.000	0.000	−0.001	0.002
		(0.029)	(0.031)	(0.030)	(0.030)	(0.052)	(0.040)	(0.028)	(0.070)	(0.049)
	$\hat{\beta}_1$	1.000	1.001	0.999	1.005	1.019	1.000	1.007	1.020	1.001
		(0.053)	(0.053)	(0.053)	(0.055)	(0.119)	(0.054)	(0.056)	(0.113)	(0.055)
	Time:									
	Loop	6.846	4.995		6.435	38.929		6.486	49.120	
	# Iterations	3	4		5	5		5	5	
	Total	26.624	21.849	2.248	36.496	205.426	5.626	47.796	279.975	14.784
	% Asymmetry	0.102	0.102	0.102	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

TABLE A2.2.6: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.2$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.164	0.163	0.214	0.325	−0.121	0.250	0.293	−0.380	0.203
		(0.214)	(0.218)	(0.268)	(0.556)	(0.828)	(0.900)	(0.702)	(1.394)	(1.440)
	$\hat{\beta}_0$	0.002	0.001	0.001	−0.006	−0.002	0.006	−0.009	0.002	0.014
		(0.119)	(0.131)	(0.115)	(0.137)	(0.273)	(0.184)	(0.142)	(0.305)	(0.231)
	$\hat{\beta}_1$	1.026	1.071	0.990	1.100	1.103	1.010	1.098	1.148	1.013
		(0.251)	(0.271)	(0.254)	(0.291)	(0.270)	(0.258)	(0.274)	(0.392)	(0.258)
	Time:									
	Loop	0.062	0.033		0.064	0.035		0.064	0.037	
	# Iterations	4	4		5	5		5	5	
	Total	0.303	0.189	0.061	0.346	0.221	0.064	0.365	0.242	0.067
	% Asymmetry	0.273	0.273	0.273	0.152	0.152	0.152	0.147	0.147	0.147
1000	$\hat{\alpha}$	0.237	0.185	0.236	0.257	−0.087	0.226	0.366	−0.344	0.196
		(0.241)	(0.188)	(0.239)	(0.518)	(0.933)	(0.814)	(0.614)	(1.827)	(1.333)
	$\hat{\beta}_0$	0.000	0.000	0.000	0.003	0.000	0.001	0.000	0.004	−0.002
		(0.033)	(0.035)	(0.033)	(0.038)	(0.065)	(0.050)	(0.040)	(0.105)	(0.074)
	$\hat{\beta}_1$	1.002	1.006	0.999	1.010	1.036	1.001	1.016	1.033	1.006
		(0.082)	(0.080)	(0.080)	(0.076)	(0.164)	(0.075)	(0.085)	(0.135)	(0.079)
	Time:									
	Loop	1.300	0.488		1.474	2.163		1.546	5.685	
	# Iterations	4	4		5	5		5	5	
	Total	5.288	2.398	0.502	7.908	11.903	0.943	9.723	32.795	1.964
	% Asymmetry	0.130	0.130	0.130	0.106	0.106	0.106	0.125	0.125	0.125
2000	$\hat{\alpha}$	0.227	0.174	0.224	0.300	−0.081	0.221	0.386	−0.340	0.190
		(0.238)	(0.194)	(0.237)	(0.538)	(0.992)	(0.831)	(0.647)	(1.730)	(1.249)
	$\hat{\beta}_0$	−0.001	−0.001	−0.001	0.000	0.000	0.001	−0.001	0.001	−0.002
		(0.023)	(0.024)	(0.023)	(0.026)	(0.045)	(0.035)	(0.029)	(0.072)	(0.047)
	$\hat{\beta}_1$	1.000	1.003	0.999	1.009	1.029	1.003	1.007	1.022	1.000
		(0.053)	(0.053)	(0.053)	(0.067)	(0.125)	(0.052)	(0.061)	(0.098)	(0.056)
	Time:									
	Loop	5.646	4.368		6.266	38.941		6.251	48.360	
	# Iterations	4	4		5	5		5	5	
	Total	22.771	19.424	1.836	36.008	203.903	5.603	46.456	271.254	14.558
	% Asymmetry	0.103	0.103	0.103	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

TABLE A2.2.7: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.5$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.375	0.393	0.581	0.528	0.132	0.795	0.537	−0.091	0.739
		(0.260)	(0.238)	(0.397)	(0.595)	(0.852)	(1.233)	(0.660)	(1.539)	(1.476)
	$\hat{\beta}_0$	0.004	0.005	0.000	−0.009	0.032	0.007	−0.002	−0.015	−0.005
		(0.090)	(0.106)	(0.083)	(0.124)	(0.264)	(0.282)	(0.124)	(0.456)	(0.295)
	$\hat{\beta}_1$	1.006	1.092	0.870	1.099	1.114	0.990	1.109	1.162	1.002
		(0.241)	(0.301)	(0.285)	(0.305)	(0.324)	(0.253)	(0.314)	(0.407)	(0.265)
	Time:									
	Loop	0.064	0.033		0.065	0.035		0.066	0.037	
	# Iterations	5	5		5	5		5	5	
	Total	0.354	0.206	0.062	0.375	0.227	0.063	0.385	0.246	0.066
	% Asymmetry	0.274	0.274	0.274	0.151	0.151	0.151	0.147	0.147	0.147
1000	$\hat{\alpha}$	0.683	0.459	0.706	0.509	0.316	0.721	0.463	0.073	0.744
		(0.294)	(0.111)	(0.315)	(0.569)	(0.855)	(0.853)	(0.672)	(1.682)	(1.274)
	$\hat{\beta}_0$	0.000	0.000	0.000	0.000	0.004	0.001	−0.002	−0.004	−0.003
		(0.020)	(0.025)	(0.020)	(0.037)	(0.066)	(0.048)	(0.041)	(0.114)	(0.072)
	$\hat{\beta}_1$	0.999	1.006	0.972	1.010	1.060	1.000	1.018	1.053	0.999
		(0.093)	(0.075)	(0.080)	(0.078)	(0.199)	(0.075)	(0.099)	(0.211)	(0.082)
	Time:									
	Loop	1.469	0.521		1.499	2.183		1.562	5.623	
	# Iterations	5	5		5	5		5	5	
	Total	7.873	2.937	0.517	8.560	12.478	0.938	10.004	32.083	1.964
	% Asymmetry	0.131	0.131	0.131	0.106	0.106	0.106	0.126	0.126	0.126
2000	$\hat{\alpha}$	0.702	0.462	0.721	0.547	0.390	0.731	0.485	0.293	0.698
		(0.301)	(0.104)	(0.321)	(0.557)	(0.814)	(0.832)	(0.692)	(1.647)	(1.338)
	$\hat{\beta}_0$	−0.001	−0.001	−0.001	0.002	0.002	0.002	0.001	0.005	0.003
		(0.014)	(0.018)	(0.014)	(0.025)	(0.041)	(0.032)	(0.030)	(0.065)	(0.056)
	$\hat{\beta}_1$	0.998	1.003	0.984	1.009	1.058	1.000	1.011	1.033	0.999
		(0.057)	(0.056)	(0.058)	(0.055)	(0.190)	(0.053)	(0.067)	(0.176)	(0.052)
	Time:									
	Loop	7.442	5.147		6.422	38.466		6.764	48.674	
	# Iterations	5	5		5	5		5	5	
	Total	40.022	27.738	1.980	38.652	208.172	5.454	48.511	274.855	14.525
	% Asymmetry	0.102	0.102	0.102	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

TABLE A2.2.8: Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.8$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.377	0.637	1.609	0.839	0.411	1.928	0.767	0.166	1.897
		(0.472)	(0.249)	(15.527)	(0.699)	(0.698)	(1.920)	(0.885)	(1.562)	(2.322)
	$\hat{\beta}_0$	0.012	0.003	0.171	0.004	−0.004	−0.008	0.023	0.070	0.031
		(0.079)	(0.092)	(5.020)	(0.174)	(0.197)	(0.442)	(0.238)	(0.633)	(0.730)
	$\hat{\beta}_1$	0.805	0.984	0.577	0.954	0.998	0.798	1.002	1.048	0.860
		(0.296)	(0.359)	(0.478)	(0.255)	(0.255)	(0.310)	(0.279)	(0.341)	(0.284)
	Time:									
	Loop	0.064	0.033		0.065	0.034		0.065	0.037	
	# Iterations	5	6		6	5		6	5	
	Total	0.356	0.234	0.065	0.409	0.221	0.063	0.422	0.252	0.066
	% Asymmetry	0.272	0.272	0.272	0.152	0.152	0.152	0.145	0.145	0.145
1000	$\hat{\alpha}$	1.523	0.627	1.773	1.049	1.329	1.942	0.877	1.190	1.720
		(0.792)	(0.176)	(1.028)	(0.824)	(1.283)	(1.461)	(0.854)	(1.752)	(1.618)
	$\hat{\beta}_0$	0.016	0.000	−0.002	−0.005	0.005	0.003	−0.001	0.008	−0.001
		(0.046)	(0.019)	(0.078)	(0.053)	(0.176)	(0.139)	(0.066)	(0.192)	(0.172)
	$\hat{\beta}_1$	0.881	0.979	0.817	0.997	1.150	0.968	0.994	1.074	0.980
		(0.139)	(0.090)	(0.197)	(0.107)	(0.307)	(0.085)	(0.086)	(0.270)	(0.080)
	Time:									
	Loop	1.464	0.541		1.456	2.188		1.512	5.654	
	# Iterations	6	6		6	6		5	6	
	Total	9.445	3.432	0.421	8.984	13.301	0.860	10.190	33.624	1.894
	% Asymmetry	0.128	0.128	0.128	0.105	0.105	0.105	0.125	0.125	0.125
2000	$\hat{\alpha}$	1.786	0.621	1.850	1.759	1.477	1.900	1.462	1.130	1.701
		(1.034)	(0.179)	(1.089)	(1.236)	(1.190)	(1.391)	(1.227)	(1.747)	(1.590)
	$\hat{\beta}_0$	−0.001	0.010	−0.001	−0.002	−0.001	−0.005	0.002	−0.018	−0.007
		(0.031)	(0.010)	(0.065)	(0.065)	(0.133)	(0.096)	(0.065)	(0.139)	(0.119)
	$\hat{\beta}_1$	0.859	0.975	0.886	0.994	1.176	0.987	0.998	1.056	0.990
		(0.150)	(0.025)	(0.125)	(0.061)	(0.348)	(0.057)	(0.059)	(0.190)	(0.054)
	Time:									
	Loop	6.170	4.801		5.842	39.105		5.834	47.706	
	# Iterations	6	6		5	6		5	6	
	Total	38.760	30.506	1.885	36.167	227.035	5.528	43.368	282.436	14.560
	% Asymmetry	0.100	0.100	0.100	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

B2. Empirical application

TABLE B2.1: Descriptive statistics for the variables included in the empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas

	Mean	Std. Dev.	Min	Q_1	Median	Q_3	Max	N
BCI	0.147	0.354	0.000	0.000	0.000	0.000	1.000	4,848
AQI_{min}	0.131	0.086	0.000	0.060	0.120	0.190	0.430	4,848
AQI_{max}	1.514	0.931	0.380	1.120	1.430	1.710	22.120	4,848
% days O_3	0.480	0.276	0.000	0.312	0.468	0.682	1.000	4,848
% days $PM_{2.5}$	0.406	0.273	0.000	0.192	0.386	0.584	1.000	4,848
% days PM_{10}	0.030	0.101	0.000	0.000	0.000	0.011	1.000	4,848
% days CO	0.008	0.048	0.000	0.000	0.000	0.000	0.738	4,848
% days SO_2	0.052	0.126	0.000	0.000	0.000	0.019	0.962	4,848
% days NO_2	0.024	0.056	0.000	0.000	0.000	0.019	0.499	4,848
% days Above Moderate	0.370	0.213	0.000	0.202	0.344	0.504	0.966	4,848
% days Exceptional Events	0.024	0.092	0.000	0.000	0.000	0.000	0.940	4,848

FIGURE B2.1: Centroids of the U.S. Metropolitan Statistical Areas included in the empirical application

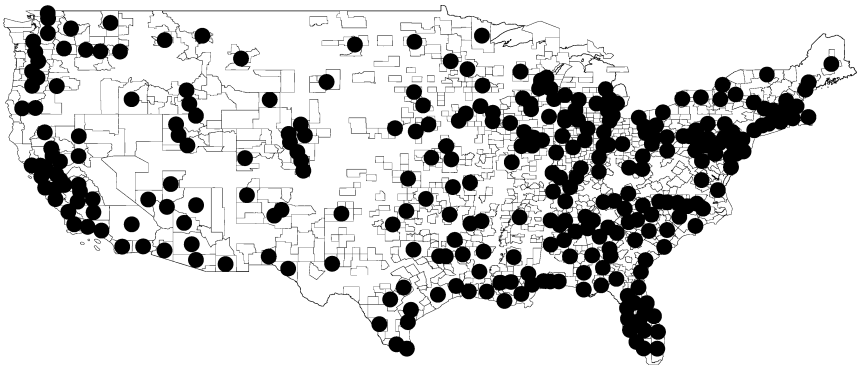


TABLE B2.2: Spatial lag Probit estimation results for the empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas

	Dependent variable: BCI					
	UNRESTRICTED MODEL			RESTRICTED MODEL		
	(iGMMa)	(iGMM)	(LGMM)	(iGMMa)	(iGMM)	(LGMM)
Intercept	0.046	0.178	−0.423	−0.059	0.194	−0.328
	(0.806)	(0.932)	(1.136)	(0.796)	(0.900)	(1.144)
AQI_{min}	−2.370**	−2.589**	−3.022***	−2.154**	−2.339**	−2.749***
	(1.039)	(1.061)	(0.990)	(1.025)	(1.047)	(0.989)
AQI^2_{min}	6.630**	8.078**	8.850***	6.213*	7.423**	8.419**
	(3.354)	(3.319)	(3.261)	(3.336)	(3.299)	(3.280)
AQI_{max}	−0.022	−0.033	0.024			
	(0.085)	(0.090)	(0.094)			
AQI^2_{max}	0.004	0.004	−0.003			
	(0.005)	(0.006)	(0.008)			
% days O_3	0.776	0.780	0.674	0.703***	0.640**	0.643***
	(0.589)	(0.622)	(0.553)	(0.247)	(0.250)	(0.236)
% days $PM_{2.5}$	0.396	0.366	0.349	0.445*	0.325	0.435**
	(0.586)	(0.611)	(0.573)	(0.232)	(0.231)	(0.215)
% days PM_{10}	0.965	0.967	0.884			
	(0.717)	(0.750)	(0.632)			
% days SO_2	−0.717	−0.524	−0.917			
	(0.676)	(0.698)	(0.662)			
% days NO_2	0.400	0.395	0.626			
	(0.942)	(0.964)	(0.897)			
% days Above Moderate	0.334	0.231	0.421			
	(0.352)	(0.375)	(0.327)			
% days Exceptional Events	0.116	0.241	0.056			
	(0.359)	(0.371)	(0.373)			
Spatial Lag ($\hat{\alpha}$)	0.771***	0.686***	0.954**	0.768***	0.711***	1.051***
	(0.051)	(0.056)	(0.375)	(0.052)	(0.055)	(0.388)
Observations	4,848	4,848	4,848	4,848	4,848	4,848
# Neighbors (average)	16	16	16	16	16	16
# Iterations	11	14	–	15	18	–
Total Time (in seconds)	33.999	123.689	7.102	34.928	154.002	6.891
# Instruments	141	141	141	113	113	113
Wald test (overall sig.)	24.155	23.004	30.584	11.123	12.490	15.435
(p-value)	(0.012)	(0.018)	(0.001)	(0.025)	(0.014)	(0.004)
Wald test (excl. restr.)[1]	–	–	–	12.783	9.999	15.781
(p-value)	(–)	(–)	(–)	(0.078)	(0.189)	(0.027)
Hansen's J test	85.178	49.565	1124.975	77.899	44.497	1530.180
(p-value)	(0.911)	≈(1.000)	≈(0.000)	(0.638)	≈(1.000)	≈(0.000)
McFadden R^2	0.038	0.061	−0.428	0.037	0.058	−0.662
$\rho^2(\hat{Y}, Y)$	0.056	0.061	0.033	0.055	0.059	0.003
$\%(\hat{Y} = Y)$	0.861	0.856	0.679	0.859	0.856	0.559

NOTE: Robust standard errors in parentheses, based on Kelejian and Prucha (2007). Time effects and Mundlak variables were added. Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively.

[1] Wald test for exclusion restrictions. Under the null hypothesis, the coefficients for the variables AQI_{max}, AQI^2_{max}, % days PM_{10}, % days SO_2, % days NO_2, % days Above Moderate and % days Exceptional Events are jointly equal to zero.

Chapter 3

Fractional responses with spatial dependence

3.1. Introduction

Fractional responses refer to continuous variables on the interval $[0, 1]$. Depending on the nature of the data, the fractional responses can also be defined on $[0, 1)$, $(0, 1]$ and $(0, 1)$. These variables differ from the binary case, once they can assume any value between 0 and 1.

Several examples can be found in the literature, where fractional responses are used as the dependent variable. The participation rates in 401(k) pension plans (Papke and Wooldridge, 1996), test pass rates (Papke and Wooldridge, 2008), the bilateral intra-industry trade index (Proença and Faustino, 2015), the size-to-exports ratio of a firm (Wagner, 2001), the exports-to-growth ratio of a firm (Fryges, 2009), the degree of financial leverage of a firm (McDonald, 2009; Ramalho and Ramalho, 2017; Ramalho et al., 2014) and the proportion of losses covered by insurance companies (Sigrist and Stahel, 2011) are examples of dependent variables that are bounded between zero and one. However, none of the existing articles consider a spatial approach to estimation. In fact, the Spatial Econometrics literature on fractional responses is considerably scarce. To the best of the author's knowledge, only Lin and Lee (2010) and Xu and Lee (2015a) explicitly consider a spatial framework to develop models for fractional responses.

Lin and Lee (2010) estimate a linear spatial lag model to address the spatial, social and economic effects over teenage pregnancy rates in the U.S. Counties. However, considering a linear approach to model fractional responses has two major

drawbacks, similar to those of the linear probability model for dichotomous responses (see also Papke and Wooldridge, 1996). One, the predicted values of the fractional dependent variable do not necessarily lie in the closed interval $[0, 1]$. Two, the linear model fails to account for diminishing effects of the exogenous explanatory variables over the teenage pregnancy rates.

More recently, Xu and Lee (2015a) made an important theoretical contribution, addressing the specification and the estimation of spatially lagged models for fractional responses in the open interval $(0, 1)$. The authors consider a nonlinear spatial lag model with nonadditive errors, $Y_i = F\left(\alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta} + e_i\right)$, where $Y_i \in (0, 1)$ and e_i is an i.i.d. random error. The function $F(\cdot)$ is assumed to be strictly increasing and continuous on $\mathbb{R}$. Under these assumptions, the inverse, $F^{-1}(\cdot)$, exists and the responses, Y_i, can be transformed in such a way that the transformed model can handle any real value. Hence, the model can be written as $F^{-1}(Y_i) = \alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta} + e_i$ and the transformed fractional dependent variable, $F^{-1}(Y_i)$, is now a linear function of the parameters. This procedure is usually referred to as the inverse transformation approach. Estimation is based on Maximum Likelihood or Two-Stage Least Squares methods. These estimators are shown to be consistent and asymptotically normal.

Nevertheless, there are two important limitations related to Xu and Lee (2015a) inverse transformation approach. First, the inverse transformation is not well defined if Y_i takes on the values 0, 1 or both with positive probability. The inverse function $F^{-1}(Y_i) \to \pm\infty$, as $Y_i \to 0$ or $Y_i \to 1$. Second, even if the inverse transformation is well defined – i.e., $Y_i \in (0, 1)$ –, it is not possible to recover the conditional expectation of the retransformed fractional dependent variable, $\mathrm{E}(Y_i \,|\, \mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i})$, without further assumptions. This is because, in general, $\mathrm{E}(Y_i \,|\, \mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i}) \neq F(\mathrm{E}[Y_i \,|\, \mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i}])$. The conditional expectation of the retransformed fractional dependent variable is given by $\mathrm{E}(Y_i \,|\, \mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i}) = \int_{\mathcal{A}} F\left(\alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}\right) \mathrm{d}\,\mathrm{D}(\mathbf{e} \,|\, \mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i})$, where $\mathcal{A} \subseteq \mathbb{R}^N$ and $\mathrm{D}(\mathbf{e} \,|\, \mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i})$ is the distribution function of $\mathbf{e} = F^{-1}(\mathbf{Y}) - \alpha \mathbf{W}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ conditional on $(\mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i})$,

with $\mathbf{Y}_{-i}$ the $(N-1)\times 1$ vector of responses excluding the ith response. To estimate $\mathrm{E}(\mathbf{Y}\,|\,\mathbf{X},\mathbf{W},\mathbf{Y}_{-i})$, the conditional distribution $\mathrm{D}(\mathbf{e}\,|\,\mathbf{X},\mathbf{W},\mathbf{Y}_{-i})$ has to be estimated first, once the responses, $\mathbf{Y}$, are correlated with the errors, $\mathbf{e}$. Alternatively, one may consider the smearing method of Duan (1983). This method has the advantage of not requiring a fully parametric specification for the conditional distribution $\mathrm{D}(\mathbf{e}\,|\,\mathbf{X},\mathbf{W},\mathbf{Y}_{-i})$, to estimate $\mathrm{E}(Y_i\,|\,\mathbf{X},\mathbf{W},\mathbf{Y}_{-i})$. The smearing estimate for the conditional expectation is given by $\widehat{\mathrm{E}}(Y_i\,|\,\mathbf{X},\mathbf{W},\mathbf{Y}_{-i}) = N^{-1}\sum_{l=1}^{N} F\left(\hat{\alpha}\sum_{j\neq i} w_{ij}Y_j + \mathbf{X}_i\hat{\boldsymbol{\beta}} + \hat{e}_l\right)$, where $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ can be obtained from the estimators proposed by Xu and Lee (2015a), considering the regression with the transformed fractional dependent variable, and $\hat{e}_l$ are the corresponding residuals. However, Mullahy (1998), Manning and Mullahy (2001) and Manning et al. (2005) show that, under a non-spatial nonlinear framework and for the case where the error is heteroskedastic, the smearing estimate suffers from severe bias. Using similar arguments, their findings also apply to spatial nonlinear approaches. The estimates $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are identified, but heteroskedasticity yields an attenuation bias. As a consequence, the smearing estimate for $\widehat{\mathrm{E}}(Y_i\,|\,\mathbf{X},\mathbf{W},\mathbf{Y}_{-i})$ can be quite misleading, even if an estimate for the error variance is used. See also Papke and Wooldridge (1996, 2008) for further discussion on fractional responses under non-spatial frameworks.

In this chapter, two new specifications to model fractional responses with spatial dependence are proposed. Rather than applying *ad hoc* transformations or other arbitrary adjustments to the fractional dependent variable, both specifications explicitly consider a functional form for the conditional expectation of interest. Thus, responses defined in the closed interval $[0,1]$ are admitted. The suggested specifications rely on a set of non-restrictive assumptions, based on Xu and Lee (2015a).

The first specification, the Fractional Response Spatial Lag Model (FRSLM), extends the work of Papke and Wooldridge (1996) to model fractional responses with spatial data. Spatial dependence is introduced in the proposed specification through a spatial lag of the fractional dependent variable, inside a nonlinear function. In fact, the FRSLM is analogous to the specification of Xu and Lee (2015a), differing from the latter in two important ways. One, the FRSLM considers additive errors, whereas the specification of Xu and Lee (2015a) considers non-additive

errors. Hence, under the setup of the FRSLM, the inverse transformation approach is not useful, even if the responses are defined in the open interval $(0, 1)$. Two, the FRSLM allows to directly estimate $\mathrm{E}(Y_i \mid \mathbf{X}, \mathbf{W}, \mathbf{Y}_{-i})$ and the corresponding partial effects, without requiring additional operations, such as multivariate integration or smearing-type estimators. Estimation can be addressed by simple parametric procedures, such as Two-Stage Nonlinear Least Squares. Furthermore, the predicted values lie inside the admissible interval.

The second specification, the approximate Fractional Response Spatial Lag Model (aFRSLM), consists in a first order series approximation of the FRSLM around the spatial lag parameter equal to zero. The usefulness of this approach is threefold. First, it allows to write the FRSLM as an approximate reduced form, with a tractable analytic expression, comparable to a partially linear model. Note that, for general nonlinear simultaneous models, the expression for the reduced form (if it exists) does not have a known functional form and can only be obtained through computationally complex numerical methods. Second, it allows to straightforwardly obtain an approximate estimate for $\mathrm{E}(Y_i \mid \mathbf{X}, \mathbf{W})$ and for the corresponding partial effects. Just as in the FRSLM, no additional operations are required and the estimation can be addressed by simple parametric methods. Third, the partial effects based on the approximate reduced form can be expressed as a sum of nonlinear functions of the exogenous explanatory variables and their spatial lags. This specification is of particular interest, regarding policy analysis, especially if the sample of spatial units is viewed as resulting from a steady-state relationship between the responses and the exogenous explanatory variables (see also LeSage and Pace, 2010). In this way, policy changes can be interpreted as movements to another steady-state solution. A similar interpretation is not possible for the FRSLM, once the corresponding partial effects depend on both exogenous explanatory variables and endogenous spatially lagged responses. The aFRSLM, however, has one important limitation. Predictions may fall outside the admissible interval. Nevertheless, simulations show that the proportion of predicted responses $\hat{Y}_i \notin [0, 1]$ is negligible.

Other approaches to model fractional responses are available. Examples are the two-part models (see Cook et al., 2008; Ramalho et al., 2011; Ramalho and

da Silva, 2009) and the two-limit Tobit models (see amongst others, Kieschnick and McCullough, 2003; McDonald, 2009; Sigrist and Stahel, 2011; Wagner, 2001). But they rely on certain characteristics of the fractional dependent variable, such as piled-up observations at the corners 0 and/or 1. In practice, this may not be necessarily the case for the majority of applications using fractional responses and these approaches end up to be logically inconsistent. See also Wooldridge (2010) for a discussion on this matter.

Estimation is addressed by the iterative Generalized Method of Moments (iGMM) procedure of Klier and McMillen (2008), revisited in Chapter 2. The spatial heteroskedasticity and spatial autocorrelation (spatial HAC) robust estimator of Kelejian and Prucha (2007) is considered, to produce valid inference for the asymptotic covariance estimator of the GMM estimator for the unknown parameter vector.

A detailed Monte Carlo simulation study will show that the iGMM estimator is accurate and precise to estimate the unknown parameter vector, especially for the FRSLM. In addition, for higher levels of spatial dependence and denser spatial weighting matrix, the iGMM estimator for the aFRSLM proves to be more accurate than the FRSLM to estimate the spatial lag parameter. Furthermore, the partial effects obtained from the estimation of both the FRSLM and the aFRSLM tend to be quite accurate and precise. The performance of a Two-Stage Least Squares (2SLS) estimator, for the Linear Spatial Lag Model (LSLM), is also assessed and compared with the iGMM estimator for the proposed specifications. Results show that the 2SLS parameter estimates and the estimates for the partial effects can be quite misleading, as expected.

The remainder of the chapter is organized as follows. Section 3.2 introduces the two new specifications (FRSLM and aFRSLM) for spatially lagged fractional responses and deduces their corresponding partial effects. Section 3.3 presents the GMM estimation procedure and the gradients are deduced. Section 3.4 conducts an extensive Monte Carlo simulation study. The structure of the simulated fractional dependent variable is assessed. Next, the statistical properties of both the iGMM estimator and 2SLS estimator are discussed for each of the two specifications and the linear model, respectively. In addition, the statistical properties of the corresponding

partial effects are discussed, as well, using the previous estimates. Finally, section 3.5 concludes. The Monte Carlo simulation results are summarized in Section A3.

3.2. Specifications and quantities of interest for spatially lagged fractional responses

In this section, two specifications to model spatially lagged fractional responses are presented and the underlying assumptions are stated. Also, the corresponding partial effects are deduced. The first specification is the Fractional Response Spatial Lag Model (FRSLM) and the second specification is the approximate Fractional Response Spatial Lag Model (aFRSLM). These specifications can handle responses that take on values in the closed interval $[0, 1]$. This is because no transformations are applied to the fractional dependent variable. The FRSLM considers the atractive functional forms developed by Papke and Wooldridge (1996) and introduces spatial dependence into their suggested specifications through a spatial lag of the fractional dependent variable. The FRSLM is also related to the specification of Xu and Lee (2015a), differing in the way the errors enter the expression. The former considers additive errors, whereas the latter considers non-additive errors. The aFRSLM consists in a first order series approximation of the FRSLM around the spatial lag parameter equal to zero. Under this specification, the FRSLM can be written as an approximate reduced form, with a tractable analytic expression, that is generally not possible for the majority of nonlinear simultaneous models. Additionally, the partial effects can be approximated by sums of nonlinear functions of exogenous explanatory variables and their spatial lags. In this way, measuring the effects of policy changes becomes more intuitive, as the reduced form can be interpreted as a steady-state relationship between the responses and the exogenous explanatory variables.

3.2.1. The Fractional Response Spatial Lag Model (FRSLM)

The Fractional Response Spatial Lag Model (FRSLM) follows as:

$$Y_i = G\left(\alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta}\right) + u_i, \qquad i, j = 1, 2, \ldots, N \tag{56}$$

where Y_i is a bounded dependent variable on the interval $[0, 1]$, for the ith spatial unit, with $i = 1, 2, \ldots, N$ and N denoting the total number of spatial units. The function $G(\eta)$ is a link function defined in the open interval $(0, 1)$, for all $\eta \in \mathbb{R}$, and η is usually referred to as the index. The coefficients w_{ij} are non-negative scalars that correspond to the spatial weights of unit j on unit i, with $i \neq j$ and $j = 1, 2, \ldots, N$. Also, by convention, $w_{ii} = 0$, for all i. The scalar α is the spatial lag parameter. he $1 \times K$ vector $\mathbf{X}_i$ includes the observations for a set of K exogenous explanatory variables and a constant, for the unit i. The $K \times 1$ vector $\boldsymbol{\beta}$ is the corresponding vector of regression parameters. The disturbance term, u_i, is a random error, for the ith spatial unit, defined on the closed interval $[-G(\eta_i), 1 - G(\eta_i)]$.

Stacking over the spatial units yields:

$$\mathbf{Y} = G(\alpha \mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta}) + \mathbf{u} \tag{57}$$

where $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_N]^\intercal$ is an $N \times 1$ vector of fractional responses. The $N \times K$ matrix of exogenous explanatory variables is $\mathbf{X} = [\mathbf{X}_1^\intercal, \mathbf{X}_2^\intercal, \ldots, \mathbf{X}_N^\intercal]^\intercal$, with $\mathbf{X}_i = [1, x_{i2}, \ldots, x_{iK}]$, for $i = 1, 2, \ldots, N$, and the corresponding $K \times 1$ parameter vector is $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_K]^\intercal$. The $N \times 1$ vector of random errors is $\mathbf{u} = [u_1, u_2, \ldots, u_N]^\intercal$. The matrix $\mathbf{W}$ is the $N \times N$ spatial weighting matrix, with generic element w_{ij}. By definition, $\mathbf{W}$ is a non-stochastic non-negative matrix with zeros on its main diagonal.

Before presenting additional details, it is useful to list a set of assumptions that underlie the proposed model. These assumptions are commonly used in the literature of nonlinear models with spatial dependence (see e.g. Conley, 1999; Jenish and Prucha, 2009, 2012) and follow along the lines of Xu and Lee (2015a). They focus on the geographical setting, the structure of the spatial weighting matrix, on the properties of the link function and on the statistical properties of the error term.

Assumption 3.1. The spatial units are located in a region $D_N \subset D \subset \mathbb{R}^d$, where $\lim_{N\to\infty} |D_N| = \infty$ and $\mathbb{R}^d$ is the finite dimensional Euclidean space of dimension d. The distance between every two spatial units is larger than or equal to a positive constant, c^*.

Assumption 3.2. The elements of $\mathbf{W}_N$ are uniformly bounded and the sequence of spatial weighting matrices $\{\mathbf{W}_N\}_{N\in\mathbb{N}}$ is uniformly bounded in both row sums and column sums.

Assumption 3.3. The link function, $G(\eta)$, is strictly increasing and twice continuously differentiable, for all $\eta \in \mathbb{R}$.

Assumption 3.4. The first derivative of the link function, w.r.t. the index η, $g(\eta) = G'(\eta)$, is a bounded Lipschitz function.

Assumption 3.5. There is a constant ζ such that, $\zeta = \sup_{\alpha\in A}|\alpha| \times \sup_{\eta} g(\eta) \times \sup_N \|\mathbf{W}_N\|_\infty < 1$, where $A \subset \mathbb{R}$ is the compact parameter space of α, $\|\mathbf{W}_N\|_\infty = \max_{i=1,2,\ldots,N} \sum_{j=1}^{n} |w_{ij}|$ is the row sum norm.

Assumption 3.6. The random error, u_i, has zero mean and is independent of x_{ik}, for all $i = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, K$.

Assumption 3.1 establishes that one cannot have an infinite number of spatial units in a bounded space. Therefore, increasing-domain asymptotics is implied, while infill-type asymptotics are ruled out[7]. Assumption 3.2 involve standard boundedness conditions for $\mathbf{W}$. Assumption 3.3 ensures the existence and continuity of the first and second derivative of the link function, $g(\eta)$ and $g'(\eta)$, respectively, for all $\eta \in \mathbb{R}$. Also, $g(\eta)$ is strictly positive, for all $\eta \in \mathbb{R}$. Assumption 3.4 implies that $g'(\eta)$ is bounded, for all $\eta \in \mathbb{R}$, once $g(\eta)$ is bounded and Lipschitz. Assumption 3.5 is related to model stability. Provided that $\mathbf{W}$ is uniformly bounded in both row sums and column sums and that $g(\eta)$ is a bounded function, ζ is finite. For the case where the working spatial weighting matrix is row standardized, $\|\mathbf{W}\|_\infty = 1$, the constant, ζ, verifies $\zeta = \sup_{\alpha\in A}|\alpha| \times \sup_{\eta} g(\eta)$. Thus, the condition on ζ restricts the parameter space of α, in the sense that it is satisfied as long as $|\alpha| < 1/\sup_{\eta} g(\eta)$. Moreover, under Assumption 3.5, the contraction mapping theorem holds. Assumption 3.6 considers a convenient exogenous and homoskedastic setting for the error term.

In most applications with fractional dependent variables, the link function is usually specified as a Probit function and $G(\eta) = \Phi(\eta)$, the standard Normal CDF,

[7]The statistical properties of a Method of Moments estimator, under both increasing-domain asymptotics and infill asymptotics, will be addressed in a simulation study (see Section 3.4 for details).

or as a Logistic function and $G(\eta) = \Lambda(\eta)$, the standard Logistic CDF. Both specifications are suitable for the proposed framework, once they are defined in the open interval $(0, 1)$, and satisfy Assumption 3.3 and Assumption 3.4. Similarly, the spatial weighting matrix can be defined according to a distance-based or a contiguity-based criterion, once the conditions in Assumption 3.2 hold. This is because dependence decreases as distance between two spatial units increases. Also, it cannot be the case that a given spatial unit has infinitely many neighbors. Hence, these criteria ensure that both row sums and column sums of $\mathbf{W}$ are uniformly bounded.

Given the specifications for $G(\eta)$ and $\mathbf{W}$, and considering that the working $\mathbf{W}$ is row normalized, the stability condition, in Assumption 3.5, depends exclusively on the boundedness condition for the first derivative of the link function, $g(\eta)$. Thus, under the Probit specification, $\sup_\eta \phi(\eta) = 1/\sqrt{2\pi}$, where $\phi(\eta)$ is the standard Normal PDF, and, under the Logistic (or Logit) specification, $\sup_\eta \lambda(\eta) = 1/4$, where $\lambda(\eta)$ is the standard Logistic PDF. In this way, the parameter space for α is $\alpha_{\text{Probit}} \in \left(-\sqrt{2\pi}, \sqrt{2\pi}\right)$ and $\alpha_{\text{Logit}} \in (-4, 4)$, respectively.

Note that other specifications for $G(\eta)$ are available, that do not necessarily involve distribution functions (or even known functions). Nevertheless, they must verify the conditions in Assumption 3.3 and Assumption 3.4, to be applicable to the proposed framework.

The assumptions on $G(\eta)$ also have important implications with respect to the identification and interpretation of the partial effects. A strictly increasing and (twice) continuously differentiable link function ensures that the partial effects exist and that their signs remain unchanged for any arbitrary change in η (see also Wooldridge, 2010; Xu and Lee, 2015a). The general partial effect for the kth explanatory variable over the ith spatial unit response is given by:

$$\frac{\partial Y_i}{\partial x_{ik}} = g\left(\alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta}\right) \times \left(\alpha \sum_{j \neq i} w_{ij} \frac{\partial Y_j}{\partial x_{ik}} + \beta_k\right) \tag{58}$$

where $g(\cdot)$ is the first derivative of the function $G(\cdot)$, w.r.t. the index. Note that the partial derivative $\partial u_i / \partial x_{ik} = 0$, for all $k = 1, 2, \ldots, K$, due to Assumption 3.6.

More compactly, in matrix notation:

$$\begin{aligned} \boldsymbol{\Delta}_k &= \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W} \boldsymbol{\Delta}_k + \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \beta_k \\ \Leftrightarrow \boldsymbol{\Delta}_k &= \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W}\right]^{-1} \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \beta_k, \qquad \forall k = 1, 2, \ldots, K \end{aligned} \tag{59}$$

where $\mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})}$ is an $N \times N$ diagonal matrix, whose diagonal elements are $g(\eta_i)$, with $\eta_i = \alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta}$. The $N \times N$ matrix $\left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W}\right]^{-1}$ is the FRSLM link augmented spatial lag operator inverse, with $\mathbf{I}$ the $N \times N$ identity matrix. Note that, under the condition in Assumption 3.5, $\zeta < 1$, hence the matrix $\left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W}\right]$ is nonsingular.

A particular feature of the partial effects, under spatial frameworks, is that a unit change in a given explanatory variable produces different responses, for each spatial unit. In other words, each element of the $N \times N$ matrix of partial effects, for the kth explanatory variable, $\boldsymbol{\Delta}_k$, differs in both rows and columns. Thus, $\boldsymbol{\Delta}_k$ is not symmetric. Element-wise interpretation of the partial effects is possible, but it is not common in most of the applications using spatial data. On this matter, LeSage and Pace (2009) suggest to summarize the partial effects into five measures. The first measure is the Average Direct Effect:

$$\begin{aligned} ADE_k &= \frac{1}{N} \sum_{i=1}^{N} \left\{ \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W}\right]^{-1} \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \right\}_{ii} \beta_k \\ &= \frac{1}{N} \operatorname{tr}\left(\left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W}\right]^{-1} \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \right) \beta_k \end{aligned} \tag{60}$$

the average of the diagonal elements of $\boldsymbol{\Delta}_k$, where $\operatorname{tr}(\cdot)$ is the trace of a matrix. The second measure is the Average Total Effect To an observation:

$$ATET_{k,j} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W}\right]^{-1} \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \right\}_{ij} \beta_k \tag{61}$$

the average of the jth column of $\boldsymbol{\Delta}_k$. The third measure is the Average Total Effect From an observation:

$$ATEF_{k,i} = \frac{1}{N} \sum_{j=1}^{N} \left\{ \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \mathbf{W}\right]^{-1} \mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})} \right\}_{ij} \beta_k \tag{62}$$

the average of the ith row of $\mathbf{\Delta}_k$. The fourth measure is the Average Total Effect:

$$ATE_k = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left[\mathbf{I} - \alpha\,\mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})}\mathbf{W}\right]^{-1}\mathbf{D}_{\mathbf{g}(\boldsymbol{\eta})}\right\}_{ij}\beta_k \tag{63}$$

the average of all elements of $\mathbf{\Delta}_k$. Finally, the fifth measure is the Average Indirect Effect:

$$AIE_k = ATE_k - ADE_k \tag{64}$$

the average of the off-diagonal elements of $\mathbf{\Delta}_k$.

The previous measures for the partial effects can also be evaluated at specific values of the explanatory variables, say means, medians, maximums, minimums or at different quantiles. For example, when evaluating the partial effects at the means, the ith explanatory variables, $x_{i2}, \ldots, x_{iK}$, in (58) are replaced by their corresponding sample averages, $\bar{x}_2, \ldots, \bar{x}_K$.

For the case where policy analysis is concerned, the partial effects deduced in (58) and (59) may be difficult to interpret. This is because, the effects over the responses, due to a unit change in a given explanatory variable, depend on the all values of the exogenous explanatory variables and on endogenous the spatially lagged responses, as well. The notion of policy analysis implies that the endogenous variables can be expressed as a function of the exogenous explanatory variables, usually referred to as the reduced form. In other words, as LeSage and Pace (2010) suggest, it implies that the observed sample of spatial units can be viewed as resulting from a steady-state relationship between the responses and the exogenous explanatory variables. Thus, the effects of policy change can be interpreted as convergence to another steady-state, that only depends (eventually in a nonlinear way) on the values of the exogenous explanatory variables. However, due to nonlinearity and the simultaneous nature of the FRSLM in (56) and (57), a reduced form exists[8], but it

[8] Blundell and Matzkin (2014) and Matzkin (2015) show that, under a nonparametric simultaneous nonadditive framework and considering a set of assumptions analogous to those of Xu and Lee (2015a), a reduced form exists and it is observationally equivalent to the corresponding structural form. In fact, here, the reduced form exists if the model is stable, which is equivalent to say that Assumption 3.5 holds. Nevertheless, even when simple examples with nonlinear simultaneous models are considered, the reduced form expression and the reduced form parameters can be analytically difficult to interpret.

is not analytically tractable, contrary to the linear case.

Next, a new specification is presented, that allows to express the FRSLM in (56) and (57) as an approximate reduced form. This specification addresses the issue related to the tractability of the reduced form under the general nonlinear simultaneous framework presented in this section and, consequently, it allows policy makers to interpret the approximate partial effects as steady-state changes, that are uniquely driven by the exogenous explanatory variables.

3.2.2. The approximate Fractional Response Spatial Lag Model (aFRSLM)

Consider the first-order series expansion of the FRSLM link function $G(\cdot)$ around $\alpha = 0$:

$$G\left(\alpha \sum_{i \neq j} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta}\right) \approx G\left(\mathbf{X}_i \boldsymbol{\beta}\right) + \alpha\, g\left(\mathbf{X}_i \boldsymbol{\beta}\right) \sum_{i \neq j} w_{ij} Y_j \tag{65}$$

then the approximate Fractional Response Spatial Lag Model (aFRSLM) follows as:

$$Y_i \approx G\left(\mathbf{X}_i \boldsymbol{\beta}\right) + \alpha\, g\left(\mathbf{X}_i \boldsymbol{\beta}\right) \sum_{i \neq j} w_{ij} Y_j + u_i, \qquad i = 1, 2, \ldots, N \tag{66}$$

where the index of the link function is now $\mathbf{X}_i \boldsymbol{\beta}$, for each $i = 1, 2, \ldots, N$. Stacking over the spatial units yields:

$$\mathbf{Y} \approx G\left(\mathbf{X}\boldsymbol{\beta}\right) + \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})} \mathbf{W}\mathbf{Y} + \mathbf{u} \Leftrightarrow \mathbf{Y} \approx \left[\mathbf{I} - \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})} \mathbf{W}\right]^{-1} G\left(\mathbf{X}\boldsymbol{\beta}\right) + \mathbf{v} \tag{67}$$

where $\mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})}$ is an $N \times N$ diagonal matrix, whose diagonal elements are $g\left(\mathbf{X}_i \boldsymbol{\beta}\right)$. The new error term is given by $\mathbf{v} = \left[\mathbf{I} - \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})} \mathbf{W}\right]^{-1} \mathbf{u}$, where the $N \times N$ matrix $\left[\mathbf{I} - \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})} \mathbf{W}\right]^{-1}$ is the aFRSLM link augmented spatial lag operator inverse. Under this specification, the fractional responses, $\mathbf{Y}$, depend on every element of the spatial weighting matrix, $\mathbf{W}$, on every element of the matrix of the K exogenous explanatory variables, $\mathbf{X}$, and on every element of the reduced form disturbances, $\mathbf{v}$. In addition, Assumption 3.1 to Assumption 3.6 are maintained.

The link function, $G(\cdot)$, can be specified as a Probit function or a Logit function, analogous to the FRSLM. In the same manner, the stability condition, $\zeta < 1$, from Assumption 3.5, remains valid.

However, it is important to note that, under the setting of the aFRSLM, the predicted values, $\hat{Y}_i$, do not necessarily lie in the open interval $(0, 1)$. This is because the ith row of the matrix product $\left[\mathbf{I} - \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]^{-1} G(\mathbf{X}\boldsymbol{\beta})$ is bounded between $(1+\zeta)^{-1}$ and $(1-\zeta)^{-1}$, for all $i = 1, 2, \ldots, N$ (see Proposition 9.4.13. of Bernstein, 2009). For values of $\zeta > 0.5$, it can be especially problematic, once first order approximations generally start to diverge from the true values[9].

As previously mentioned there are two major advantages related to the aFRSLM specification. First, it allows to express the FRSLM as an approximate reduced form:

$$\mathrm{E}(\mathbf{Y} \mid \mathbf{X}) \approx \left[\mathbf{I} - \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]^{-1} G(\mathbf{X}\boldsymbol{\beta}) \tag{68}$$

where the right hand side of the equation is a tractable and estimable nonlinear function of the parameters. Note that, the conditional expectation $\mathrm{E}(\mathbf{v} \mid \mathbf{X}) = \mathbf{0}$ due to Assumption 3.6. In addition, the aFRSLM can be expressed as a sum of nonlinear functions of the exogenous explanatory variables and their spatial lags. To see this, consider the series expansion of the inverse:

$$\begin{aligned}\left[\mathbf{I} - \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]^{-1} &= \mathbf{I} + \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W} + \alpha^2\, \mathbf{D}^2_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}^2 + \alpha^3\, \mathbf{D}^3_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}^3 + \ldots \\ &= \sum_{h=0}^{\infty} \alpha^h\, \mathbf{D}^h_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}^h\end{aligned} \tag{69}$$

which converges absolutely for $\zeta < 1$. Hence, the conditional expectation in (68) can be written as:

$$\begin{aligned}\mathrm{E}(\mathbf{Y} \mid \mathbf{X}, \mathbf{W}) \approx\; & G(\mathbf{X}\boldsymbol{\beta}) + \alpha\, \mathbf{D}_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W} G(\mathbf{X}\boldsymbol{\beta}) \\ & + \alpha^2\, \mathbf{D}^2_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}^2 G(\mathbf{X}\boldsymbol{\beta}) + \alpha^3\, \mathbf{D}^3_{g(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}^3 G(\mathbf{X}\boldsymbol{\beta}) + \ldots\end{aligned} \tag{70}$$

[9] See Klier and McMillen (2008) for further discussion on the properties of a GMM estimator for a linearized version of a model for a spatially lagged binary dependent variable, based on a spatially lagged latent variable framework. See also Chapter 2 for further discussion on the accuracy of an approximation of the spatial lag operator inverse.

Second, using the conditional expectation in equation (68) allows to interpret the partial effects as changes in the steady-state relationship between the responses and the exogenous explanatory variables. The matrix of the approximated partial effects is given by:

$$\begin{aligned} \boldsymbol{\Delta}_k \approx & \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]^{-1} \mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})}\beta_k \\ & + \alpha \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]^{-1} \mathbf{D}_{\mathbf{g}'(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\, \mathbf{G}^*(\mathbf{X}\boldsymbol{\beta})\, \beta_k \\ \Leftrightarrow \boldsymbol{\Delta}_k \approx & \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]^{-1} \\ & \times \left[\mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})} + \alpha\, \mathbf{D}_{\mathbf{g}'(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\, \mathbf{G}^*(\mathbf{X}\boldsymbol{\beta})\right] \beta_k, \quad \forall k = 1, 2, \ldots, K \end{aligned} \tag{71}$$

where $\mathbf{G}^*(\mathbf{X}\boldsymbol{\beta}) = \left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]^{-1} \mathbf{G}\left(\mathbf{X}\boldsymbol{\beta}\right)$ and $\mathbf{D}_{\mathbf{g}'(\mathbf{X}\boldsymbol{\beta})}$ is an $N \times N$ diagonal matrix, whose diagonal elements are $g'(\mathbf{X}_i\boldsymbol{\beta})$. The matrix $\left[\mathbf{I} - \alpha\, \mathbf{D}_{\mathbf{g}(\mathbf{X}\boldsymbol{\beta})}\mathbf{W}\right]$ is non-singular, once the condition $\zeta < 1$ holds.

The partial effects in (71) can be interpreted as effects of policy changes (see also LeSage and Pace, 2010) and summarized according to the five measures proposed by LeSage and Pace (2009), the ADE, ATET, ATEF, ATE and AIE, deduced in the previous section.

3.3. GMM estimation

The estimation of the proposed models for spatially lagged fractional responses considers the Generalized Method of Moments (GMM) approach presented on Chapter 2, based on the works of Pinkse and Slade (1998) and Klier and McMillen (2008). Under the setting of the GMM, it is assumed that the unknown parameter vector $\boldsymbol{\Theta} = (\alpha, \boldsymbol{\beta})$ satisfy the following moment condition:

$$\mathrm{E}\left(\mathbf{Z}^\intercal \mathbf{u}\right) = \mathbf{0} \tag{72}$$

where $\mathbf{Z} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}]$, as suggested by Kelejian and Prucha (1998). The GMM estimates for the unknown parameter vector, $\boldsymbol{\Theta}$, are obtained by minimizing the objective function:

$$\mathcal{Q}\left(\boldsymbol{\Theta}\right) = \mathbf{u}^\intercal \mathbf{Z} \left(\mathbf{Z}^\intercal \mathbf{Z}\right)^{-1} \mathbf{Z}^\intercal \mathbf{u} \tag{73}$$

and the GMM estimator reduces to nonlinear two stages least squares (N2SLS). As the minimization problem in (73) does not have a closed formula, the iterative procedure of Klier and McMillen (2008) is used. In addition, the spatial heteroskedasticity and spatial autocorrelation robust estimator of Kelejian and Prucha (2007) is considered, to overcome potential biases in the estimated asymptotic covariance matrix estimator of the (iterative) GMM estimator. See Section 2.2 in Chapter 2 for details.

The individual gradients for the FRSLM are:

$$(\mathbf{\Gamma}_{\boldsymbol{\beta}})_i = \frac{\partial u_i}{\partial \boldsymbol{\beta}^{\intercal}} = -\, g\left(\alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta}\right) \mathbf{X}_i, \qquad i = 1, 2, \ldots, N \tag{74}$$

and

$$(\mathbf{\Gamma}_{\alpha})_i = \frac{\partial u_i}{\partial \alpha} = -\, g\left(\alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i \boldsymbol{\beta}\right) \sum_{j \neq i} w_{ij} Y_j, \qquad i = 1, 2, \ldots, N \tag{75}$$

where $g(\cdot)$ is the first derivative of the function $G(\cdot)$, w.r.t. the index.

The individual gradients for the aFRSLM are:

$$\begin{aligned} (\mathbf{\Gamma}_{\boldsymbol{\beta}})_i = \frac{\partial u_i}{\partial \boldsymbol{\beta}^{\intercal}} = &-\, g\left(\mathbf{X}_i \boldsymbol{\beta}\right) \mathbf{X}_i \\ &-\, \alpha\, g'\left(\mathbf{X}_i \boldsymbol{\beta}\right) \mathbf{X}_i \left(\sum_{i \neq j} w_{ij} Y_j\right), \qquad i = 1, 2, \ldots, N \end{aligned} \tag{76}$$

and

$$(\mathbf{\Gamma}_{\alpha})_i = \frac{\partial u_i}{\partial \alpha} = -\, g\left(\mathbf{X}_i \boldsymbol{\beta}\right) \left(\sum_{i \neq j} w_{ij} Y_j\right), \qquad i = 1, 2, \ldots, N \tag{77}$$

where $g'(\cdot)$ is the first derivative of the function $g(\cdot)$, w.r.t. the index.

Next, the results for an extensive Monte Carlo simulation study are presented and discussed. The iterative GMM (iGMM) estimator, presented in this section, is used to estimate the FRSLM and the aFRSLM for a set of experiments. Its statistical properties, are assessed in twofold. One, in terms of the accuracy and precision to estimate the true values for the parameters of interest. Two, in terms of the performance to estimate the direct effects and the indirect effects. In addition, they are compared with the statistical properties of the Two-Stage Least Squares

(2SLS) estimator, for the Linear Spatial Lag Model (LSLM).

3.4. Monte Carlo simulations

In this section, an extensive Monte Carlo simulation study is conducted. The structure of the simulated fractional dependent variable is presented, for each set of experiments. The statistical properties of the iterative GMM (iGMM) estimator for the proposed approaches to model spatially lagged fractional responses – the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM) – are addressed and compared in detail. The statistical properties of the iGMM estimator are also compared with those from the Two-Stage Least Squares (2SLS) estimator, for the Linear Spatial Lag Model (LSLM). The adequacy of the linear approach to model fractional responses is assessed and discussed. The estimates obtained from the iGMM and 2SLS estimation of the corresponding models are also used to estimate two measures for the partial effects: the Average Direct Effect (ADE) and the Average Indirect Effect (AIE). Their statistical properties are addressed and compared in detail, as well. All the estimates are summarized in terms of bias and root mean squared error, considering a large variety of sampling designs.

3.4.1. Simulation design

The design for the simulation study is based on a simplified version of the FRSLM in (56), with a single explanatory variable. The link function is assumed to be the Probit function, $\Phi(\cdot)$, and the design for the explanatory variable follows along the lines of Xu and Lee (2015a), $\mathbf{x} \sim 1.5\,(\mathbf{I} - 0.2\mathbf{W})^{-1}\mathcal{N}(\mathbf{0}, \mathbf{I})$, to induce spatial correlation. The generation of the fractional dependent variable considers the following model:

$$Y_i = \mathcal{B}\varepsilon^{-1}(\nu_i; \mu_{Y_i}, \psi_Y), \qquad i = 1, 2, \ldots, N \tag{78}$$

where $\mathcal{B}\varepsilon^{-1}(\cdot)$ is the quantile function of the Beta distribution, with parameters μ_{Y_i} and ψ_Y. The parameter μ_{Y_i} is the simulated average of Y_i, differing across i,

and the parameter ψ_Y is a fixed dispersion parameter, for all i.[10] The scalar ν_i is a randomly drawn quantile from a $\mathcal{U}(0,1)$ distribution. The mean of Y_i conditional on the regressors is given by:

$$\mu_{Y_i} = \Phi\left(\alpha \sum_{j \neq i} w_{ij} Y_j + \beta_0 + \beta_1 x_i\right), \qquad i = 1, 2, \ldots, N \tag{79}$$

and the variance of Y_i conditional on the regressors is given by:

$$\sigma^2_{Y_i} = \frac{\mu_{Y_i}(1 - \mu_{Y_i})}{\psi_Y + 1}, \qquad i = 1, 2, \ldots, N \tag{80}$$

Due to the simultaneous nature of the FRSLM, the fractional dependent variable, Y_i, can only be obtained iteratively. Xu and Lee (2015a) show that an iterative procedure converges, once the conditions of the contraction mapping theorem hold, under the assumptions presented in Section 3.2.1. Here, the following iterative procedure is considered:

1. For each i, given the probability ν_i, generate $Y_i^{(0)}$ from $\mathcal{B}\varepsilon^{-1}\left(\nu_i; \mu_{Y_i}^{(0)}, \psi_Y\right)$, with $\mu_{Y_i}^{(0)} = \Phi(\beta_0 + x_i\beta_1)$.

2. Construct the next iteration, $Y_i^{(1)}$, from $\mathcal{B}\varepsilon^{-1}\left(\nu_i; \mu_{Y_i}^{(1)}, \psi_Y\right)$, with

$$\mu_{Y_i}^{(1)} = \Phi\left(\alpha \sum_{j \neq i} w_{ij} Y_j^{(0)} + \beta_0 + x_i\beta_1\right)$$

   and using the same probability, ν_i.

3. Repeat step 2., using the values of Y_i from the last iteration, until $\left|Y_i^{(s)} - Y_i^{(s-1)}\right| < 10^{-8}$, for all i.

Note that, even though $\mathcal{B}\varepsilon^{-1}(\nu_i; \mu_{Y_i}, \psi_Y)$ is defined on the open interval $(0,1)$, it is possible that, numerically, for some i, Y_i can take on the values zero and one, due

[10] See also Paolino (2001), Ferrari and Cribari-Neto (2004) and Ramalho et al. (2011) for applications with non-spatial data, using the mean-dispersion parameterization of the Beta distribution

to rounding error of the machine. This is especially meaningful for the case where ψ_Y is close to zero[11].

The spatial weighting matrix, $\mathbf{W}$, is constructed according to the two stage setting described in Section 2.6.1 of Chapter 2. Firstly, the N spatial units are randomly drawn points in the unit square. Secondly, an initial spatial weighting matrix, $\mathbf{W}_0$, is constructed based on a nearest neighbor criterion; the working spatial weighting matrix, $\mathbf{W}$, is the row standardized $\mathbf{W}_0$. The number of nearest neighbors is given by $\delta \times N$, where δ is the matrix density (the complement of sparsity), the proportion of non-zero elements in $\mathbf{W}$. In this way, the large sample properties of both the iGMM estimator for the proposed models and the 2SLS estimator for the LSLM can be addressed according to the frameworks of increasing-domain asymptotics (fixed δ, increasing N) and of infill asymptotics (increasing δ, fixed N).

The sample sizes $N = 100$, $N = 500$ and $N = 1000$ are considered. The matrix density parameter, δ, varies over the set $\{0.01, 0.1, 0.2\}$. The regression parameters are held fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial lag parameter, α, is designed to vary over the set $\{0, 1, 1.5, 2\}$, once these values satisfy the condition $\zeta < 1$ in Assumption 3.5, provided that $\sup_\eta \phi(\eta) = 1/\sqrt{2\pi}$ and $\sup_n \|\mathbf{W}\|_\infty = 1$. The dispersion parameter, ψ_Y, varies over the set $\{0.1, 1, 10\}$. The matrix of instruments used in the estimation is given by $\mathbf{Z} = [\mathbf{X} \ \mathbf{WX} \ \mathbf{W}^2\mathbf{X}]$. For each experiment, 1000 replications are used. The experiments were performed in a Linux based server, with 64 GB of RAM and composed by 24 AMD Opteron CPUs, ranging from 0.8 GHz to 2.1 GHz.

For each set of experiments, the sampling distribution of the vector of simulated responses, $\mathbf{Y}$, is reported. The parameter estimates, $\hat{\alpha}$, $\hat{\beta}_0$ and $\hat{\beta}_1$, are reported, as well, for the FRSLM, the aFRSLM and the LSLM. They are summarized by the mean and the root mean squared error (RMSE). The estimates for both the FRSLM and aFRSLM are obtained from the iGMM estimator, presented in Section 3.3, and the estimates for the LSLM are obtained from the 2SLS estimator. The 2SLS parameter estimates are of particular interest, as the adequacy of considering a linear

[11] As $\psi_Y \to 0$, the conditional variance of Y_i converges to $\mu_{Y_i}(1 - \mu_{Y_i})$, which is equal to the conditional variance of a dichotomous $\{0, 1\}$ dependent variable.

approach to estimate models for fractional responses can be assessed. The estimates for the Average Direct Effect (ADE) and the Average Indirect Effect (AIE) are also reported for the FRSLM, the aFRSLM and the LSLM. They are summarized by the mean absolute bias (MAB) and the root mean squared error (RMSE). The true values for the ADE and the AIE are obtained from the matrix of partial effects of the FRSLM – see equation (59) –, evaluated at the true values of the parameters α, β_0 and β_1. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$, to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM and their corresponding partial effects. All the calculations were performed using R.

3.4.2. Results

The results of the Monte Carlo experiments are presented in Section A3. The sampling distributions of the vector of simulated responses, $\mathbf{Y}$, are shown in Section A3.1. The simulation results on the statistical properties of the iterative GMM (iGMM) estimator and the Two-Stage Least Squares (2SLS) estimator are detailed in Section A3.2. The simulation results on the statistical properties of the Average Direct Effect (ADE) and of the Average Indirect Effect (AIE), are detailed in Section A3.3. The simulation results are organized according to the values of the dispersion parameter, ψ_Y, and the true values of the spatial lag parameter, α. Also, they are organized according to usefulness of the model specification: first, the starting point, the Linear Spatial Lag Model (LSLM), second, the true model, the Fractional Response Spatial Lag Model (FRSLM), third, the proposed alternative, the approximate Fractional Response Spatial Lag Model (aFRSLM).

The pattern of the sampling distributions of the vector of simulated responses, $\mathbf{Y}$, tends to be an "U"-shaped curve. This is because the exogenous explanatory variable is designed in such a way that persistence (in the form of spatial correlation) is induced in the simulated responses. Hence, even for the case where the $\alpha = 0$, there is a source of spatial dependence in $\mathbf{Y}$. Even so, the pattern of the sampling distributions changes progressively to a reverse "J"-shaped (right-skewed) curve, as ψ_Y increases ($\sigma^2_{Y_i}$ decreases) and $\alpha \to 0$. In fact, as ψ_Y increases, the density of

observations at the corners zero and one decreases, with the decreasing rate being particularly fast at the corner one, as $\alpha \to 0$. The constant term (held fixed at $\beta_0 = -1$) plays an important role on this phenomenon. Under the current simulation design, the negative sign of β_0 implies that the mean of Y_i conditional on the regressor x_i – see equation (79) – tends to be skewed toward zero. This effect becomes larger, as $\alpha \to 0$, once the magnitude of β_0 is greater than the magnitude of the term $\beta_1 x_i$, for almost all i.

With regard to the statistical properties of the estimators, results are consistent with similar simulation studies (see Xu and Lee, 2015a, and Chapter 2). They can be summarized in fourfold. One, the 2SLS estimator is substantially more biased than the iGMM estimator (for both the FRSLM and the aFRSLM). Two, the iGMM estimator for the FRSLM tends to be the least biased. Three, as ψ_Y decreases ($\sigma^2_{Y_i}$ increases), the estimates for the parameters of interest tend to be more biased. Four, the true value of α does not seem to have a significant effect on the accuracy of estimators. One important exception is the iGMM estimator for the aFRSLM, when $\alpha \geq 1.5$. In this case, the estimates for β_0 and β_1 become more biased. However, this effect is mitigated in the estimation of the ADE and the AIE.

For the case where N is fixed and δ is increasing (the spatial weighting matrix, $\mathbf{W}$, becomes dense), the estimates for α and β_0 tend to display an increasing bias, whereas the estimates for β_1 tend to display a decreasing bias. Two other results are of particular importance. First, for the case where $\alpha = 0$ and as δ increases, both the iGMM estimator and the 2SLS estimator tend to estimate spurious spatial dependence, i.e., $\hat{\alpha} > 0$. Such distortions were already mentioned in Chapter 2 and are discussed in detail by Lahiri (1996) and, more recently, by Lee (2004). Second, as $\mathbf{W}$ becomes dense, the iGMM estimator generally performs better for the approximated model than for the true model, when estimating $\hat{\alpha}$. The bias of the aFRSLM estimates for α tends to decrease as δ goes from 0.01 to 0.1. Similarly, the bias of the aFRSLM estimates for $\hat{\alpha}$ tends to decrease as δ goes from 0.1 to 0.2, but only for the case where the true value of $\alpha = 2$. These results are consistent with the findings from Chapter 2.

For the case where δ is fixed and N is increasing, the bias of the estimates for

the parameters of interest tends to decrease, especially when $\delta = 0.01$. For the case where $\delta \geq 0.1$, the bias of the estimates for β_1 also tends to decrease, but the bias of the estimates for α and β_0 exhibit a contradicting behavior. In some cases, the bias increases, while in other cases the bias decreases. This can be explained by the effect of increasing-domain and infill asymptotics operating together. As a result, the rate of convergence for the various parameters of interest can be different and possibly slower than $\sqrt{N}$ (see Lee, 2004, , also noted in Chapter 2).

In terms of the RMSEs of the estimated parameters, results show that the iGMM estimator is more precise than the 2SLS estimator. There are, however, two exceptions. One, the case where $\psi_Y = 1$, $\delta \geq 0.1$ and the true value of $\alpha \leq 1.5$. Two, the case where $\psi_Y = 10$, $\delta \geq 0.1$ and the true value of $\alpha = 0$. Here, the 2SLS estimator is more precise than the iGMM estimator, when estimating $\hat{\alpha}$. Results also show that the iGMM estimator for the FRSLM tends to produce smaller RMSEs for both $\hat{\alpha}$ and $\hat{\beta}_1$, than the iGMM estimator for the aFRSLM. To the contrary, the iGMM estimator for the FRSLM tends to produce higher RMSEs for $\hat{\beta}_0$, than the iGMM estimator for the aFRSLM.

For the case where N is fixed and δ is increasing, the RMSEs largely increase, especially for $\hat{\alpha}$. This draws further attention to the estimation issues under infill asymptotics. For the case where δ is fixed and N is increasing, the RMSEs steadily decrease and the decreasing rate lowers substantially as δ increases. Once again, this is a consequence of the combined effect of increasing-domain and infill asymptotics.

Now, focusing on the statistical properties of the estimated partial effects, the results can be summarized in fourfold. One, the iGMM estimator for the FRSLM tends to provide the most accurate estimates for the ADE, while the iGMM estimator for the aFRSLM tends to provide the most accurate estimates for the AIE. Two, the 2SLS estimator provides the least accurate estimates for both the ADE and AIE. Three, the iGMM estimator for the FRSLM displays a significant amount of small sample bias for the ADE and the AIE, especially when $\delta = 0.01$. Nevertheless, such bias tends to decrease significantly, as both N and δ increase. Four, the dispersion parameter, ψ_Y, and the true value of α do not seem to have a significant effect on the accuracy of the iGMM estimates for both the ADE and the AIE.

Conversely, the 2SLS estimates for the ADE and the AIE become severely biased, as the true value of α increases.

For the case where N is fixed and δ is increasing, the estimates for the AIE tend to display an increasing bias, whereas the estimates for the ADE exhibit an heterogeneous pattern. Under the FRSLM, the iGMM estimates for the ADE do not change much. However, under the aFRSLM, the bias tends to increase, as δ is increasing. Surprisingly, the 2SLS estimates for the ADE tend to display a decreasing bias. This occurs mainly due to the fact that the bias at $\delta = 0.01$ is, in some cases, extremely large. As a consequence, this result must be viewed with caution, once the 2SLS estimates for the ADE (but also for the AIE) tend to be fairly more biased than the iGMM estimates for both the FRSLM and aFRSLM.

For the case where δ is fixed and N is increasing, both the estimates for the ADE and the AIE tend to display a decreasing bias. There are, however, two cases where the bias for $\widehat{AIE}$ significantly increases. First, for the iGMM estimator for the FRSLM, when $\delta = 0.2$ and as N goes from 500 to 1000. Second, for the 2SLS estimator, when $\psi_Y \geq 1$ and the true value of $\alpha = 2$.

In terms of the RMSEs of the estimated partial effects, they tend to be substantially large for the 2SLS estimator, while being small for the iGMM estimator. In fact, the most precise estimates for the ADE are those obtained from the iGMM estimation of the FRSLM and the most precise estimates for the AIE are those obtained from the iGMM estimation of the aFRSLM. Results also show that the inequality $\text{RMSE}\left(\widehat{ADE}\right) < \text{RMSE}\left(\widehat{AIE}\right)$ holds for all estimators.

For the case where N is fixed and δ is increasing, the RMSEs for the AIE tend to increase. This issue becomes more severe for the 2SLS estimator, as ψ_Y increases, whereas it becomes more severe for the iGMM estimator, as ψ_Y decreases. With regard to the RMSEs for the ADE, they tend to strictly decrease, for the 2SLS estimator, but tend to decrease as δ goes from 0.01 to 0.1 and increase as δ goes from 0.1 to 0.2, for the iGMM estimator.

For the case where δ is fixed and N is increasing, the RMSEs for the ADE and the RMSEs for the AIE tend to decrease. Exceptions are the iGMM estimator for the FRSLM ($\delta = 0.2$ and as N goes from 500 to 1000) and the 2SLS estimator

($\psi_Y \geq 1$ and the true value of $\alpha = 2$). Here, the RMSEs for the AIE tend to increase (as well as the corresponding bias). An important remark is related to the increasing precision of the iGMM estimator for the aFRSLM, when $\delta = 0.01$. In this case, the RMSEs for both the ADE and the AIE significantly decrease.

The iGMM estimator for the aFRSLM proves to be especially useful to estimate the AIE, as it performs better than the iGMM estimator for the FRSLM, in terms of accuracy and precision, for the majority of the simulations. The LSLM proves to be quite misleading to estimate both the ADE and AIE, as expected.

3.5. Conclusions

In this chapter two new specifications to model fractional responses with spatial dependence were proposed. No transformations are applied to the responses, hence observations at the boundaries, zero and one, can be handled. The setup for the proposed specifications rely on a set of assumptions that are commonly used in the literature. Most of these assumptions were presented and discussed in detail by Xu and Lee (2015a).

The first specification, the Fractional Response Spatial Lag Model (FRSLM), extends the approach of Papke and Wooldridge (1996) to spatial frameworks and generalizes the approach of Xu and Lee (2015a) to accommodate responses defined in the closed interval $[0, 1]$. The FRSLM considers the attractive functional forms developed by Papke and Wooldridge (1996), while introducing spatial dependence into their specification through a spatial lag of the fractional dependent variable. In this way, the corresponding conditional expectation is modeled directly through a nonlinear function (with possibly known functional form) that includes a set of exogenous explanatory variables and a spatial lag of the fractional dependent variable. Moreover, the errors enter the model expression in an additive way, as opposed to Xu and Lee (2015a). The partial effects are deduced and summarized according to the five measures suggested by LeSage and Pace (2009) – the Average Direct Effect (ADE), the Average Total Effect To an observation (ATET), the Average Total Effect From an observation (ATEF), the Average Total Effect (ATE) and the Average

Indirect Effect (AIE). However, their interpretation can be troublesome, especially if policy analysis is concerned. This is because the effects of exogenous changes do not uniquely depend on the values of the exogenous explanatory variables.

The second specification, the approximate Fractional Response Spatial Lag Model (aFRSLM), consists in a first order series approximation of the nonlinear function from the FRSLM, around the spatial lag parameter equal to zero. Under this approach, the FRSLM can be written as an approximate reduced form, with a tractable analytic formula, and the true partial effects can be approximated by sums of nonlinear functions of the exogenous explanatory variables and their spatially lagged values. This approach is particularly useful once the reduced form expression for nonlinear simultaneous models, such as the FRSLM, has no known or tractable analytic formula and can only be obtained through complex numerical methods. In addition, it allows to use the approximate partial effects for policy analysis and to interpret them as approximate measures for policy changes. The policy makers do not require prior knowledge of the neighbors responses, to control for exogenous unit-specific and/or neighbor-specific changes.

The iterative Generalized Method of Moments (iGMM) estimator of Klier and McMillen (2008) is used to estimate both the FRSLM and the aFRSLM. In addition, the Two-Stages Least Squares (2SLS) estimator is also considered to estimate the Linear Spatial Lag Model (LSLM), a popular starting point for analyzing nonlinear models. The statistical properties of both the iGMM estimator and the 2SLS estimator are assessed through an extensive Monte Carlo simulation study, where the distribution of the simulated data is typically an U-shaped curve. The simulation results are consistent with the existing literature.

The adequacy of iGMM estimator for both the FRSLM and the aFRSLM is shown, regarding the estimation of the unknown parameter vector and the corresponding partial effects, namely the ADE and the AIE. Increasing the variability of the responses tends to produce more biased estimates, but this effect appears to be mitigated in the estimation of the ADE and the AIE. The iGMM estimates for the FRSLM are, generally, the most accurate and precise, for both the unknown parameter vector and the relevant partial effects. Even so, the iGMM estimator for

the aFRSLM tends to perform better than the iGMM estimator for the FRSLM, when estimating the spatial lag parameter for sampling designs with high levels of spatial dependence and denser spatial weighting matrices, but it tends to be more biased when estimating the regression parameters. Nevertheless, the iGMM estimator for the aFRSLM proves to be superior to the iGMM estimator for the FRSLM, when estimating the AIE. The limitations of the 2SLS estimator for the LSLM were pointed out.

The proposed specifications admit several extensions. First, it would be interesting to introduce endogeneity through the explanatory variables and/or the spatial weighting matrix. This will raise additional complications on the functional form of the link function, that has to be properly accommodated into the GMM estimation, to ensure consistency. Second, it would be interesting to add higher-order spatial lags to the FRSLM specification. This will affect the performance of the iGMM estimator for the aFRSLM and it may be useful to develop alternative ways to obtain a tractable expression for the reduced form. Third, it would be interesting to generalize the proposed specifications to spatial panel data frameworks. The possibility of incorporating temporal and spatio-temporal effects is still widely unexplored. In addition, there are very few works that address the performance the corresponding spatial panel estimators, for different treatments of the unobserved heterogeneity. Fourth, it would be interesting to generalize the proposed specifications to more flexible semiparametric or nonparametric approaches, where the assumptions presented in Section 3.2.1 may not hold.

All the algorithms and estimation procedures used in this chapter will be made available in an R package.

APPENDIXES

A3. Simulation results

A3.1. Histograms for the simulated dependent variable

TABLE A3.1.1: Data structure for the simulated dependent variable, with $\alpha = 0$ and $\psi_Y = 0.1$

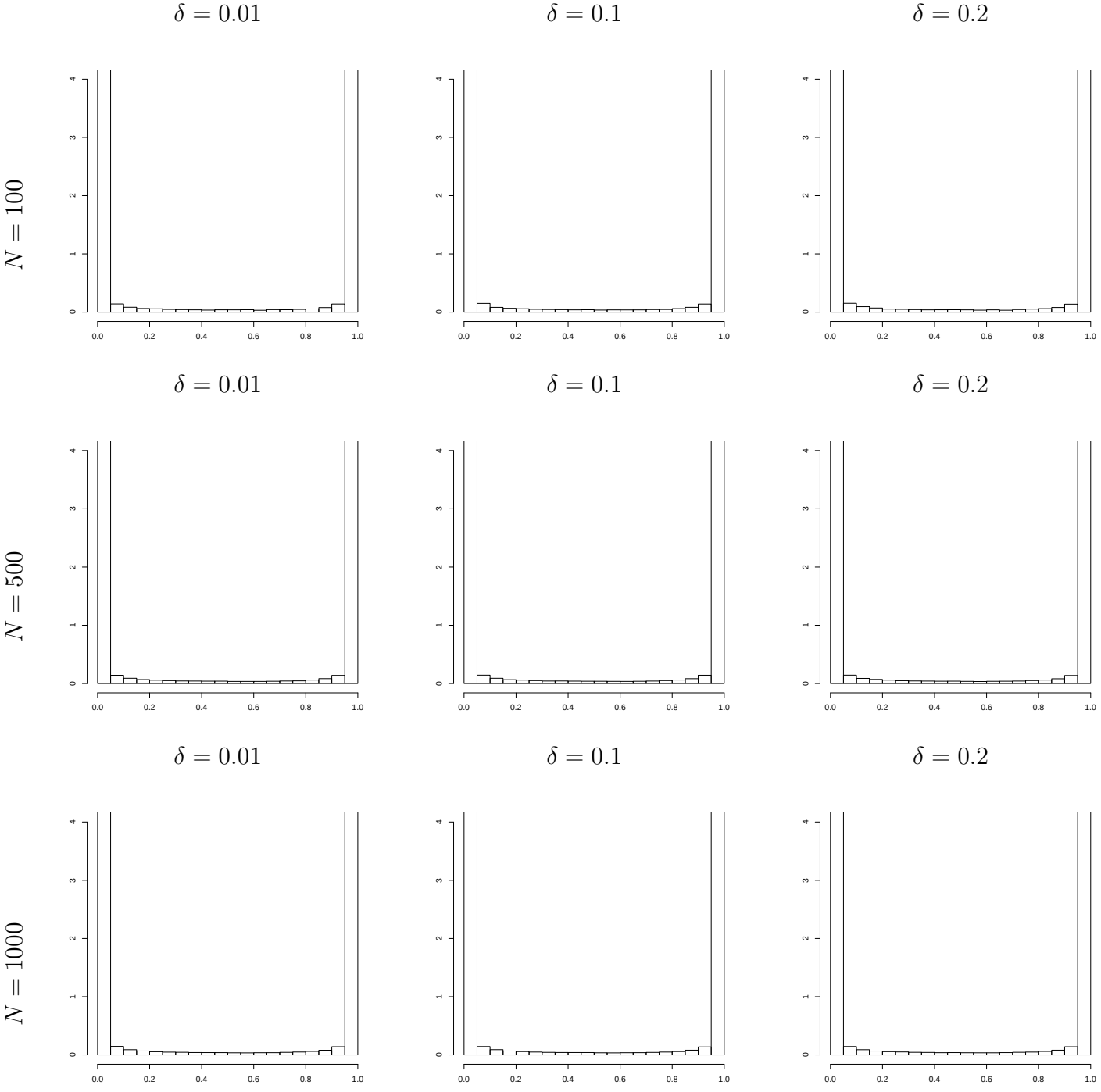


TABLE A3.1.2: Data structure for the simulated dependent variable, with $\alpha = 1$ and $\psi_Y = 0.1$

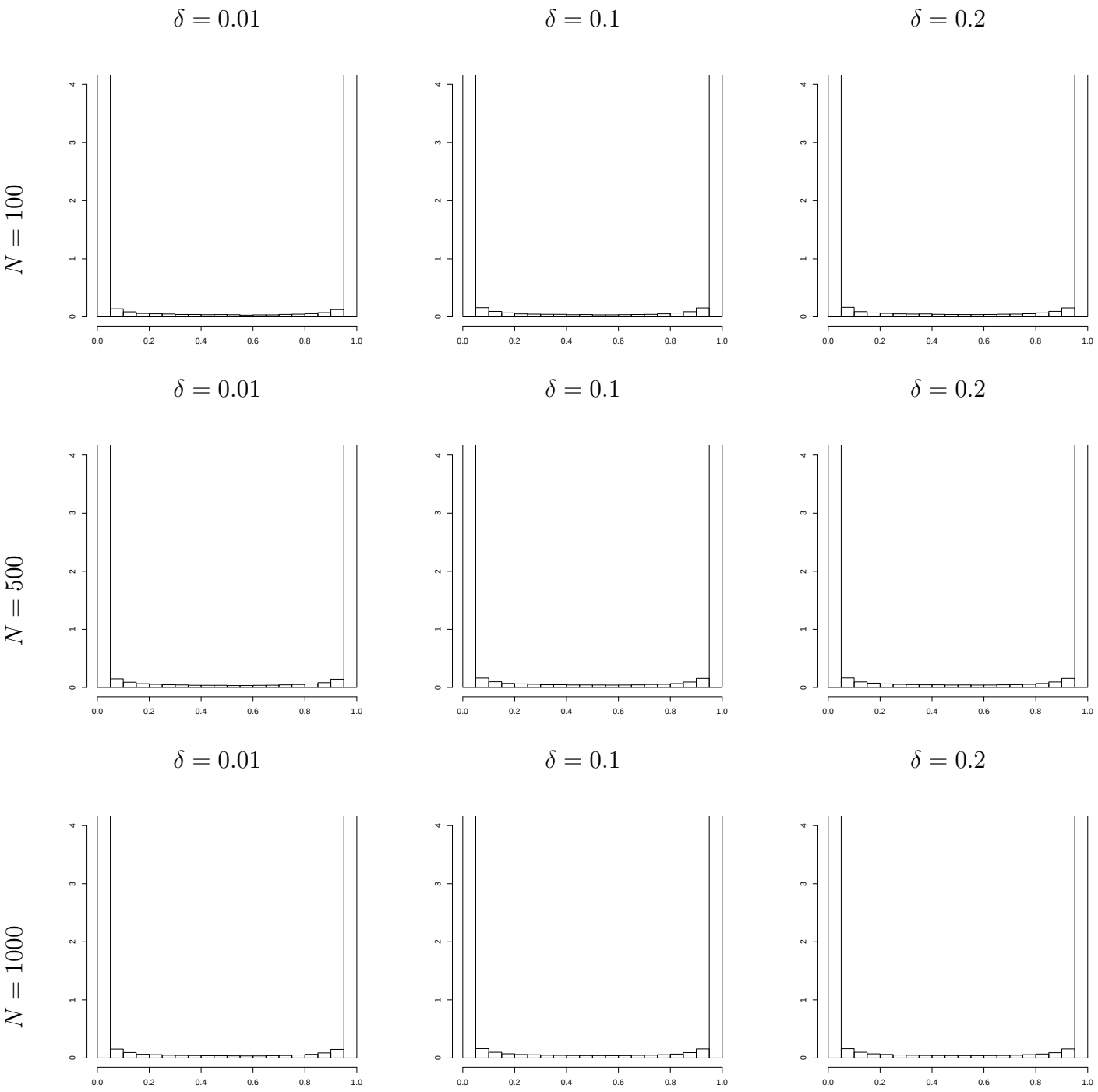


TABLE A3.1.3: Data structure for the simulated dependent variable, with $\alpha = 1.5$ and $\psi_Y = 0.1$

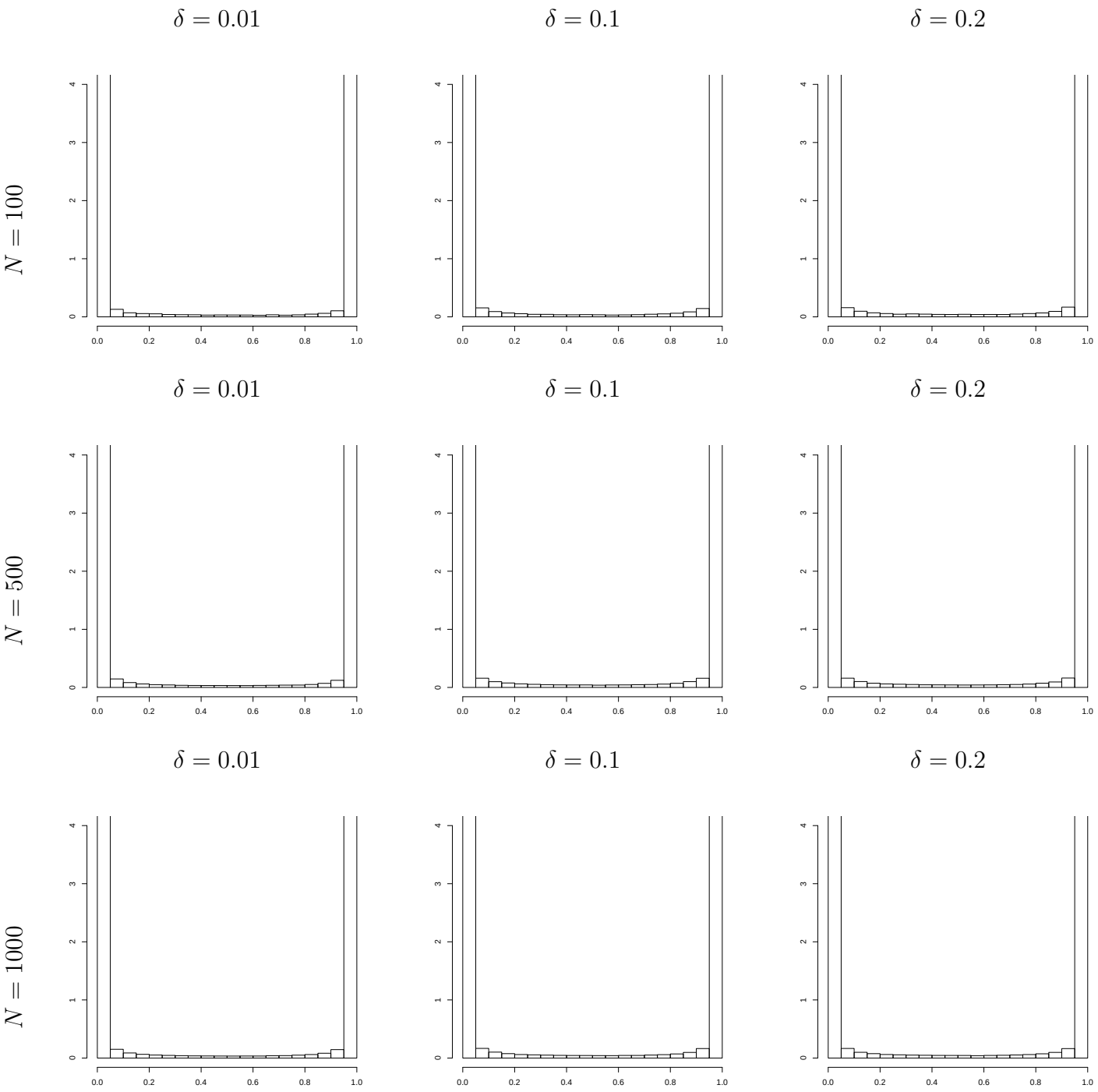


TABLE A3.1.4: Data structure for the simulated dependent variable, with $\alpha = 2$ and $\psi_Y = 0.1$

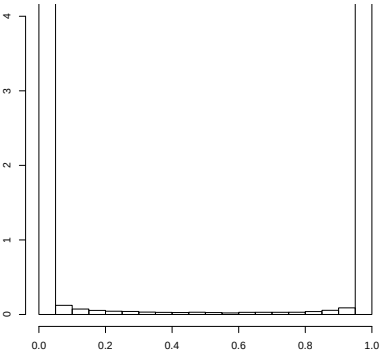
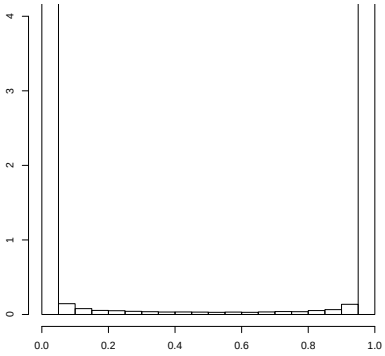
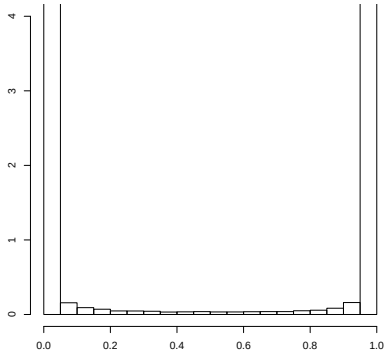
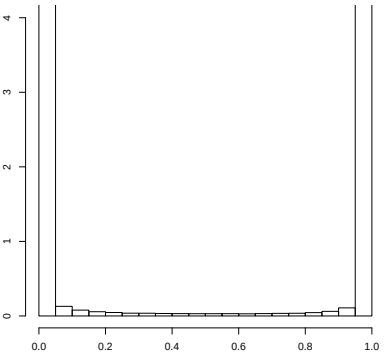
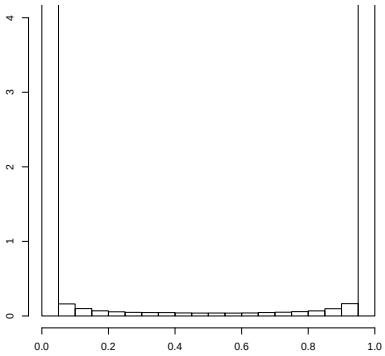
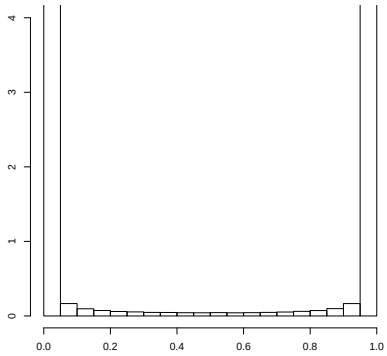
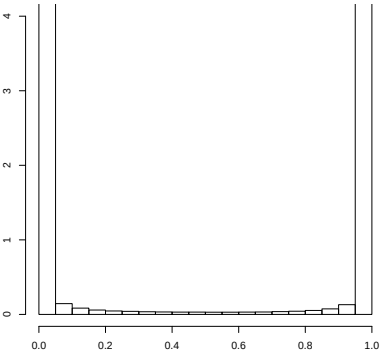
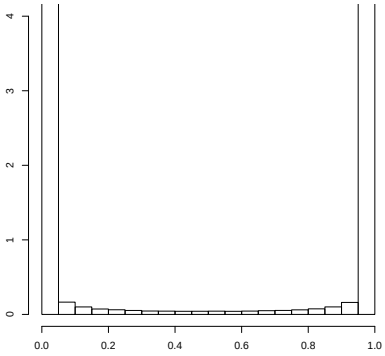
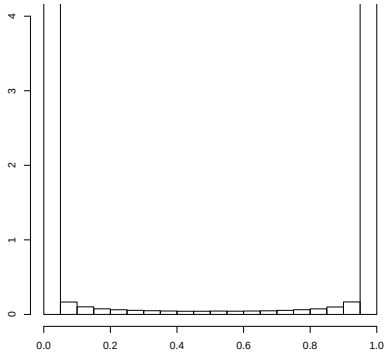
	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.2$
$N = 100$			
$N = 500$			
$N = 1000$			

TABLE A3.1.5: Data structure for the simulated dependent variable, with $\alpha = 0$ and $\psi_Y = 1$

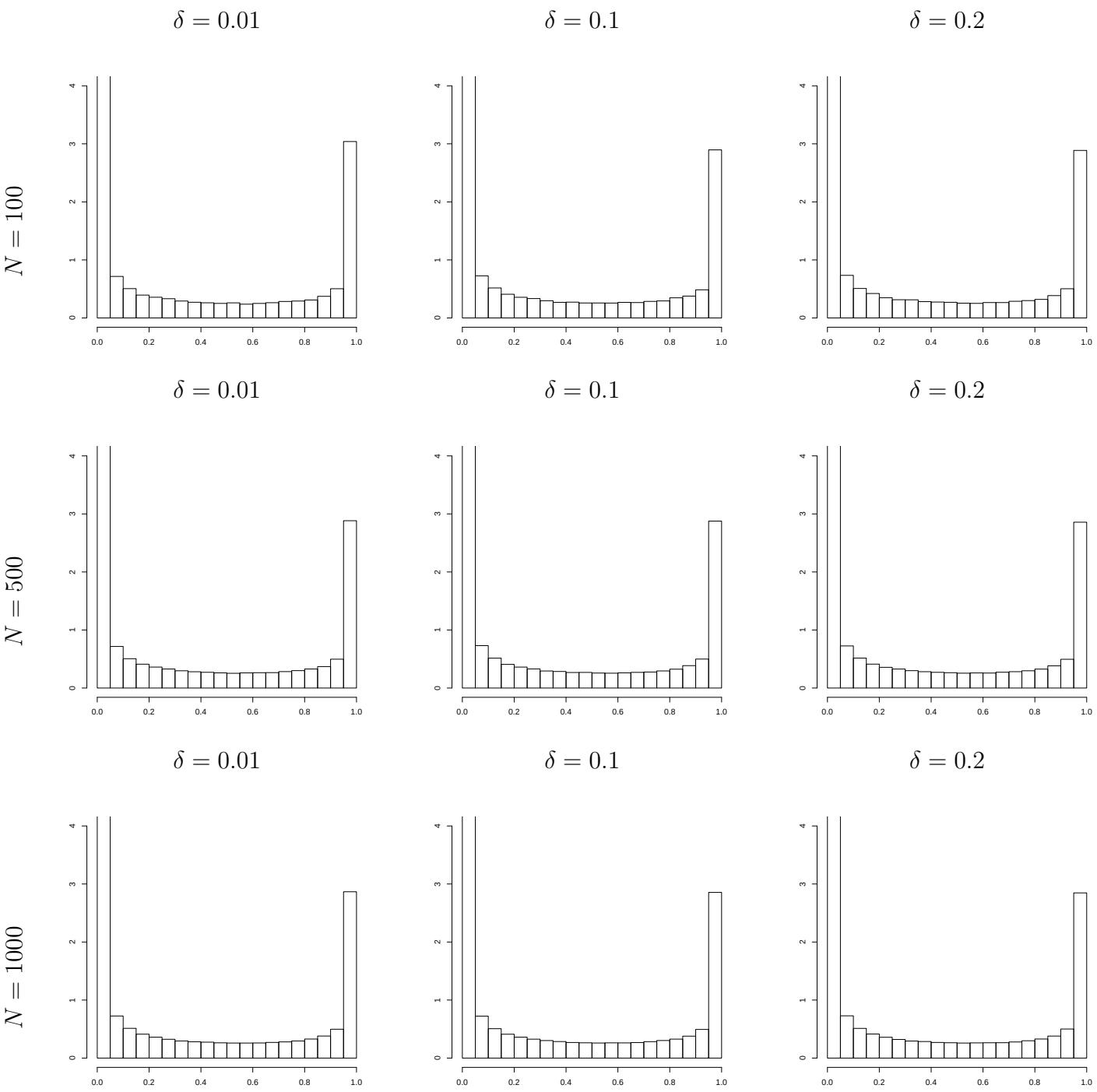


TABLE A3.1.6: Data structure for the simulated dependent variable, with $\alpha = 1$ and $\psi_Y = 1$

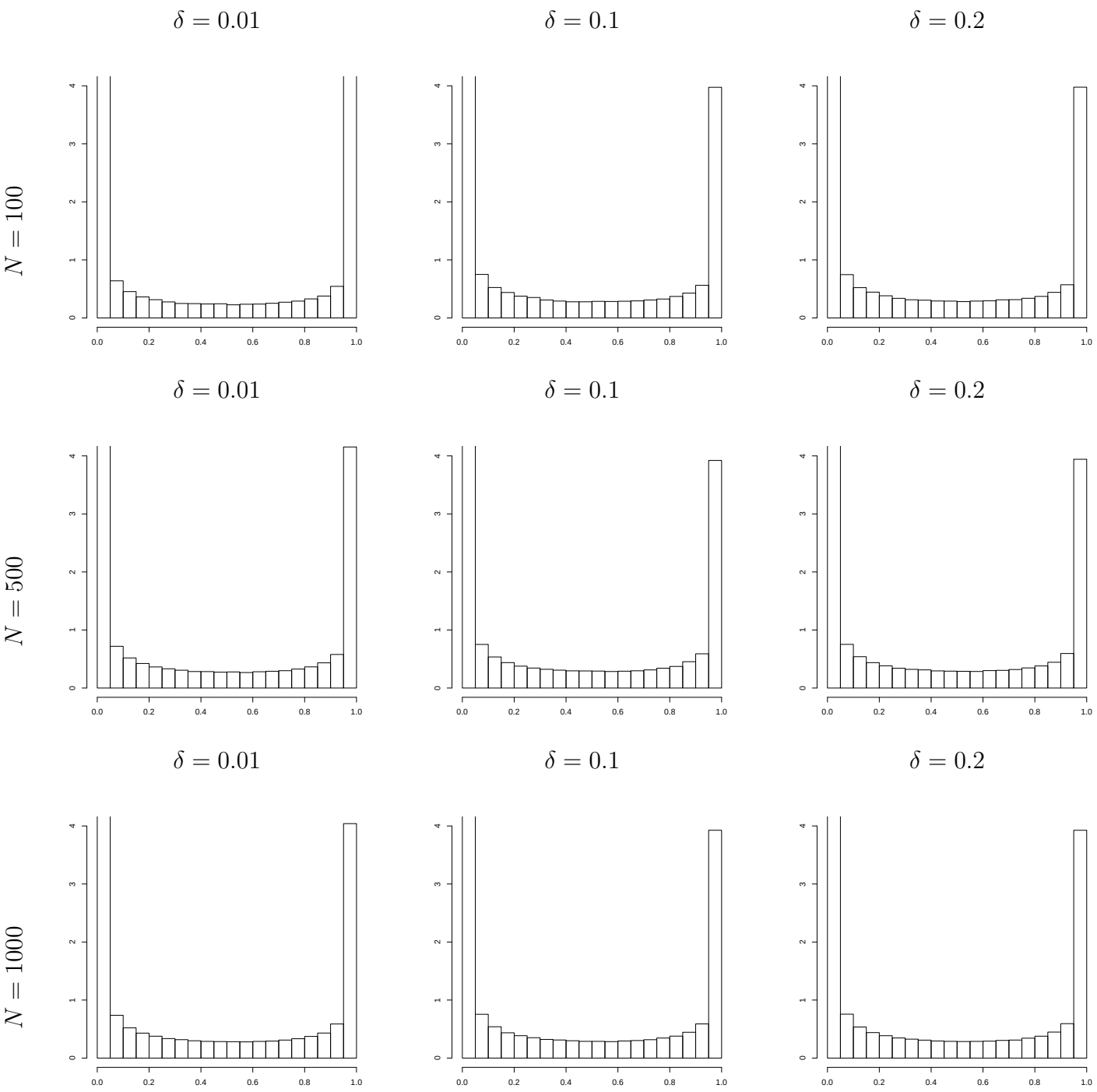


TABLE A3.1.7: Data structure for the simulated dependent variable, with $\alpha = 1.5$ and $\psi_Y = 1$

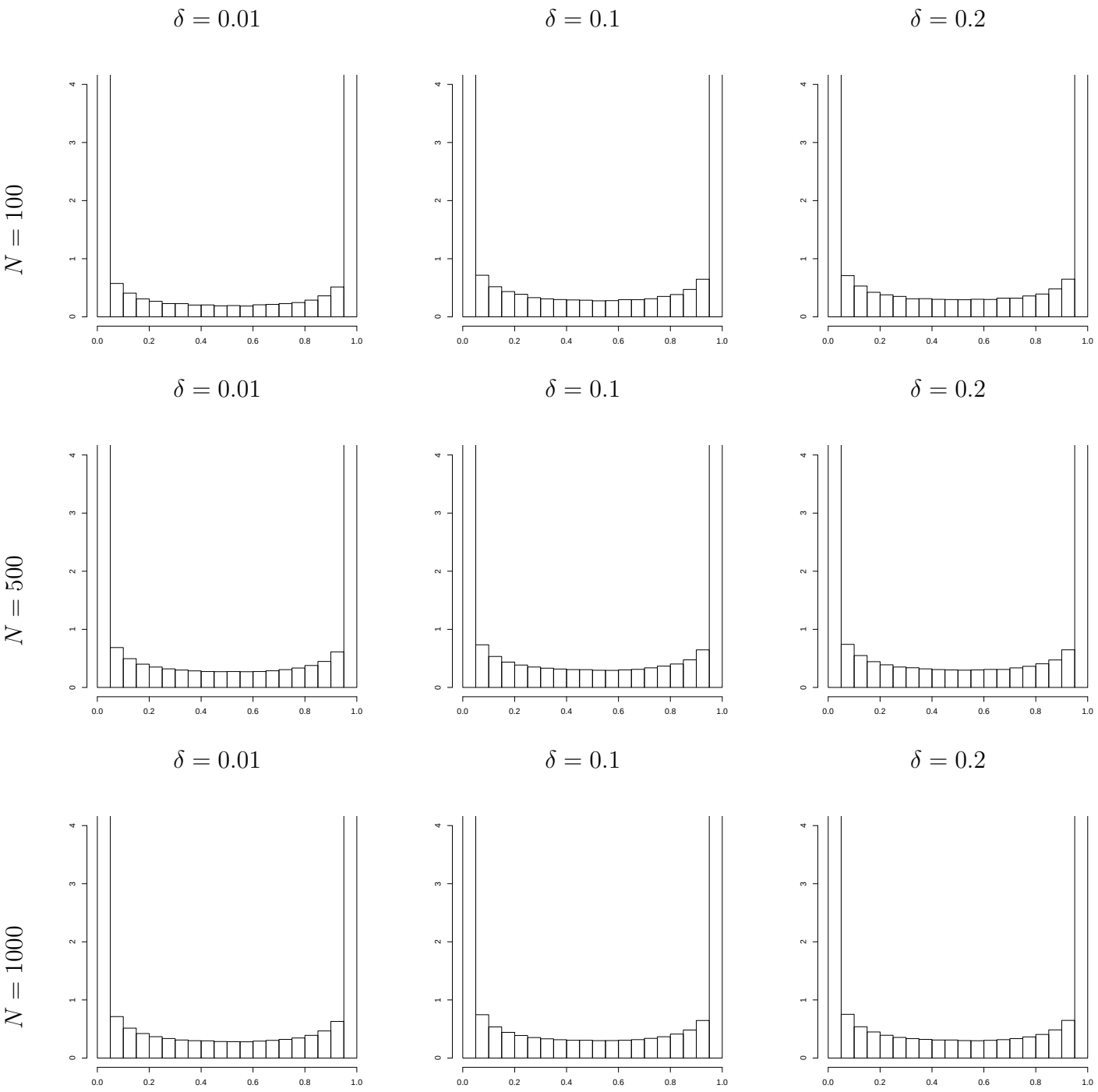


TABLE A3.1.8: Data structure for the simulated dependent variable, with $\alpha = 2$ and $\psi_Y = 1$

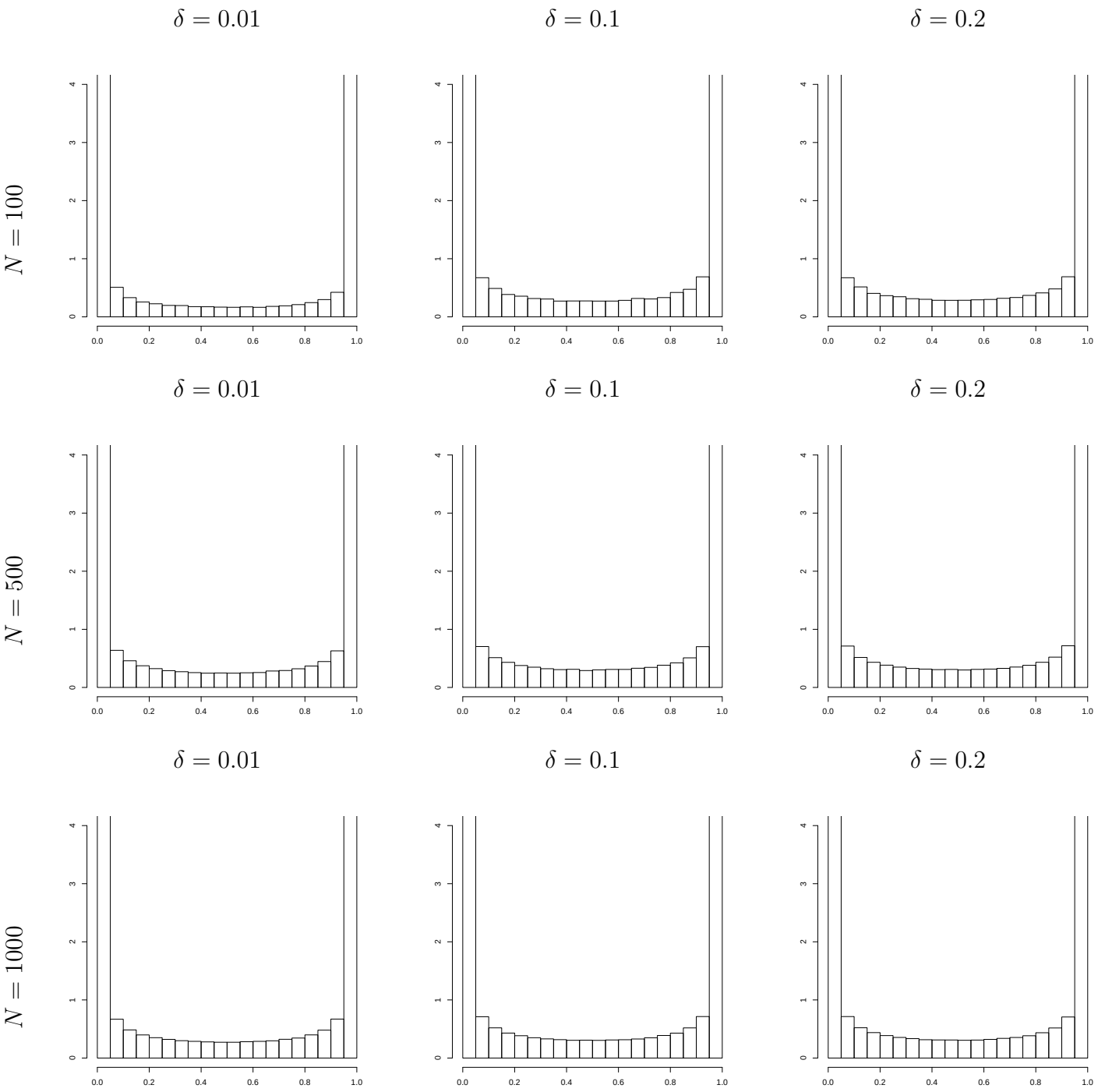


TABLE A3.1.9: Data structure for the simulated dependent variable, with $\alpha = 0$ and $\psi_Y = 10$

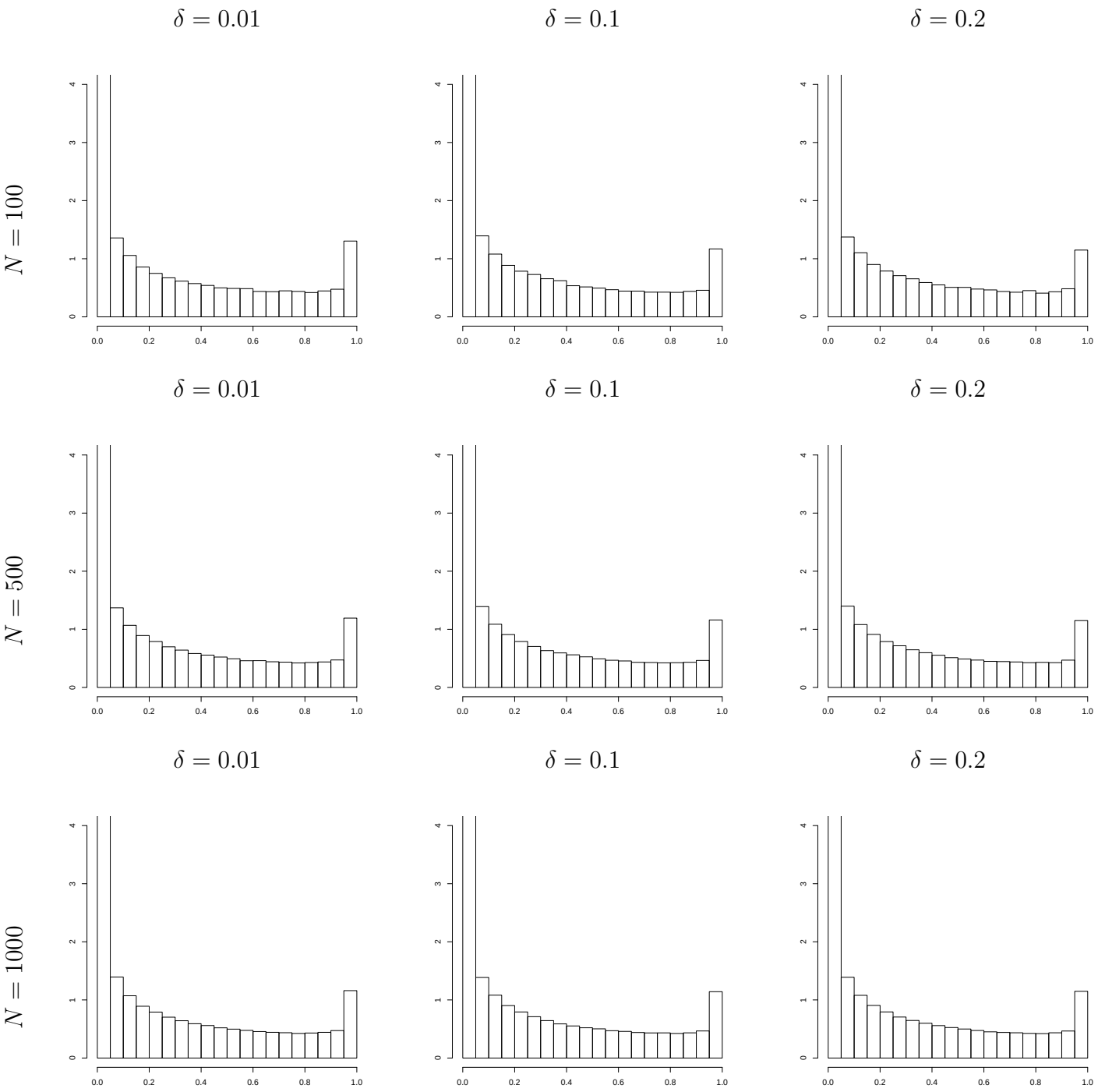


TABLE A3.1.10: Data structure for the simulated dependent variable, with $\alpha = 1$ and $\psi_Y = 10$

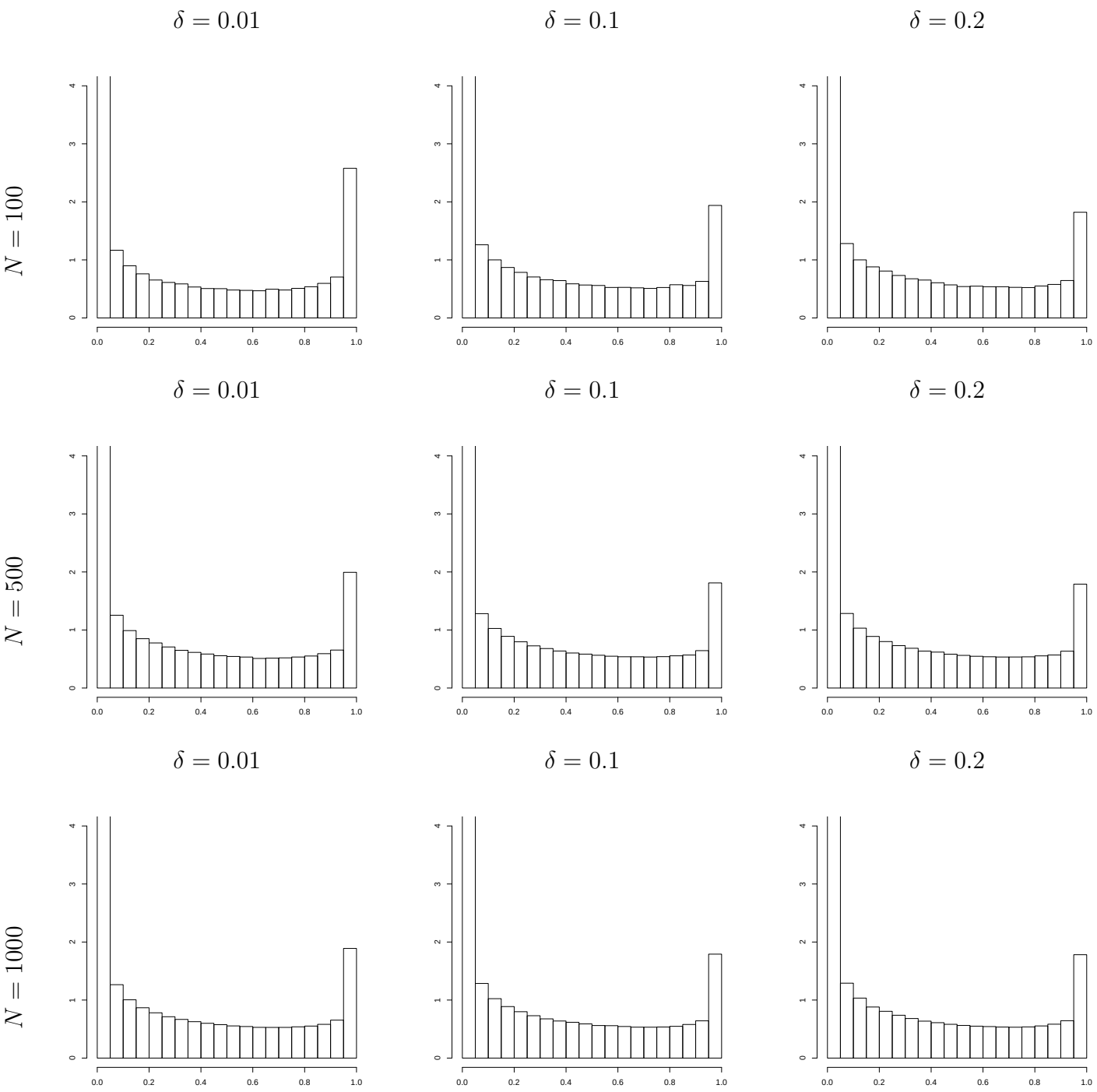


TABLE A3.1.11: Data structure for the simulated dependent variable, with $\alpha = 1.5$ and $\psi_Y = 10$

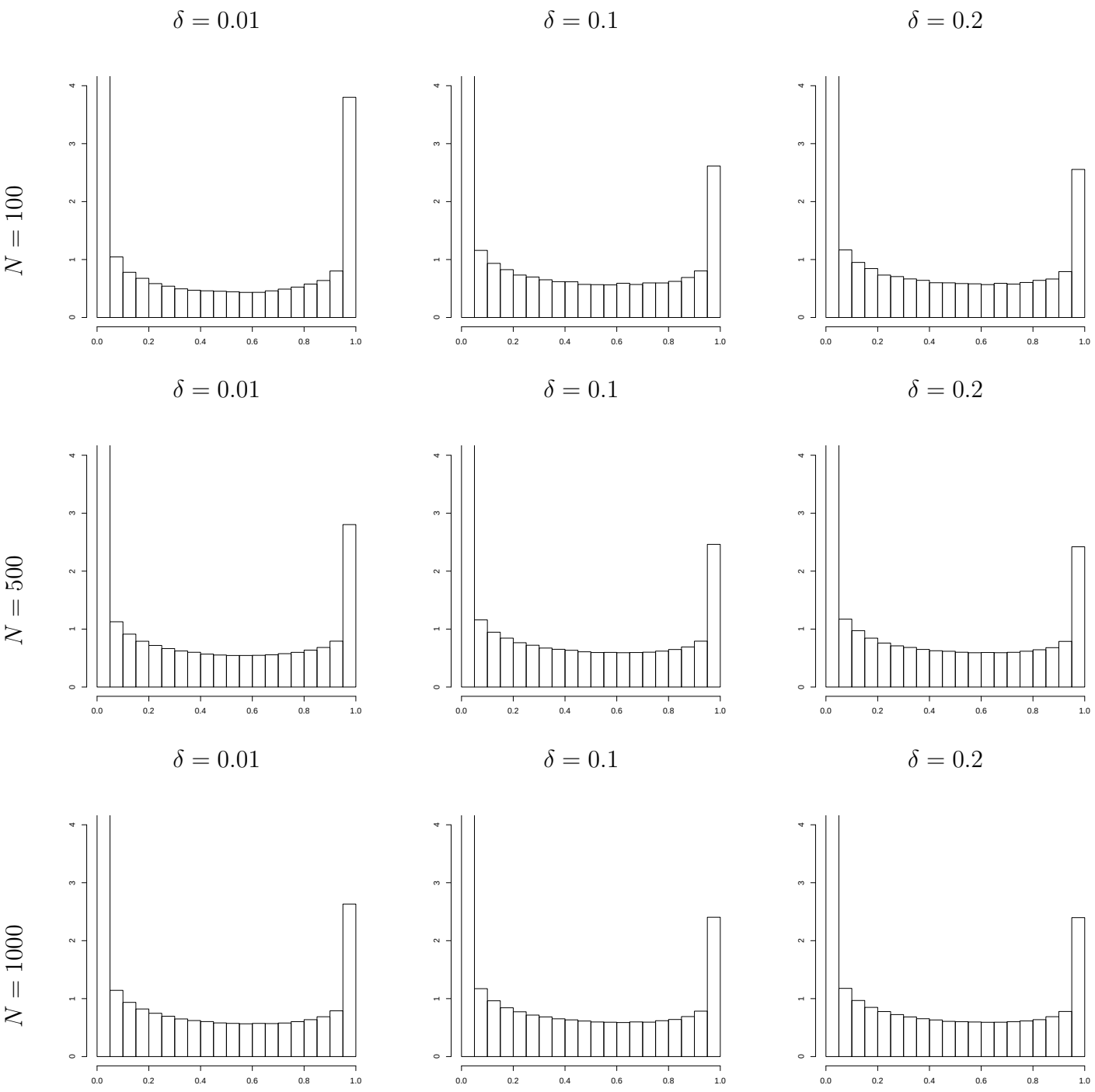
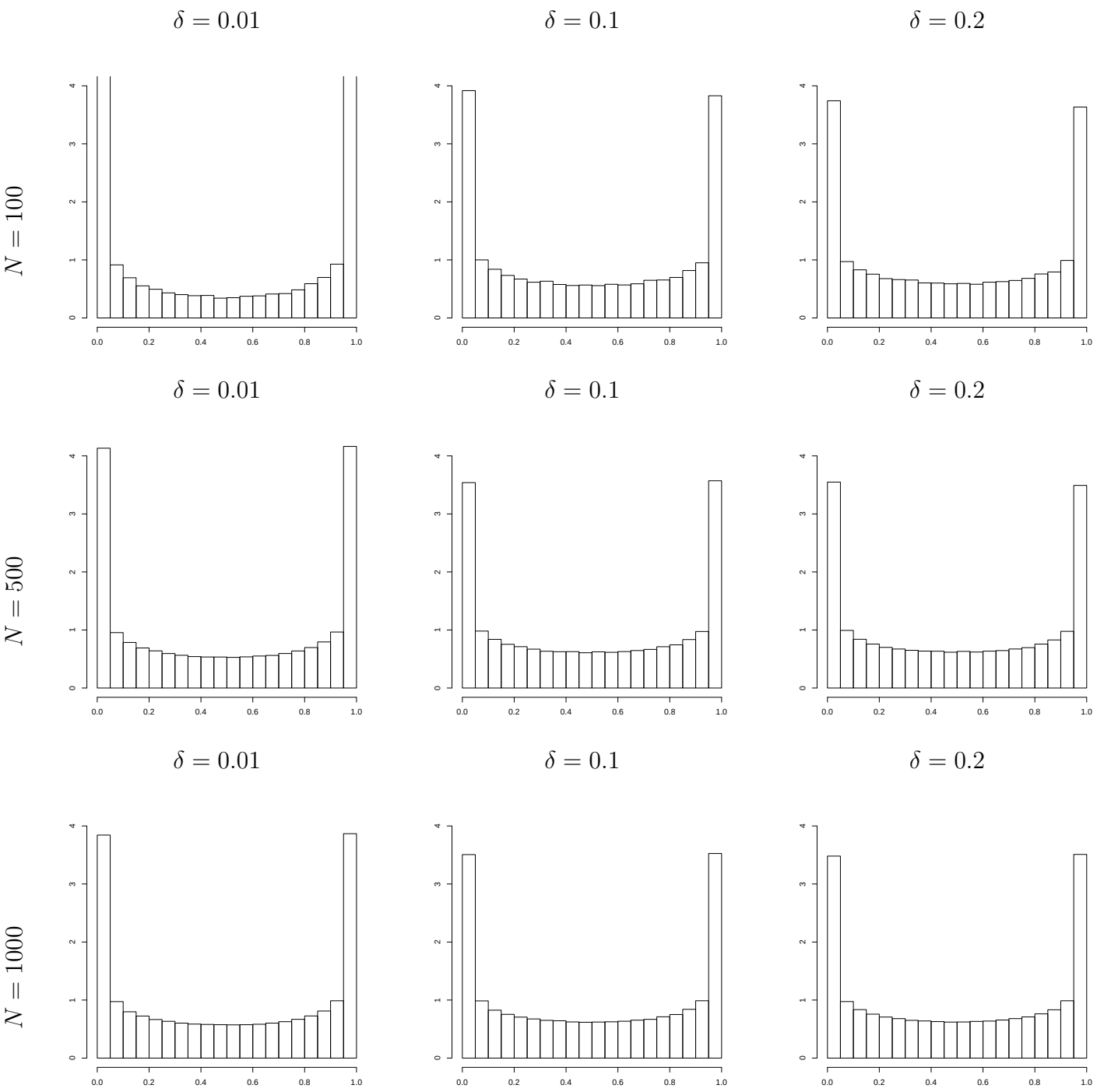


TABLE A3.1.12: Data structure for the simulated dependent variable, with $\alpha = 2$ and $\psi_Y = 10$



A3.2. GMM estimation

TABLE A3.2.1: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 0$ and $\psi_Y = 0.1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	−0.002	−0.072	0.000	−0.197	−0.178	0.191	−0.423	0.187	0.591
		(0.181)	(0.799)	(0.616)	(0.611)	(2.371)	(2.000)	(0.937)	(3.527)	(2.852)
	$\hat{\beta}_0$	0.741	−1.103	−1.056	0.782	−1.066	−1.000	0.839	−1.143	−1.030
		(1.744)	(0.428)	(0.374)	(1.793)	(0.782)	(0.605)	(1.858)	(1.056)	(0.765)
	$\hat{\beta}_1$	0.466	1.140	1.067	0.475	1.126	0.960	0.474	1.084	0.865
		(0.536)	(0.424)	(0.360)	(0.528)	(0.475)	(0.362)	(0.529)	(0.337)	(0.468)
500	$\hat{\alpha}$	−0.023	−0.049	−0.018	−0.168	0.057	0.330	−0.473	−0.129	0.336
		(0.155)	(0.564)	(0.533)	(0.563)	(1.923)	(1.764)	(0.956)	(3.441)	(2.919)
	$\hat{\beta}_0$	0.736	−0.995	−0.986	0.777	−1.022	−0.994	0.863	−0.964	−0.920
		(1.737)	(0.183)	(0.180)	(1.784)	(0.564)	(0.475)	(1.879)	(0.996)	(0.694)
	$\hat{\beta}_1$	0.473	1.009	0.990	0.475	1.007	0.905	0.474	0.999	0.817
		(0.528)	(0.087)	(0.086)	(0.526)	(0.089)	(0.174)	(0.527)	(0.096)	(0.273)
1000	$\hat{\alpha}$	−0.017	−0.024	0.002	−0.213	−0.165	0.127	−0.415	0.201	0.518
		(0.148)	(0.542)	(0.522)	(0.586)	(1.984)	(1.753)	(0.917)	(3.102)	(2.793)
	$\hat{\beta}_0$	0.732	−0.995	−0.988	0.790	−0.949	−0.929	0.848	−1.049	−0.964
		(1.733)	(0.167)	(0.165)	(1.798)	(0.577)	(0.498)	(1.863)	(0.903)	(0.653)
	$\hat{\beta}_1$	0.473	1.000	0.985	0.474	0.998	0.893	0.474	0.992	0.821
		(0.528)	(0.063)	(0.066)	(0.526)	(0.061)	(0.172)	(0.527)	(0.072)	(0.260)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.2: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1$ and $\psi_Y = 0.1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.520	1.147	0.970	0.321	1.013	1.128	0.101	1.315	1.419
		(0.514)	(0.861)	(0.563)	(0.829)	(1.860)	(1.719)	(1.179)	(2.729)	(2.513)
	$\hat{\beta}_0$	0.731	−1.170	−0.953	0.793	−1.080	−0.925	0.867	−1.171	−0.943
		(1.734)	(0.562)	(0.290)	(1.804)	(0.778)	(0.536)	(1.891)	(1.037)	(0.653)
	$\hat{\beta}_1$	0.458	1.138	0.846	0.513	1.110	0.877	0.519	1.083	0.807
		(0.545)	(0.450)	(0.340)	(0.490)	(0.415)	(0.351)	(0.484)	(0.281)	(0.343)
500	$\hat{\alpha}$	0.505	0.997	0.997	0.314	0.955	1.129	0.019	0.952	1.171
		(0.515)	(0.451)	(0.439)	(0.826)	(1.561)	(1.589)	(1.260)	(2.653)	(2.398)
	$\hat{\beta}_0$	0.734	−0.999	−0.931	0.794	−0.988	−0.907	0.900	−0.994	−0.859
		(1.735)	(0.195)	(0.161)	(1.802)	(0.568)	(0.463)	(1.922)	(0.956)	(0.658)
	$\hat{\beta}_1$	0.504	0.997	0.887	0.519	1.006	0.848	0.519	1.008	0.785
		(0.497)	(0.084)	(0.149)	(0.482)	(0.083)	(0.214)	(0.481)	(0.089)	(0.289)
1000	$\hat{\alpha}$	0.504	0.970	0.998	0.322	0.929	1.103	0.108	1.116	1.290
		(0.514)	(0.429)	(0.443)	(0.818)	(1.498)	(1.560)	(1.148)	(2.545)	(2.481)
	$\hat{\beta}_0$	0.730	−0.990	−0.940	0.790	−0.974	−0.903	0.866	−1.042	−0.892
		(1.730)	(0.168)	(0.146)	(1.798)	(0.550)	(0.461)	(1.885)	(0.920)	(0.652)
	$\hat{\beta}_1$	0.511	0.995	0.908	0.520	0.999	0.851	0.519	0.997	0.772
		(0.489)	(0.056)	(0.121)	(0.481)	(0.061)	(0.203)	(0.482)	(0.067)	(0.289)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.3: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1.5$ and $\psi_Y = 0.1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.767	1.728	1.385	0.626	1.595	1.584	0.400	1.772	1.694
		(0.756)	(0.929)	(0.716)	(0.974)	(1.692)	(1.641)	(1.311)	(2.635)	(2.189)
	$\hat{\beta}_0$	0.712	−1.227	−0.899	0.785	−1.092	−0.860	0.879	−1.150	−0.848
		(1.716)	(0.625)	(0.283)	(1.797)	(0.777)	(0.524)	(1.905)	(1.145)	(0.658)
	$\hat{\beta}_1$	0.426	1.114	0.644	0.525	1.107	0.782	0.534	1.094	0.748
		(0.577)	(0.471)	(0.469)	(0.478)	(0.347)	(0.353)	(0.469)	(0.295)	(0.389)
500	$\hat{\alpha}$	0.760	1.449	1.363	0.625	1.432	1.572	0.352	1.487	1.557
		(0.750)	(0.421)	(0.396)	(0.974)	(1.365)	(1.571)	(1.359)	(2.248)	(2.184)
	$\hat{\beta}_0$	0.735	−0.987	−0.857	0.787	−0.971	−0.850	0.899	−0.999	−0.804
		(1.736)	(0.201)	(0.196)	(1.796)	(0.581)	(0.485)	(1.924)	(0.936)	(0.666)
	$\hat{\beta}_1$	0.504	0.977	0.760	0.537	1.001	0.774	0.541	1.010	0.736
		(0.496)	(0.090)	(0.261)	(0.464)	(0.082)	(0.284)	(0.460)	(0.087)	(0.329)
1000	$\hat{\alpha}$	0.772	1.464	1.422	0.637	1.413	1.531	0.417	1.376	1.574
		(0.737)	(0.380)	(0.363)	(0.954)	(1.262)	(1.377)	(1.299)	(2.248)	(2.175)
	$\hat{\beta}_0$	0.730	−0.985	−0.873	0.782	−0.964	−0.849	0.874	−0.950	−0.801
		(1.731)	(0.173)	(0.174)	(1.790)	(0.533)	(0.455)	(1.898)	(0.940)	(0.664)
	$\hat{\beta}_1$	0.522	0.989	0.793	0.540	0.998	0.783	0.540	0.998	0.728
		(0.478)	(0.062)	(0.224)	(0.461)	(0.057)	(0.264)	(0.460)	(0.061)	(0.333)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.4: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 2$ and $\psi_Y = 0.1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.982	2.207	1.713	0.897	2.182	1.955	0.664	2.381	2.113
		(1.034)	(1.068)	(0.841)	(1.178)	(1.643)	(1.618)	(1.503)	(2.686)	(2.171)
	$\hat{\beta}_0$	0.685	−1.255	−0.873	0.789	−1.121	−0.782	0.924	−1.195	−0.758
		(1.689)	(0.603)	(0.250)	(1.803)	(0.836)	(0.562)	(1.961)	(1.350)	(0.728)
	$\hat{\beta}_1$	0.389	1.079	0.476	0.519	1.093	0.666	0.535	1.105	0.645
		(0.614)	(0.442)	(0.552)	(0.484)	(0.329)	(0.464)	(0.468)	(0.303)	(0.499)
500	$\hat{\alpha}$	0.991	1.930	1.661	0.936	1.933	1.919	0.725	2.095	1.997
		(1.015)	(0.423)	(0.490)	(1.126)	(1.190)	(1.366)	(1.434)	(2.242)	(1.988)
	$\hat{\beta}_0$	0.736	−1.004	−0.795	0.781	−0.973	−0.774	0.893	−1.046	−0.740
		(1.737)	(0.216)	(0.246)	(1.791)	(0.600)	(0.507)	(1.922)	(1.114)	(0.683)
	$\hat{\beta}_1$	0.493	0.965	0.607	0.544	1.004	0.667	0.550	1.013	0.631
		(0.508)	(0.094)	(0.406)	(0.456)	(0.083)	(0.381)	(0.450)	(0.085)	(0.427)
1000	$\hat{\alpha}$	1.031	1.946	1.727	0.898	1.836	1.856	0.691	2.012	1.912
		(0.976)	(0.365)	(0.423)	(1.168)	(1.209)	(1.358)	(1.457)	(2.006)	(1.857)
	$\hat{\beta}_0$	0.730	−0.987	−0.796	0.806	−0.916	−0.741	0.905	−1.008	−0.724
		(1.731)	(0.189)	(0.239)	(1.817)	(0.608)	(0.533)	(1.931)	(1.005)	(0.679)
	$\hat{\beta}_1$	0.517	0.975	0.637	0.550	0.999	0.679	0.552	1.000	0.644
		(0.483)	(0.065)	(0.373)	(0.451)	(0.060)	(0.368)	(0.448)	(0.060)	(0.409)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.5: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 0$ and $\psi_Y = 1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.005	−0.009	0.009	−0.208	−0.229	0.024	−0.399	0.035	0.330
		(0.161)	(0.470)	(0.435)	(0.560)	(1.682)	(1.410)	(0.894)	(2.368)	(2.218)
	$\hat{\beta}_0$	0.732	−1.054	−1.034	0.782	−0.998	−0.981	0.837	−1.046	−0.977
		(1.734)	(0.262)	(0.230)	(1.790)	(0.507)	(0.466)	(1.854)	(0.740)	(0.604)
	$\hat{\beta}_1$	0.466	1.056	1.027	0.478	1.060	0.972	0.477	1.042	0.890
		(0.536)	(0.230)	(0.204)	(0.525)	(0.230)	(0.225)	(0.525)	(0.193)	(0.248)
500	$\hat{\alpha}$	−0.001	−0.008	0.005	−0.150	0.003	0.180	−0.348	0.154	0.447
		(0.143)	(0.391)	(0.385)	(0.505)	(1.418)	(1.327)	(0.806)	(2.212)	(2.041)
	$\hat{\beta}_0$	0.729	−1.006	−1.001	0.770	−1.007	−0.991	0.826	−1.055	−1.007
		(1.730)	(0.134)	(0.131)	(1.776)	(0.420)	(0.379)	(1.839)	(0.644)	(0.509)
	$\hat{\beta}_1$	0.474	1.009	1.000	0.476	1.008	0.943	0.476	1.011	0.893
		(0.527)	(0.064)	(0.065)	(0.524)	(0.070)	(0.116)	(0.524)	(0.069)	(0.182)
1000	$\hat{\alpha}$	−0.014	−0.001	0.014	−0.172	−0.030	0.149	−0.371	−0.024	0.267
		(0.139)	(0.412)	(0.406)	(0.516)	(1.459)	(1.333)	(0.839)	(2.432)	(2.187)
	$\hat{\beta}_0$	0.732	−1.002	−0.998	0.777	−0.994	−0.980	0.832	−0.995	−0.949
		(1.733)	(0.128)	(0.127)	(1.782)	(0.426)	(0.371)	(1.846)	(0.699)	(0.569)
	$\hat{\beta}_1$	0.474	1.003	0.995	0.476	1.005	0.941	0.476	1.002	0.870
		(0.526)	(0.044)	(0.045)	(0.524)	(0.044)	(0.114)	(0.524)	(0.054)	(0.207)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.6: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1$ and $\psi_Y = 1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.562	1.074	0.971	0.371	1.011	1.130	0.138	1.020	1.217
		(0.466)	(0.516)	(0.379)	(0.756)	(1.236)	(1.341)	(1.115)	(2.033)	(2.045)
	$\hat{\beta}_0$	0.731	−1.054	−0.911	0.767	−1.044	−0.953	0.853	−1.038	−0.917
		(1.734)	(0.316)	(0.194)	(1.775)	(0.488)	(0.382)	(1.872)	(0.751)	(0.549)
	$\hat{\beta}_1$	0.461	1.051	0.831	0.512	1.053	0.900	0.517	1.044	0.840
		(0.541)	(0.253)	(0.245)	(0.490)	(0.183)	(0.228)	(0.486)	(0.183)	(0.278)
500	$\hat{\alpha}$	0.510	1.009	1.015	0.367	0.946	1.060	0.137	0.988	1.143
		(0.506)	(0.313)	(0.311)	(0.754)	(1.140)	(1.177)	(1.101)	(1.681)	(1.819)
	$\hat{\beta}_0$	0.732	−1.006	−0.947	0.774	−0.986	−0.924	0.859	−0.997	−0.899
		(1.733)	(0.138)	(0.120)	(1.780)	(0.417)	(0.359)	(1.876)	(0.613)	(0.502)
	$\hat{\beta}_1$	0.506	1.006	0.906	0.519	1.007	0.886	0.520	1.004	0.841
		(0.494)	(0.061)	(0.120)	(0.481)	(0.064)	(0.165)	(0.480)	(0.061)	(0.224)
1000	$\hat{\alpha}$	0.511	1.001	1.029	0.347	0.845	0.960	0.153	0.919	1.108
		(0.502)	(0.297)	(0.309)	(0.782)	(1.209)	(1.241)	(1.072)	(1.909)	(1.876)
	$\hat{\beta}_0$	0.727	−1.002	−0.956	0.781	−0.946	−0.892	0.851	−0.972	−0.881
		(1.727)	(0.119)	(0.104)	(1.788)	(0.441)	(0.391)	(1.867)	(0.689)	(0.535)
	$\hat{\beta}_1$	0.513	1.003	0.919	0.520	1.001	0.886	0.522	1.002	0.838
		(0.487)	(0.044)	(0.099)	(0.480)	(0.043)	(0.161)	(0.479)	(0.053)	(0.228)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.7: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1.5$ and $\psi_Y = 1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.823	1.621	1.323	0.659	1.569	1.618	0.486	1.591	1.698
		(0.693)	(0.633)	(0.447)	(0.923)	(1.053)	(1.298)	(1.191)	(1.664)	(1.754)
	$\hat{\beta}_0$	0.730	−1.075	−0.828	0.774	−1.052	−0.890	0.842	−1.056	−0.874
		(1.733)	(0.381)	(0.238)	(1.782)	(0.475)	(0.370)	(1.863)	(0.719)	(0.530)
	$\hat{\beta}_1$	0.440	1.082	0.652	0.527	1.050	0.790	0.537	1.051	0.770
		(0.562)	(0.338)	(0.384)	(0.475)	(0.196)	(0.291)	(0.465)	(0.184)	(0.322)
500	$\hat{\alpha}$	0.784	1.488	1.404	0.659	1.433	1.526	0.467	1.397	1.503
		(0.724)	(0.303)	(0.295)	(0.924)	(0.980)	(1.088)	(1.212)	(1.687)	(1.720)
	$\hat{\beta}_0$	0.731	−0.994	−0.869	0.774	−0.975	−0.873	0.852	−0.959	−0.818
		(1.731)	(0.143)	(0.164)	(1.781)	(0.418)	(0.362)	(1.871)	(0.707)	(0.558)
	$\hat{\beta}_1$	0.511	0.996	0.775	0.539	1.008	0.807	0.541	1.009	0.772
		(0.489)	(0.065)	(0.240)	(0.461)	(0.061)	(0.239)	(0.459)	(0.065)	(0.288)
1000	$\hat{\alpha}$	0.785	1.490	1.452	0.659	1.420	1.493	0.437	1.347	1.458
		(0.723)	(0.267)	(0.260)	(0.921)	(0.940)	(1.035)	(1.245)	(1.714)	(1.793)
	$\hat{\beta}_0$	0.726	−0.994	−0.887	0.774	−0.966	−0.866	0.865	−0.938	−0.806
		(1.727)	(0.123)	(0.142)	(1.781)	(0.397)	(0.347)	(1.885)	(0.716)	(0.578)
	$\hat{\beta}_1$	0.526	0.998	0.801	0.542	1.003	0.814	0.542	1.004	0.772
		(0.475)	(0.042)	(0.208)	(0.458)	(0.046)	(0.226)	(0.458)	(0.043)	(0.284)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.8: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 2$ and $\psi_Y = 1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	1.064	2.201	1.622	0.942	2.069	1.910	0.780	2.082	1.924
		(0.948)	(0.793)	(0.618)	(1.114)	(1.045)	(1.183)	(1.348)	(1.601)	(1.554)
	$\hat{\beta}_0$	0.707	−1.124	−0.790	0.780	−1.039	−0.788	0.861	−1.050	−0.762
		(1.710)	(0.430)	(0.275)	(1.791)	(0.548)	(0.449)	(1.887)	(0.815)	(0.585)
	$\hat{\beta}_1$	0.396	1.070	0.473	0.524	1.046	0.656	0.536	1.056	0.666
		(0.606)	(0.369)	(0.547)	(0.478)	(0.179)	(0.399)	(0.466)	(0.180)	(0.406)
500	$\hat{\alpha}$	1.032	1.975	1.682	0.959	1.943	1.900	0.770	1.948	1.947
		(0.973)	(0.281)	(0.406)	(1.097)	(0.893)	(1.014)	(1.354)	(1.400)	(1.495)
	$\hat{\beta}_0$	0.737	−0.989	−0.783	0.774	−0.971	−0.793	0.870	−0.975	−0.759
		(1.738)	(0.151)	(0.240)	(1.782)	(0.452)	(0.406)	(1.892)	(0.701)	(0.554)
	$\hat{\beta}_1$	0.500	0.984	0.606	0.547	1.008	0.678	0.551	1.005	0.667
		(0.501)	(0.065)	(0.401)	(0.453)	(0.060)	(0.359)	(0.450)	(0.059)	(0.385)
1000	$\hat{\alpha}$	1.051	1.967	1.725	0.961	1.915	1.881	0.793	1.975	1.957
		(0.953)	(0.245)	(0.354)	(1.096)	(0.855)	(0.989)	(1.330)	(1.470)	(1.565)
	$\hat{\beta}_0$	0.728	−0.982	−0.790	0.773	−0.956	−0.788	0.856	−0.989	−0.764
		(1.729)	(0.130)	(0.228)	(1.782)	(0.431)	(0.393)	(1.877)	(0.739)	(0.566)
	$\hat{\beta}_1$	0.523	0.989	0.643	0.551	1.004	0.686	0.553	1.002	0.662
		(0.477)	(0.043)	(0.362)	(0.449)	(0.042)	(0.350)	(0.447)	(0.040)	(0.385)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.9: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 0$ and $\psi_Y = 10$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.021	0.001	0.004	−0.120	−0.044	−0.001	−0.246	−0.009	0.068
		(0.126)	(0.184)	(0.183)	(0.398)	(0.651)	(0.631)	(0.644)	(1.016)	(1.065)
	$\hat{\beta}_0$	0.732	−1.005	−1.003	0.756	−0.997	−0.991	0.793	−1.004	−0.988
		(1.733)	(0.091)	(0.091)	(1.759)	(0.196)	(0.192)	(1.802)	(0.314)	(0.307)
	$\hat{\beta}_1$	0.464	1.010	1.006	0.475	1.009	0.992	0.477	1.006	0.969
		(0.538)	(0.070)	(0.070)	(0.527)	(0.070)	(0.069)	(0.524)	(0.098)	(0.115)
500	$\hat{\alpha}$	0.000	0.001	0.003	−0.112	−0.004	0.027	−0.230	0.025	0.088
		(0.105)	(0.157)	(0.157)	(0.392)	(0.589)	(0.576)	(0.626)	(0.919)	(0.957)
	$\hat{\beta}_0$	0.729	−1.003	−1.002	0.758	−1.001	−0.996	0.791	−1.009	−0.994
		(1.729)	(0.054)	(0.054)	(1.761)	(0.173)	(0.169)	(1.799)	(0.268)	(0.277)
	$\hat{\beta}_1$	0.473	1.002	1.001	0.476	1.002	0.988	0.476	1.001	0.967
		(0.527)	(0.028)	(0.028)	(0.524)	(0.027)	(0.036)	(0.524)	(0.041)	(0.074)
1000	$\hat{\alpha}$	−0.006	0.003	0.006	−0.092	−0.012	0.018	−0.238	−0.025	0.038
		(0.110)	(0.165)	(0.165)	(0.373)	(0.571)	(0.561)	(0.616)	(0.978)	(1.004)
	$\hat{\beta}_0$	0.730	−1.001	−1.000	0.752	−0.997	−0.992	0.795	−0.993	−0.975
		(1.730)	(0.053)	(0.053)	(1.756)	(0.167)	(0.164)	(1.802)	(0.282)	(0.288)
	$\hat{\beta}_1$	0.474	1.001	0.999	0.476	1.001	0.988	0.475	1.001	0.964
		(0.526)	(0.019)	(0.019)	(0.524)	(0.023)	(0.032)	(0.525)	(0.022)	(0.081)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.10: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1$ and $\psi_Y = 10$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.559	1.005	0.958	0.429	0.972	1.024	0.323	0.954	1.032
		(0.456)	(0.166)	(0.163)	(0.655)	(0.498)	(0.546)	(0.839)	(0.774)	(0.898)
	$\hat{\beta}_0$	0.732	−1.004	−0.919	0.749	−0.999	−0.953	0.786	−0.984	−0.932
		(1.733)	(0.097)	(0.110)	(1.754)	(0.191)	(0.166)	(1.796)	(0.287)	(0.268)
	$\hat{\beta}_1$	0.473	1.004	0.850	0.515	1.008	0.920	0.520	1.006	0.910
		(0.529)	(0.066)	(0.169)	(0.487)	(0.060)	(0.117)	(0.481)	(0.069)	(0.137)
500	$\hat{\alpha}$	0.516	0.999	1.015	0.432	0.971	1.032	0.312	0.956	1.036
		(0.492)	(0.136)	(0.143)	(0.641)	(0.453)	(0.493)	(0.839)	(0.748)	(0.848)
	$\hat{\beta}_0$	0.729	−1.000	−0.953	0.750	−0.991	−0.956	0.792	−0.986	−0.942
		(1.729)	(0.058)	(0.067)	(1.754)	(0.168)	(0.151)	(1.801)	(0.271)	(0.249)
	$\hat{\beta}_1$	0.507	1.000	0.912	0.521	1.003	0.928	0.521	1.001	0.913
		(0.493)	(0.027)	(0.094)	(0.479)	(0.025)	(0.090)	(0.480)	(0.033)	(0.122)
1000	$\hat{\alpha}$	0.518	1.006	1.037	0.437	0.996	1.052	0.301	0.944	1.018
		(0.489)	(0.125)	(0.137)	(0.635)	(0.442)	(0.495)	(0.861)	(0.789)	(0.890)
	$\hat{\beta}_0$	0.725	−1.001	−0.961	0.748	−0.999	−0.961	0.797	−0.980	−0.932
		(1.725)	(0.050)	(0.056)	(1.751)	(0.161)	(0.146)	(1.807)	(0.288)	(0.269)
	$\hat{\beta}_1$	0.514	0.999	0.922	0.521	1.001	0.926	0.522	0.997	0.909
		(0.486)	(0.018)	(0.082)	(0.479)	(0.018)	(0.092)	(0.478)	(0.040)	(0.126)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.11: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1.5$ and $\psi_Y = 10$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	0.822	1.512	1.299	0.740	1.498	1.502	0.616	1.483	1.534
		(0.686)	(0.170)	(0.266)	(0.805)	(0.425)	(0.479)	(0.983)	(0.716)	(0.816)
	$\hat{\beta}_0$	0.727	−1.008	−0.839	0.740	−1.001	−0.897	0.791	−0.994	−0.890
		(1.728)	(0.100)	(0.178)	(1.744)	(0.190)	(0.182)	(1.801)	(0.313)	(0.273)
	$\hat{\beta}_1$	0.455	1.008	0.678	0.530	1.006	0.808	0.538	1.007	0.814
		(0.547)	(0.075)	(0.333)	(0.472)	(0.062)	(0.219)	(0.464)	(0.065)	(0.225)
500	$\hat{\alpha}$	0.793	1.504	1.428	0.730	1.469	1.506	0.628	1.471	1.511
		(0.712)	(0.123)	(0.151)	(0.814)	(0.390)	(0.430)	(0.978)	(0.691)	(0.781)
	$\hat{\beta}_0$	0.727	−1.001	−0.884	0.745	−0.987	−0.902	0.784	−0.990	−0.890
		(1.727)	(0.061)	(0.126)	(1.748)	(0.166)	(0.164)	(1.794)	(0.288)	(0.262)
	$\hat{\beta}_1$	0.514	0.999	0.785	0.541	1.000	0.826	0.542	1.000	0.821
		(0.486)	(0.025)	(0.219)	(0.459)	(0.024)	(0.189)	(0.458)	(0.031)	(0.208)
1000	$\hat{\alpha}$	0.797	1.505	1.472	0.733	1.484	1.531	0.615	1.432	1.475
		(0.708)	(0.116)	(0.123)	(0.808)	(0.394)	(0.425)	(0.982)	(0.688)	(0.758)
	$\hat{\beta}_0$	0.721	−1.001	−0.898	0.742	−0.993	−0.911	0.791	−0.972	−0.881
		(1.721)	(0.053)	(0.109)	(1.746)	(0.166)	(0.156)	(1.800)	(0.286)	(0.272)
	$\hat{\beta}_1$	0.527	0.999	0.805	0.542	1.000	0.828	0.542	1.000	0.825
		(0.473)	(0.019)	(0.197)	(0.459)	(0.020)	(0.185)	(0.458)	(0.020)	(0.201)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

TABLE A3.2.12: Simulation results for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 2$ and $\psi_Y = 10$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	$\hat{\alpha}$	1.072	2.027	1.566	0.998	1.999	1.843	0.923	1.997	1.922
		(0.934)	(0.212)	(0.498)	(1.032)	(0.395)	(0.580)	(1.151)	(0.619)	(0.803)
	$\hat{\beta}_0$	0.719	−1.012	−0.774	0.752	−1.002	−0.816	0.792	−0.999	−0.815
		(1.720)	(0.124)	(0.247)	(1.757)	(0.207)	(0.257)	(1.805)	(0.318)	(0.314)
	$\hat{\beta}_1$	0.415	1.007	0.496	0.528	1.008	0.658	0.539	1.008	0.675
		(0.587)	(0.094)	(0.512)	(0.473)	(0.064)	(0.361)	(0.462)	(0.060)	(0.353)
500	$\hat{\alpha}$	1.048	1.994	1.693	1.027	1.966	1.850	0.917	1.968	1.906
		(0.955)	(0.113)	(0.342)	(0.999)	(0.347)	(0.428)	(1.154)	(0.588)	(0.741)
	$\hat{\beta}_0$	0.729	−0.997	−0.793	0.740	−0.982	−0.817	0.794	−0.984	−0.816
		(1.730)	(0.061)	(0.215)	(1.744)	(0.175)	(0.228)	(1.806)	(0.296)	(0.299)
	$\hat{\beta}_1$	0.506	0.998	0.621	0.548	1.002	0.680	0.552	1.003	0.686
		(0.494)	(0.027)	(0.382)	(0.452)	(0.024)	(0.331)	(0.448)	(0.027)	(0.337)
1000	$\hat{\alpha}$	1.063	1.997	1.759	1.035	1.993	1.873	0.924	1.948	1.858
		(0.939)	(0.107)	(0.274)	(0.991)	(0.345)	(0.394)	(1.148)	(0.598)	(0.716)
	$\hat{\beta}_0$	0.722	−0.997	−0.807	0.736	−0.997	−0.826	0.793	−0.974	−0.802
		(1.723)	(0.056)	(0.199)	(1.740)	(0.173)	(0.218)	(1.804)	(0.301)	(0.314)
	$\hat{\beta}_1$	0.527	0.998	0.646	0.551	1.001	0.678	0.554	0.999	0.687
		(0.474)	(0.018)	(0.356)	(0.449)	(0.019)	(0.332)	(0.446)	(0.022)	(0.334)

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM and the iGMM estimates of the aFRSLM.

A3.3. Partial effects

TABLE A3.3.1: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 0$ and $\psi_Y = 0.1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	0.287	0.020	3.011	0.328	0.017	0.133	0.318	0.021	0.049
		(0.289)	(0.088)	(64.306)	(0.330)	(0.038)	(2.772)	(0.321)	(0.040)	(0.090)
	AIE	0.043	0.022	2.176	0.165	0.084	0.377	0.224	0.259	0.074
		(0.059)	(0.088)	(44.169)	(0.188)	(0.302)	(9.228)	(0.238)	(1.265)	(0.103)
500	ADE	0.274	0.005	0.009	0.292	0.005	0.024	0.297	0.008	0.039
		(0.275)	(0.007)	(0.011)	(0.294)	(0.007)	(0.035)	(0.298)	(0.044)	(0.056)
	AIE	0.035	0.016	0.024	1.548	0.079	0.058	0.395	0.163	0.065
		(0.043)	(0.020)	(0.030)	(4.441)	(0.395)	(0.067)	(0.669)	(0.490)	(0.073)
1000	ADE	0.288	0.004	0.008	0.286	0.004	0.023	0.281	0.005	0.038
		(0.289)	(0.005)	(0.010)	(0.286)	(0.005)	(0.032)	(0.282)	(0.009)	(0.053)
	AIE	0.054	0.015	0.023	0.174	0.060	0.060	0.257	0.193	0.066
		(0.063)	(0.019)	(0.028)	(0.310)	(0.092)	(0.068)	(0.324)	(0.955)	(0.074)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.2: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1$ and $\psi_Y = 0.1$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	0.385	0.026	71.173	0.329	0.016	0.077	0.316	0.021	0.059
		(0.391)	(0.092)	(2035.701)	(0.333)	(0.032)	(0.968)	(0.317)	(0.052)	(0.083)
	AIE	0.325	0.043	69.205	1.385	0.120	0.107	0.661	0.286	0.071
		(0.351)	(0.204)	(1980.330)	(2.862)	(0.473)	(0.924)	(0.989)	(1.114)	(0.092)
500	ADE	0.326	0.005	0.030	0.318	0.005	0.039	0.311	0.007	0.052
		(0.327)	(0.007)	(0.221)	(0.319)	(0.008)	(0.054)	(0.312)	(0.024)	(0.069)
	AIE	0.507	0.025	0.088	0.596	0.130	0.063	0.397	0.347	0.092
		(0.627)	(0.032)	(0.926)	(0.964)	(1.024)	(0.074)	(0.871)	(2.324)	(0.769)
1000	ADE	0.327	0.003	0.021	0.318	0.004	0.038	0.323	0.004	0.055
		(0.327)	(0.004)	(0.025)	(0.318)	(0.005)	(0.050)	(0.323)	(0.008)	(0.071)
	AIE	0.476	0.035	0.048	0.471	0.090	0.063	1.122	0.362	0.082
		(0.513)	(0.348)	(0.348)	(0.921)	(0.199)	(0.072)	(2.163)	(2.992)	(0.478)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.3: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1.5$ and $\psi_Y = 0.1$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	1.012	0.086	34.989	0.448	0.030	0.066	0.411	0.030	0.067
		(1.155)	(0.851)	(704.173)	(0.479)	(0.350)	(0.354)	(0.502)	(0.216)	(0.098)
	AIE	1.922	0.116	50.257	2.423	0.291	0.089	3.165	0.747	0.089
		(2.508)	(0.984)	(1023.225)	(3.164)	(3.008)	(0.311)	(5.502)	(9.331)	(0.167)
500	ADE	0.449	0.005	1.096	0.307	0.008	0.058	0.324	0.006	0.064
		(0.457)	(0.007)	(24.807)	(0.312)	(0.110)	(0.142)	(0.325)	(0.009)	(0.081)
	AIE	2.094	0.038	5.372	8.888	0.388	0.314	1.469	0.317	0.066
		(2.554)	(0.048)	(122.072)	(18.196)	(8.016)	(8.007)	(2.970)	(1.227)	(0.086)
1000	ADE	0.369	0.004	0.102	0.382	0.003	0.051	0.328	0.015	0.065
		(0.371)	(0.005)	(1.877)	(0.411)	(0.006)	(0.063)	(0.329)	(0.351)	(0.081)
	AIE	1.422	0.034	0.575	1.199	0.155	0.056	1.419	8.179	0.070
		(1.677)	(0.042)	(14.583)	(1.723)	(1.437)	(0.071)	(2.793)	(245.927)	(0.241)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.4: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 2$ and $\psi_Y = 0.1$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	2.004	0.119	14.574	0.420	0.033	0.065	0.332	0.033	0.082
		(4.594)	(0.808)	(271.508)	(0.564)	(0.194)	(0.088)	(0.340)	(0.130)	(0.100)
	AIE	4.314	0.181	23.322	2.420	0.450	0.084	1.069	0.576	0.091
		(10.099)	(1.354)	(482.821)	(3.370)	(4.377)	(0.125)	(1.492)	(3.017)	(0.128)
500	ADE	0.465	0.008	0.052	0.342	0.005	0.068	0.330	0.008	0.082
		(0.529)	(0.009)	(0.122)	(0.344)	(0.008)	(0.081)	(0.335)	(0.036)	(0.100)
	AIE	5.643	0.068	0.130	5.025	0.178	0.062	4.559	0.675	0.079
		(6.742)	(0.079)	(0.570)	(7.359)	(0.545)	(0.088)	(8.210)	(3.594)	(0.110)
1000	ADE	0.813	0.004	0.049	0.316	0.005	0.068	0.362	0.005	0.079
		(1.080)	(0.005)	(0.052)	(0.322)	(0.043)	(0.091)	(0.372)	(0.018)	(0.094)
	AIE	9.495	0.059	0.062	13.538	0.400	0.279	30.212	1.019	0.359
		(12.448)	(0.071)	(0.067)	(26.386)	(5.672)	(5.620)	(90.022)	(9.356)	(5.564)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.5: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 0$ and $\psi_Y = 1$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	0.302	0.011	0.026	0.331	0.011	0.022	0.308	0.013	0.034
		(0.305)	(0.015)	(0.282)	(0.335)	(0.017)	(0.048)	(0.330)	(0.031)	(0.059)
	AIE	0.067	0.013	0.040	0.339	0.049	0.054	3.506	0.143	0.064
		(0.089)	(0.017)	(0.495)	(0.535)	(0.132)	(0.087)	(10.724)	(0.831)	(0.077)
500	ADE	0.293	0.004	0.007	0.284	0.004	0.018	0.289	0.005	0.027
		(0.293)	(0.005)	(0.009)	(0.285)	(0.006)	(0.024)	(0.290)	(0.018)	(0.040)
	AIE	0.046	0.011	0.017	0.112	0.042	0.050	0.207	0.106	0.062
		(0.056)	(0.014)	(0.022)	(0.125)	(0.060)	(0.059)	(0.291)	(0.434)	(0.071)
1000	ADE	0.283	0.003	0.006	0.281	0.003	0.017	0.288	0.004	0.029
		(0.283)	(0.003)	(0.008)	(0.282)	(0.004)	(0.024)	(0.289)	(0.019)	(0.042)
	AIE	0.056	0.012	0.018	0.151	0.043	0.048	0.347	0.095	0.062
		(0.073)	(0.015)	(0.022)	(0.166)	(0.067)	(0.057)	(0.556)	(0.257)	(0.070)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.6: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1$ and $\psi_Y = 1$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	0.373	0.015	25.607	0.368	0.011	0.032	0.329	0.013	0.046
		(0.378)	(0.083)	(734.984)	(0.389)	(0.020)	(0.043)	(0.332)	(0.027)	(0.064)
	AIE	0.332	0.026	37.803	3.151	0.072	0.057	0.325	0.199	0.062
		(0.362)	(0.162)	(1082.711)	(8.880)	(0.220)	(0.073)	(0.574)	(0.927)	(0.076)
500	ADE	0.338	0.004	0.019	0.323	0.004	0.031	0.321	0.004	0.041
		(0.338)	(0.004)	(0.025)	(0.324)	(0.005)	(0.041)	(0.325)	(0.008)	(0.056)
	AIE	0.484	0.016	0.035	1.466	0.064	0.055	3.098	0.154	0.062
		(0.509)	(0.020)	(0.070)	(2.414)	(0.150)	(0.065)	(6.988)	(0.745)	(0.073)
1000	ADE	0.325	0.003	0.021	0.325	0.003	0.030	0.318	0.003	0.041
		(0.326)	(0.003)	(0.023)	(0.325)	(0.003)	(0.040)	(0.318)	(0.010)	(0.057)
	AIE	0.659	0.015	0.030	3.794	0.062	0.056	1.113	0.171	0.062
		(0.781)	(0.020)	(0.035)	(10.769)	(0.087)	(0.066)	(2.479)	(0.913)	(0.073)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.7: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1.5$ and $\psi_Y = 1$.

δ		0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	1.295	0.023	248.084	0.356	0.011	0.046	0.500	0.012	0.057
		(1.313)	(0.076)	(6 118.716)	(0.379)	(0.029)	(0.060)	(0.589)	(0.023)	(0.073)
	AIE	2.547	0.033	412.301	3.001	0.107	0.054	4.449	0.208	0.061
		(2.633)	(0.082)	(10 417.626)	(5.267)	(0.525)	(0.070)	(7.192)	(0.685)	(0.081)
500	ADE	0.513	0.004	0.046	0.335	0.004	0.046	0.328	0.004	0.055
		(0.538)	(0.005)	(0.201)	(0.337)	(0.005)	(0.056)	(0.330)	(0.006)	(0.070)
	AIE	2.858	0.020	0.105	0.854	0.076	0.046	3.190	0.218	0.057
		(3.560)	(0.025)	(1.044)	(1.228)	(0.112)	(0.061)	(4.033)	(0.995)	(0.079)
1000	ADE	0.398	0.002	0.035	0.317	0.003	0.045	0.328	0.003	0.055
		(0.400)	(0.003)	(0.037)	(0.320)	(0.003)	(0.054)	(0.328)	(0.013)	(0.070)
	AIE	2.544	0.019	0.036	1.602	0.071	0.045	0.570	0.245	0.058
		(2.820)	(0.025)	(0.039)	(2.179)	(0.101)	(0.058)	(0.761)	(1.420)	(0.079)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.8: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 2$ and $\psi_Y = 1$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	12.659	0.107	512.238	0.529	0.013	0.060	0.369	0.015	0.069
		(33.113)	(1.748)	(15 665.081)	(0.603)	(0.038)	(0.074)	(0.394)	(0.047)	(0.085)
	AIE	27.203	0.130	485.927	6.066	0.162	0.063	4.682	0.406	0.076
		(70.357)	(1.755)	(14 702.239)	(7.432)	(1.029)	(0.084)	(6.731)	(2.821)	(0.101)
500	ADE	0.835	0.004	2.028	0.424	0.003	0.064	0.343	0.004	0.070
		(0.958)	(0.006)	(55.751)	(0.453)	(0.004)	(0.074)	(0.346)	(0.005)	(0.085)
	AIE	9.608	0.026	12.140	3.881	0.103	0.044	7.800	0.234	0.065
		(11.131)	(0.036)	(347.873)	(5.535)	(0.151)	(0.064)	(17.019)	(0.568)	(0.091)
1000	ADE	0.312	0.003	0.051	0.310	0.002	0.063	0.318	0.003	0.072
		(0.360)	(0.003)	(0.053)	(0.316)	(0.003)	(0.073)	(0.319)	(0.009)	(0.086)
	AIE	6.417	0.025	0.021	10.899	0.101	0.038	7.823	0.411	0.063
		(7.571)	(0.031)	(0.026)	(23.829)	(0.182)	(0.060)	(17.077)	(2.664)	(0.092)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.9: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 0$ and $\psi_Y = 10$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	0.306	0.004	0.005	0.279	0.004	0.009	0.287	0.005	0.013
		(0.310)	(0.006)	(0.007)	(0.281)	(0.006)	(0.012)	(0.288)	(0.007)	(0.020)
	AIE	0.038	0.005	0.009	0.141	0.018	0.026	0.191	0.029	0.037
		(0.049)	(0.007)	(0.011)	(0.166)	(0.024)	(0.033)	(0.225)	(0.051)	(0.046)
500	ADE	0.283	0.002	0.003	0.281	0.002	0.008	0.283	0.002	0.012
		(0.284)	(0.002)	(0.003)	(0.281)	(0.002)	(0.010)	(0.284)	(0.004)	(0.017)
	AIE	0.057	0.004	0.007	0.155	0.017	0.024	0.133	0.027	0.036
		(0.068)	(0.006)	(0.009)	(0.170)	(0.022)	(0.030)	(0.147)	(0.036)	(0.045)
1000	ADE	0.287	0.001	0.003	0.287	0.001	0.007	0.287	0.001	0.012
		(0.287)	(0.002)	(0.003)	(0.287)	(0.002)	(0.010)	(0.287)	(0.002)	(0.017)
	AIE	0.026	0.005	0.007	0.064	0.016	0.024	0.220	0.027	0.036
		(0.037)	(0.006)	(0.009)	(0.077)	(0.021)	(0.030)	(0.272)	(0.039)	(0.045)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.10: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1$ and $\psi_Y = 10$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	0.438	0.004	1.674	0.316	0.004	0.022	0.315	0.004	0.026
		(0.443)	(0.005)	(22.033)	(0.319)	(0.005)	(0.027)	(0.317)	(0.006)	(0.034)
	AIE	0.450	0.007	2.635	0.462	0.025	0.036	0.820	0.041	0.044
		(0.470)	(0.009)	(34.329)	(0.764)	(0.034)	(0.044)	(1.618)	(0.057)	(0.053)
500	ADE	0.349	0.002	0.018	0.313	0.001	0.023	0.317	0.002	0.027
		(0.350)	(0.002)	(0.019)	(0.314)	(0.002)	(0.027)	(0.317)	(0.003)	(0.034)
	AIE	0.538	0.007	0.030	0.518	0.023	0.034	0.594	0.041	0.044
		(0.596)	(0.009)	(0.032)	(0.755)	(0.030)	(0.040)	(0.812)	(0.062)	(0.051)
1000	ADE	0.333	0.001	0.020	0.318	0.001	0.024	0.320	0.001	0.027
		(0.334)	(0.001)	(0.021)	(0.318)	(0.001)	(0.028)	(0.320)	(0.004)	(0.034)
	AIE	0.625	0.007	0.028	0.733	0.024	0.033	1.993	0.042	0.044
		(0.669)	(0.008)	(0.030)	(1.080)	(0.032)	(0.039)	(3.401)	(0.079)	(0.052)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.11: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 1.5$ and $\psi_Y = 10$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	3.501	0.005	41.078	0.420	0.004	0.037	0.347	0.004	0.041
		(8.045)	(0.007)	(589.038)	(0.448)	(0.006)	(0.042)	(0.349)	(0.007)	(0.050)
	AIE	6.816	0.010	68.190	4.495	0.031	0.038	0.766	0.057	0.041
		(15.530)	(0.014)	(1177.028)	(10.486)	(0.050)	(0.045)	(1.251)	(0.150)	(0.052)
500	ADE	0.442	0.002	1.000	0.348	0.001	0.040	0.339	0.002	0.043
		(0.444)	(0.002)	(29.847)	(0.349)	(0.002)	(0.043)	(0.341)	(0.003)	(0.050)
	AIE	1.777	0.008	5.189	3.648	0.030	0.032	3.218	0.055	0.038
		(1.896)	(0.011)	(158.501)	(5.399)	(0.039)	(0.037)	(7.560)	(0.080)	(0.047)
1000	ADE	0.420	0.001	0.035	0.340	0.001	0.041	0.323	0.001	0.042
		(0.422)	(0.001)	(0.035)	(0.345)	(0.001)	(0.044)	(0.323)	(0.002)	(0.048)
	AIE	3.186	0.009	0.038	17.754	0.030	0.031	2.619	0.050	0.037
		(3.626)	(0.011)	(0.040)	(44.210)	(0.040)	(0.035)	(5.822)	(0.072)	(0.047)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

TABLE A3.3.12: Simulated bias of the estimated partial effects for the Linear Spatial Lag Model (LSLM), the Fractional Response Spatial Lag Model (FRSLM) and the approximate Fractional Response Spatial Lag Model (aFRSLM), with $\alpha = 2$ and $\psi_Y = 10$.

	δ	0.01			0.1			0.2		
N		LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM	LSLM	FRSLM	aFRSLM
100	ADE	115.762	0.013	454.341	0.461	0.004	0.054	0.797	0.005	0.060
		(354.932)	(0.052)	(10898.514)	(0.531)	(0.006)	(0.061)	(1.317)	(0.008)	(0.069)
	AIE	257.920	0.024	788.002	7.833	0.039	0.037	37.658	0.083	0.042
		(790.196)	(0.093)	(19231.881)	(9.873)	(0.057)	(0.047)	(108.317)	(0.316)	(0.055)
500	ADE	1.338	0.002	1.480	0.373	0.001	0.059	0.322	0.002	0.062
		(2.973)	(0.002)	(33.273)	(0.375)	(0.002)	(0.063)	(0.323)	(0.002)	(0.069)
	AIE	10.356	0.010	8.332	7.110	0.039	0.018	4.585	0.073	0.028
		(13.355)	(0.013)	(182.556)	(8.867)	(0.052)	(0.024)	(6.040)	(0.119)	(0.041)
1000	ADE	2.161	0.001	0.052	0.333	0.001	0.061	0.329	0.001	0.061
		(4.027)	(0.001)	(0.053)	(0.336)	(0.001)	(0.064)	(0.329)	(0.002)	(0.068)
	AIE	40.352	0.011	0.022	19.322	0.040	0.014	6.740	0.069	0.024
		(76.960)	(0.014)	(0.033)	(41.868)	(0.052)	(0.020)	(8.900)	(0.102)	(0.040)

NOTE: Simulations based on 1000 replications. Numbers are mean absolute bias w.r.t. the true measure. The estimated ADEs are compared with the ADE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication. The estimated AIEs are compared with the AIE of the FRSLM, evaluated at the true values of the parameters α, β_0 and β_1, for each replication, as well. Numbers in parentheses are root mean square errors (RMSEs). True values of the regressions parameters fixed at $\beta_0 = -1$ and $\beta_1 = 1$. The spatial weighting matrix, $\mathbf{W}$, is based on a nearest neighbor criterion. The 2SLS estimates of the LSLM are multiplied by $\sqrt{2\pi}$ to ensure comparability with the iGMM estimates of the FRSLM, the iGMM estimates of the aFRSLM and their corresponding partial effects, as well. The comparable estimated ADEs and AIEs of the LSLM are given by $\widehat{\text{ADE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ii}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}}$ and $\widehat{\text{AIE}}_{\text{LSLM}} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{I} - \sqrt{2\pi}\hat{\alpha}_{\text{2SLS}}\mathbf{W}\right)^{-1}\right\}_{ij}\sqrt{2\pi}\hat{\beta}_{1,\text{2SLS}} - \widehat{\text{ADE}}_{\text{LSLM}}$, respectively.

Chapter 4

Unobserved heterogeneity in spatial panel data models for fractional responses: an application to the proportion of High-Tech patents in the U.S. Metropolitan Statistical Areas

4.1. Introduction

The role of innovation in economic growth has been extensively studied since the seminal work of Schumpeter (1942). The popular notion of "creative destruction" establishes that knowledge, in terms of the development of new products, new production methods, new markets, and so on, is fundamental to maintain the motion of the "capital engine" and, consequently, economic development. The "entrepreneur" (the schumpeterian agent of knowledge) is assumed to have sufficient (or even perfect) information, that is rapidly materialized into innovations. The benefit of the innovation is, therefore, the profit obtained from the emerging monopoly, before the knowledge is disseminated. As a result, the "entrepreneur" would tend to have an incentive to innovate.

Three years later, Hayek (1945) rejects the conception of a single economic agent with perfect information and with a continuous propensity to innovate. In fact, knowledge is considered to be "imperfect", as it is constantly being acquired by economic agents. Instead, Hayek concentrates on the problem of the dissemination of knowledge under two opposing market frameworks. On one hand, considering an open market and free price fluctuations, economic agents endowed with knowledge and information (though not perfect), will seek profit-maximizing opportunities.

Hence, while incorporating the information on the stock of goods and the production costs into their decisions, they are, simultaneously, competing to innovate, thus promoting welfare. On the other hand, considering a monopolistic market, relying on a single economic agent (the "planner"), that manages the total "stock" of knowledge, results in an inefficient resource allocation. This is because the importance of "local knowledge" – knowledge that depends on certain advantageous conditions in space and time – tends to be ignored by aggregated measures. In this way, Hayek became the first economist to stress out the importance of analyzing knowledge as a space (and time) issue.

In the 1960s, Arrow (1962a,b) developed the theoretical and mathematical tools that led to the foundation of the economics of knowledge and innovation. The first paper addressed the hayekian problem of resource allocation to produce knowledge, under a monopolistic market framework and under a perfectly competitive market framework. Governmental funding for research and innovation is pointed out as a way to promote optimal resource allocation. It should be noted that, under this approach, the role of government is not equivalent to the that of the "planner" from Hayek. In fact, the "stock" of knowledge is allowed to be controlled by some economic agents (e.g. researchers) and institutions (e.g. research organizations and Universities). The second paper considers a general equilibrium setting and addresses the problem of knowledge and economic development, as resulting from endogenous technological change. This approach is usually referred to as the "learning-by-doing" model. Work experience (in the sense of learning) is considered a key factor to the production of knowledge and innovation. But at the same time, producing knowledge and innovation leads to a new phase of technical experience. This simultaneous process ultimately increases labor productivity and welfare, by reducing production costs and promoting efficiency, respectively. Moreover, in Arrow (1962b), the role of institutions, education and research to promote and disseminate learning, knowledge and innovation, is also emphasized.

But it was in the late 1970s that Griliches (1979) made an important contribution to the state of art of economics of knowledge and innovation, developing the, so called, "knowledge production function". Historically, the classic approaches to

measure total output and total factor productivity considered a set of trivial inputs, such as labor and capital. Nevertheless, during 1950s, several applied studies (see Griliches, 1973, for an excellent survey) found that knowledge, in terms of public and private expenditures in Research and Development (R&D), was highly correlated with the growth of output and productivity. Recognizing the relevance of such conclusions and the limitations of the classic approach to measure the output, Griliches (1979) suggests to augment the classic Cobb-Douglas production function by a proxy measure of knowledge. Accordingly, the current "stock" of knowledge is defined by a function of all present and past expenditures in R&D and a set of unobserved inputs, besides labor and capital. On this matter, Griliches (1979) also points out three major issues, regarding the measurement of R&D using industry-level data. First, the existence of a time gap between the initial investment and the measured effects in productivity. Second, as innovations become obsolete, the current "stock" of R&D should take into consideration both the current innovations and the past innovations deducted by a depreciation factor. Third, the "stock" of knowledge in a given industry cannot be derived by specific R&D expenditures, as knowledge can be disseminated, borrowed or stolen from another industry or sector. On one hand, the first and second issues can be easily accommodated in a model specification through time lags of the R&D expenditures and the adjusted value of R&D resources. On the other hand, the third issue motivated a growing interest on the measurement and modeling of the spatial spillovers from knowledge and innovation. Examples are Acs and Audretsch (1988), Acs et al. (1994), Audretsch and Feldman (1996) and Los and Verspagen (2000), considering U.S. sector-level and industry-level data; Cassiman and Veugelers (2002), Monjon and Waelbroeck (2003) and Becker and Dietz (2004), considering sector-level and industry-level data from Belgium, France and Italy, respectively.

However, most of the empirical works using the R&D expenditures to explain knowledge and innovation were highly critized. As Griliches (1973) stresses out, this was due to the fact that the R&D expenditures did not provide any economic value for the innovative outputs. In fact, they only provide an economic valuation for the resources allocated to the development of innovative products or services.

Alternatively, Griliches (1979) and Pakes and Griliches (1980) suggest to assess the adequacy of patents as a proxy measure for knowledge and innovation. But at the same time, these authors point out two important pitfalls related to this measure. One, not all innovations are patented. In fact, considering the "R&D 100 Award" database from the *Research and Development* magazine, Fontana et al. (2013) found that about 10% of the awarded innovations were not patented. Two, different patents yield different externalities. Hence, patents cannot be directly compared, in terms of their economic impact. Even so, Pakes and Griliches (1980) discuss the adequacy of using patent counts to measure knowledge. Considering a panel of U.S. industry-level data and using both patent counts and R&D expenditures to measure knowledge increments in research-intensive industries, results show that the patent counts explained most of the between variation of research-intensive industries. Since then, many applied studies turned their attention to the measurement and modeling the knowledge and innovation spillovers, using patent counts or patent citations. Examples are Jaffe (1989a,b), Jaffe et al. (1998) and Sanyal (2003), considering U.S. sector-level and industry-level data; Audretsch and Vivarelli (1996), considering industry-level data from Italy; van Meijl (1997), Autant-Bernard (2001) and Piergiovanni and Santarelli (2001), considering sector-level, regional-level and industry-level data from France, respectively.

It is important to emphasize that most of the existing literature did not explicitly considered a spatial approach to address the specification and the estimation of knowledge and innovation spillovers. Prior to the publication of Anselin's seminal book (Anselin, 1988), the theoretical framework for Spatial Econometrics was still widely unexplored. After that, applied research in Spatial Econometrics became widespread. Focusing on the analysis of knowledge and innovation, it is worth mentioning the first works that addressed modeling and estimation through a spatial approach. Anselin et al. (1997) and Acs et al. (2002), considering U.S. Metropolitan Statistical Areas innovation counts and patent data, respectively, and Fischer and Varga (2003) considering Austrian districts patent data. A spatial error model was estimated to assess the effects of industry R&D expenditures, university R&D expenditures, professional employment (highly skilled research staff) and a geographic

coincidence index (see Jaffe, 1989b), on patent counts. In Fischer and Varga (2003), spatial lags of the explanatory variables were added instead of the geographic coincidence index.

It is also important to emphasize that the majority of works considered a linear approach to estimate the log-transformed patent counts or the log-transformed patent citations. On this matter, extensive literature in Econometrics (see Cameron and Trivedi, 2013; Winkelmann, 2008) examine two possible scenarios for count data models. For the case where the patent counts are strictly positive, the log-transformation is well defined and the linear approach can be used. The estimated parameters are elasticities. For the case where there are observations with zero patent counts, the log-transformation is no longer valid. The linear approach can only be used if *ad hoc* transformations are applied to adjust the zero counts. In consequence, the estimated parameters can no longer be interpreted as elasticities. In addition, retrieving the conditional expectation of the re-transformed count variable can be difficult. For further discussion on the issues related to the re-transformation of the dependent variable see Wooldridge (2010). Nevertheless, there are two notable examples in the literature that consider an appropriate approaches to estimate models with zero patent counts. Cincera (1997) estimate a non-spatial conditional panel Poisson model, to assess the effects of R&D investments and Jaffe's geographic coincidence index on patent counts, using data from european, japanese and american firms. Wang et al. (1998) estimate a non-spatial mixed Poisson model, to assess the effects of the R&D-to-sales ratio on patent counts, using data from pharmaceutical and biomedical companies in the United States. Still, to the best of the author's knowledge, there is no applied research that considers a spatial approach to estimation and focuses on the problem of zero patent counts.

In recent years, most of the literature using spatial frameworks addresses knowledge and innovation spillover effects through the spatial interaction model. This can be interpreted as a refinement of the gravity model, to assess the regional and technological effects on the flows of knowledge and innovative activity in space. Examples are LeSage et al. (2007), Fischer and Griffith (2008) and Fischer et al. (2009), estimating a Poisson spatial interaction model, to assess the effects of geographic and

technological proxy measures on the high-tech patent citations, using data from central european high-tech industries. LeSage et al. (2007) considers a bayesian approach to estimation, while Fischer and Griffith (2008) and Fischer et al. (2009) estimate the model by maximum likelihood. Fischer et al. (2009) also consider a spatially lagged error in the specification. However, these works focus on cited patents, leaving out all the patents that were not cited and not accommodating potential selection bias.

The purpose of this chapter is threefold. First, it assesses how R&D expenditures and Human Capital proxies (such as wages and employment) influence the knowledge spillovers in the U.S. Metropolitan Statistical Areas (U.S. MSAs), between 2010 and 2015. The proportion of U.S. origin high-tech patents is used as a proxy measure for knowledge and innovative activity. This choice is obvious for three main reasons. One, the majority of the high-tech innovations are effectively patented, as Fischer et al. (2009) points out. This is because they tend to involve huge amounts of R&D expenditures, but also because patenting generates economic benefits such as monopolistic profits, controlled dissemination of critical knowledge and widespread economic impact (externalities). Hence, considering the high-tech patents as a proxy measure of knowledge and innovative activity appears to be more adequate than to consider all patents indistinctly. Two, using the proportion of U.S. origin high-tech patents instead of their counts allows to assess the relative importance of high skilled knowledge in the U.S. MSAs. Even if a particular federal or local government policy has an effect on the aggregate or regional patent counts, it may not have a significant effect on the "stock" of high-skilled knowledge and/or on the relative importance of the U.S. origin high-tech patents in the U.S. MSAs. Three, data from the U.S. Patent and Trademark Office (`https://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm`) shows that more than 90% of the U.S. origin high-tech patents are attributed to individuals, companies or institutions that are localized in the U.S. MSAs. With regard to the explanatory variables, data on the R&D expenditures at U.S. Colleges and Universities in the U.S. MSAs, by source of funding (federal government, state and local government, business and institutional), is collected from the Higher Education Research and Development Survey

(https://www.nsf.gov/statistics/srvyherd). Data on median real wages and the total number of employed in the U.S. MSAs, both by educational level – Ph.D., graduate (except Ph.D.) and non-graduate –, is collected from the American Community Survey (https://usa.ipums.org/usa/acs.shtml). The median real wages by educational level can be interpreted as proxy measures for both the R&D expenditures in industries and Human Capital (considering the income-based approach).

Second, it presents two new specifications to model the proportion of U.S. origin high-tech patents. A spatial lag model specification is considered, in order to assess spatial spillovers of the explanatory variables. The proposed specifications combine the approach of Papke and Wooldridge (2008) and the approach presented in Chapter 3. The former develops attractive functional forms to model fractional responses with panel data, whereas the latter develops attractive specifications to model spatially lagged fractional responses with purely spatial data. The fractional nature of the dependent variable is properly taken into account and observations at the corners 0 and 1 are allowed. Hence, no *ad hoc* transformations are applied to the data. The first specification is the Panel Fractional Response Spatial Lag Probit Model (PFRSLPM). The PFRSLPM extends the approach of Papke and Wooldridge (2008) to the spatial framework, introducing a spatial lag of the fractional dependent variable inside a nonlinear function. Also, the PFRSLPM extends the FRSLM specification developed in Chapter 3 to the panel data setting, introducing a time invariant unobserved effect into the specification. The second specification is the approximate Panel Fractional Response Spatial Lag Probit Model (aPFRSLPM) consists in a first order series approximation of the PFRSLPM around the spatial lag parameter equal to zero. The time invariant unobserved effect is allowed to be correlated with the explanatory variables. Hence, the treatment of the unobserved effect is addressed as a central issue. The spatial approach of Debarsy (2012) is considered to model the unobserved heterogeneity. This approach extends the classic approaches of Mundlak (1978) and Chamberlain (1980) to the spatial framework.

Third, it addresses the issues of considering a spatially lagged linear approach to model the proportion of U.S. origin high-tech patents. The estimates and the partial effects obtained from the linear approach are discussed and compared with

those obtained from the previous proposed specifications (the PFRSLPM and the aPFRSLPM).

Estimation of the proposed specifications (the PFRSLPM and the aPFRSLPM) is addressed by the iterative Generalized Method of Moments (iGMM) procedure of Klier and McMillen (2008), recently revisited in Chapter 1. The spatial heteroskedasticity and spatial autocorrelation (spatial HAC) robust estimator of Kelejian and Prucha (2007) is considered, to produce valid inference for the asymptotic covariance estimator of the GMM estimator for the unknown parameter vector.

Results show that the degree of spatial dependence of the proportion of U.S. origin high-tech patents is quite low. This can be due to spatial aggregation, monopolistic behavior within the MSAs and/or the absence of dynamic space-time effects. In addition, employment appears to play an important role in terms of the regional innovative activity. Furthermore, the estimates obtained for the PFRSLPM and the aPFRSLPM are quite similar. Using the Debarsy (2012) approach to model the unobserved heterogeneity tends to produce better results, than those obtained for the case where the unobserved heterogeneity is ignored or the spatial dimension is neglected.

The remainder of the chapter is organized as follows. Section 4.2 introduces the two new specifications (PFRSLPM and aPFRSLPM) for spatially lagged fractional responses with panel data and deduces their corresponding partial effects. Section 4.3 presents the GMM estimation procedure and a consistent asymptotic covariance estimator. The gradients for the proposed models are deduced, as well. Section 4.4 illustrates the proposed models through an empirical application on the proportion of U.S. origin high-tech patents in the U.S. Metropolitan Statistical Areas. Section 4.5 concludes. The descriptive analysis for the variables included in the empirical application are summarized in Section A4. The estimation outputs and the estimated partial effects are presented in Section B4.

4.2. Spatial Panel models for fractional responses with correlated Random Effects

In this section the spatial panel models for fractional responses are presented and the set of underlying assumptions are stated and discussed. Plus, the corresponding partial effects are deduced, as well. The first model is the Panel Fractional Response Spatial Lag Probit Model (PFRSLPM) and the second model is the approximate Panel Fractional Response Spatial Lag Probit Model (aPFRSLPM). The proposed models extend two works on this subject. One, the PFRSLPM extends the panel data approach of Papke and Wooldridge (2008) to the spatial framework, introducing spatial dependence through a spatial lag of the fractional dependent variable. Two, both the PFRSLPM and aPFRSLPM extend the specifications developed in Chapter 3 to the panel data setting, through the introduction of a time invariant unobserved effect into each specification. Complexity is added by allowing the unobserved effect to be correlated with the explanatory variables. This issue is addressed by the spatial approach of Debarsy (2012), consisting in a generalization of the classic approaches of Mundlak (1978) and Chamberlain (1980) to model the unobserved heterogeneity.

4.2.1. Panel Fractional Response Spatial Lag Probit Model

The Panel Fractional Response Spatial Lag Probit Model (PFRSLPM) with unobserved effects follows as:

$$Y_{it} = \Phi\left(\alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=1}^{K} \beta_k x_{it,k} + c_i\right) + u_{it}, \qquad i, j = 1, 2, \ldots, N,\ t = 1, 2, \ldots, T \tag{81}$$

where $0 \leq Y_{it} \leq 1$ is a fractional dependent variable for the ith spatial unit, at time t, with $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. The total number of spatial units are denoted by N and the total number of temporal observations are denoted by T. The coefficients w_{ij} do not change over time and are non-negative scalars that correspond to the spatial weights of unit j on unit i, with $i \neq j$ and $j = 1, 2, \ldots, N$. Also, by convention, $w_{ii} = 0$, for all i and t. The scalar α is the spatial lag parameter. The variables $x_{it,1}, x_{it,2}, \ldots, x_{it,K}$ are a set of K time-varying explanatory

variables (that may include time dummies), for the ith spatial unit, at time t. The corresponding regression parameters are $\beta_1, \beta_2, \ldots, \beta_K$. The individual unobserved effect or individual unobserved heterogeneity, c_i, does not change over time and may be correlated with some of the explanatory variables. The disturbance term, u_{it}, is a random error, for the ith spatial unit, at time t, defined on the closed interval $[-\Phi(\cdot), 1-\Phi(\cdot)]$. The function $\Phi(\cdot)$ is the univariate Standard Normal cumulative distribution function (CDF). For the purpose of this chapter, considering the Probit specification is not restrictive, as all the details that will be presented can be easily extended to a general link function. In addition, this specification produces computationally simple estimators for a great variety of sampling scenarios. For this reason, it is used in most of the theoretical approaches and empirical applications, as described in the introduction.

Stacking over the spatial units yields:

$$\mathbf{Y}_t = \Phi(\alpha \mathbf{W}\mathbf{Y}_t + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{c}) + \mathbf{u}_t \tag{82}$$

where $\mathbf{Y}_t = [Y_{1t}, Y_{2t}, \ldots, Y_{Nt}]^\intercal$ is an $N \times 1$ vector of fractional responses, for each t. The $N \times K$ matrix of explanatory variables is $\mathbf{X}_t = [\mathbf{X}_{1t}^\intercal, \mathbf{X}_{2t}^\intercal, \ldots, \mathbf{X}_{Nt}^\intercal]^\intercal$, for each t, with $\mathbf{X}_{it} = [x_{it,1}, x_{it,2}, \ldots, x_{it,K}]^\intercal$, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. The corresponding $K \times 1$ parameter vector is $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_K]^\intercal$. The $N \times 1$ vector of the individual unobserved effects is $\mathbf{c} = [c_1, c_2, \ldots, c_N]^\intercal$, constant for all t, and the $N \times 1$ vector of random errors is $\mathbf{u}_t = [u_{1t}, u_{2t}, \ldots, u_{Nt}]^\intercal$, for each t. The matrix $\mathbf{W}$ is the $N \times N$ spatial weighting matrix, equal for all t and with generic element w_{ij}. By definition, $\mathbf{W}$ is a non-stochastic non-negative matrix with zeros on its main diagonal.

Without further assumptions, two important issues must be emphasized. First, consider the expectation of Y_i given $(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t}, c_i)$:

$$\begin{aligned} &\mathrm{E}(Y_{it} \mid \mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t}, c_i) \\ &= \Phi\left(\alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=1}^{K} \beta_k x_{it,k} + c_i\right) + \mathrm{E}(u_{it} \mid \mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t}, c_i) \end{aligned} \tag{83}$$

where $\mathbf{Y}_{-i,t}$ is the vector of responses excluding the ith response, at time t, with $t =$

$1, 2, \ldots, T$ and $i = 1, 2, \ldots, N$. Note that, the conditional expectations $\mathrm{E}\left(c_i \,|\, \mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t}\right)$ 0 and $\mathrm{E}\left(u_{it} \,|\, \mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t}, c_i\right) \neq 0$, once $\mathrm{Cov}\left(\mathbf{WY}_t, \mathbf{c}\right) \neq 0$ and $\mathrm{Cov}\left(\mathbf{WY}_t, \mathbf{u}_t\right) \neq 0$, respectively, for all t, due to the simultaneous nature of the model. Second, the partial effects:

$$\begin{aligned}\frac{\partial Y_{it}}{\partial x_{it,k}} = \phi &\left(\alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=1}^{K} \beta_k x_{it,k} + c_i\right) \\ &\times \left(\alpha \sum_{j \neq i} w_{ij} \frac{\partial Y_{jt}}{\partial x_{it,k}} + \beta_k + \frac{\partial c_i}{\partial x_{it,k}}\right) + \frac{\partial u_{it}}{\partial x_{it,k}}\end{aligned} \tag{84}$$

are not identified, once they depend on the unobserved effect, c_i. The partial effects in (84) would only be identified if they are averaged-out across the distribution of c_i, assuming that $\partial u_{it}/\partial x_{it,k} = 0$, for all i and t.

To address the issues above, it is assumed that Assumption 3.1 and Assumption 3.2, from Chapter 3, hold. The set of additional assumptions follow along the lines of Papke and Wooldridge (2008). They focus on the structure of the data, on the treatment of the unobserved heterogeneity and on the statistical properties of the error term.

Assumption 4.1. The number of spatial units is large relative to the number of time periods, that is $N/T \to \infty$.

Assumption 4.2. The explanatory variables are strictly exogenous, conditional on the unobserved effect, that is $\mathrm{E}\left(Y_{it} \,|\, \mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t}, c_i\right) = \mathrm{E}\left(Y_{it} \,|\, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}, c_i\right)$, with $t = 1, 2, \ldots, T$.

Assumption 4.3. The unobserved heterogeneity, c_i is normally distributed, given $(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t})$, with linear expectation and constant variance, and it is allowed to be correlated with some explanatory variables and with the spatially lagged responses. The conditional distribution of c_i is modeled based on the specification of Debarsy (2012):

$$c_i \,|\, (\mathbf{X}_{i1}, \mathbf{X}_{i2}, ..., \mathbf{X}_{iT}, \mathbf{Y}_{-i,t}) \sim \mathcal{N}\left(\varphi_0 + \sum_{k_1=1}^{K_1} \varphi_{k_1,1} \bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1} \varphi_{k_1,2} \sum_{j \neq i} w_{ij} \bar{x}_{j,k_1}, \sigma^2_{\mathbf{e}}\right) \tag{85}$$

where the variables $\bar{x}_{i,1}, \bar{x}_{i,2}, \ldots, \bar{x}_{i,K_1}$ are time averages of a subset K_1 of the K explanatory variables (not including the time dummies). The term $\sum_{j \neq i} w_{ij} \bar{x}_{j,k_1}$ is the spatial lag of the time averages $\bar{x}_{i,1}, \bar{x}_{i,2}, \ldots, \bar{x}_{i,K_1}$. The scalar φ_0 is a constant. The regression parameters for the non-spatially lagged term are $\varphi_{1,1}, \varphi_{2,1}, \ldots, \varphi_{K_1,1}$ and the regression parameters for the spatially lagged term are $\varphi_{1,2}, \varphi_{2,2}, \ldots, \varphi_{K_1,2}$. The scalar parameter $\sigma^2_{\mathbf{e}}$ is the variance of the new time-invariant component, e_i. In addition, e_i is independent from the set $(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t})$, for all i and t, and independent from the random error u_{it}, for all i and t, as well.

Assumption 4.4. The random error, u_{it}, has zero mean and is independent of $x_{it,k}$ and c_i, for all $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$ and $k = 1, 2, \ldots, K$.

Assumption 4.1 is usually considered under micro-panels frameworks (see also Papke and Wooldridge, 2008). Assumption 4.2 is common in the literature on panel data models with unobserved effects. This assumption is somewhat restrictive, in the sense that it rules out the following three scenarios. One, the presence of time lags (and consequently space-time lags) of the explanatory variables. Two, correlation between one or more elements in $\mathbf{X}_{it}$ and past and future values of Y_{it}. Three, correlation between the explanatory variables and omitted time-varying variables. Assumption 4.3 provides an extension of the classic approaches of Chamberlain (1980) and Mundlak (1978) to model the unobserved heterogeneity in spatial frameworks. This assumption may be restrictive in the sense that it imposes a parametric distribution for c_i, but, at the same time, it allows a particular dependence structure between c_i and other two components. One, between c_i and $\mathbf{X}_{it}$, through the time averages of the explanatory variables. Two, between c_i and $\sum_{j \neq i} w_{ij} Y_{jt}$, through the spatially lagged time averages of the explanatory variables. Hence, it is assumed that the simultaneous correlation between c_i and $\sum_{j \neq i} w_{ij} Y_{jt}$ is modeled by $\sum_{k_1=1}^{K_1} \varphi_{k_1,2} \sum_{j \neq i} w_{ij} \bar{x}_{j,k_1}$.[12] Assumption 4.4 establishes a convenient exogenous setting for the error term.

[12] From the simulations studies presented in this thesis, it is possible to conclude that, using lower order approximations to address the simultaneous nature of models with spatially lagged dependent variables, generally leads to promising results, in terms of accuracy and efficiency.

Under the previous assumptions the conditional expectation in (83) can be written as:

$$\begin{aligned}
&\mathrm{E}\left(Y_{it} \,|\, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}, e_i\right) \\
&= \Phi\left(\alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=0}^{K} \beta_k x_{it,k} + \sum_{k_1=1}^{K_1} \varphi_{k_1,1} \bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1} \varphi_{k_1,2} \sum_{j \neq i} w_{ij} \bar{x}_{j,k_1} + e_i\right) \\
&\quad + \mathrm{E}\left(u_{it} \,|\, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}, e_i\right)
\end{aligned} \tag{86}$$

where the constant was included in the set of the explanatory variables, for simplicity. Thus, using the law of iterated expectations,

$$\begin{aligned}
&\mathrm{E}\left(Y_{it} \,|\, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}\right) = \mathrm{E}\left[\mathrm{E}\left(Y_{it} \,|\, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}, e_i\right) | \, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}\right] \\
&= \mathrm{E}\left[\Phi\left(\alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=0}^{K} \beta_k x_{it,k} \right.\right. \\
&\qquad \left.\left. + \sum_{k_1=1}^{K_1} \varphi_{k_1,1} \bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1} \varphi_{k_1,2} \sum_{j \neq i} w_{ij} \bar{x}_{j,k_1} + e_i\right) \middle| \, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}\right] \\
&\quad + \mathrm{E}\left[\mathrm{E}\left(u_{it} \,|\, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}, e_i\right) | \, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}\right] \\
&= \Phi\left(\frac{\alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=0}^{K} \beta_k x_{it,k} + \sum_{k_1=1}^{K_1} \varphi_{k_1,1} \bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1} \varphi_{k_1,2} \sum_{j \neq i} w_{ij} \bar{x}_{j,k_1}}{\sqrt{1 + \sigma_{\mathbf{e}}^2}}\right) \\
&\quad + \mathrm{E}\left(u_{it} \,|\, \mathbf{X}_{it}, \mathbf{Y}_{-i,t}\right)
\end{aligned} \tag{87}$$

and the PFRSLPM becomes:

$$Y_{it} = \Phi\left(\frac{\alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=0}^{K} \beta_k x_{it,k} + \sum_{k_1=1}^{K_1} \varphi_{k_1,1} \bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1} \varphi_{k_1,2} \sum_{j \neq i} w_{ij} \bar{x}_{j,k_1}}{\sqrt{1 + \sigma_{\mathbf{e}}^2}}\right) + u_{it}^* \tag{88}$$

with $i, j = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. The new error is now u_{it}^*, a function of the initial error u_{it}, that results from the conditions imposed on the unobserved heterogeneity. In addition, note that (88) can be viewed as the spatial extension of the classic panel data models with correlated random effects. Stacking over the

spatial units yields:

$$\mathbf{Y}_t = \Phi\left(\frac{\alpha\mathbf{W}\mathbf{Y}_t + \mathbf{X}_t\boldsymbol{\beta} + \bar{\mathbf{X}}\boldsymbol{\varphi}_1 + \mathbf{W}\bar{\mathbf{X}}\boldsymbol{\varphi}_2}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right) + \mathbf{u}_t^* \tag{89}$$

hence all the parameters are attenuated towards zero, once $\sigma_{\mathbf{e}}^2$ is strictly positive. The $N \times K_1$ matrix with the time averages of the explanatory variables is $\bar{\mathbf{X}} = \left[\bar{\mathbf{X}}_1^\intercal, \bar{\mathbf{X}}_2^\intercal, \ldots, \bar{\mathbf{X}}_N^\intercal\right]^\intercal$, with $\bar{\mathbf{X}}_i = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_K]^\intercal$, for all $i = 1, 2, \ldots, N$. The corresponding parameter vectors are the $K_1 \times 1$ vector $\boldsymbol{\varphi}_1 = [\varphi_{1,1}, \varphi_{2,1}, \ldots, \varphi_{K_1,1}]^\intercal$ and the $K_2 \times 1$ vector $\boldsymbol{\varphi}_2 = [\varphi_{1,2}, \varphi_{2,2}, \ldots, \varphi_{K_1,2}]^\intercal$, for the time averages and the spatially lagged time averages, respectively.

It is also useful to state the following assumption on the stability of the PFRSLPM:

Assumption 4.5. There is a constant ζ such that,

$$\zeta = \sup_{\alpha\in A}\left|\frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right| \times \sup_{\eta}\phi(\eta) \times \sup_{N}\|\mathbf{W}_N\|_\infty < 1 \tag{90}$$

where $A \subset \mathbb{R}$ is the compact parameter space of α, $\phi(\eta)$ is the Standard Normal probability density function (PDF), with $\eta \in \mathbb{R}$, and $\|\mathbf{W}_N\|_\infty = \max_{i=1,2,\ldots,N}\sum_{j=1}^{N}|w_{ij}|$ is the row sum norm.

Assumption 4.5 is analogous to Assumption 3 in Xu and Lee (2015a) and to Assumption 3.5 in Chapter 3. Under the previous assumptions and the specification of the PFRSLPM, $\mathbf{W}$ is uniformly bounded and $\phi(\eta)$ is a bounded function, thus ζ is finite. For the case where the working $\mathbf{W}$ is row standardized, $\|\mathbf{W}\|_\infty = 1$, and because the link function is the Standard Normal CDF, $\sup_\eta \phi(\eta) = (2\pi)^{-1/2}$, hence the constant $\zeta = \sup_{\alpha\in A}|\alpha| \times [(2\pi)(1+\sigma_{\mathbf{e}}^2)]^{-1/2}$. This poses a restriction in the parameter space of α, in the sense that, the condition $\zeta < 1$ is only satisfied for the values of α that verify the inequality $|\alpha| < [(2\pi)(1+\sigma_{\mathbf{e}}^2)]^{1/2}$. Therefore, the stability of the PFRSLPM does not depend exclusively on the parameter space of α, contrary to the pure spatial case (see Xu and Lee, 2015a, and Chapter 3). In fact, after simple algebra, it is obvious that the model is stable if and only if:

$$\alpha^2 - 2\pi\sigma_{\mathbf{e}}^2 < 2\pi \tag{91}$$

where α and $\sigma_{\mathbf{e}}^2$ can only take values that satisfy the inequality above. For the case where $\sigma_{\mathbf{e}}^2 = 0$ and no time-invariant component is present, stability is ensured for the values of $|\alpha| < (2\pi)^{1/2}$ (see Xu and Lee, 2015a, and Chapter 3). For the case where $\sigma_{\mathbf{e}}^2 > 0$, the parameter space of α becomes wider. Note that, in terms of interpretation and comparability, the degree of spatial dependence has to be multiplied by $[(2\pi)(1+\sigma_{\mathbf{e}}^2)]^{1/2}$ to be comparable with the usual -1 and 1 admissible range for the spatial lag parameter, in spatial linear models.

Having defined the set of assumptions that allow to accommodate the correlation between c_i with the explanatory variables and with the spatially lagged responses, the partial effects can now be properly identified. The general partial effect for the kth explanatory variable over the ith spatial unit response, at time t, is given by:

$$\frac{\partial Y_{it}}{\partial x_{it,k}} = \phi\left(\frac{\alpha\sum_{j\neq i} w_{ij}Y_{jt} + \sum_{k=0}^{K}\beta_k x_{it,k} + \sum_{k_1=1}^{K_1}\varphi_{k_1,1}\bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1}\varphi_{k_1,2}\sum_{j\neq i} w_{ij}\bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right) \times \left(\frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\sum_{j\neq i} w_{ij}\frac{\partial Y_{jt}}{\partial x_{it,k}} + \frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right) \tag{92}$$

Note that the partial derivative $\partial u_{it}^*/\partial x_{it,k} = 0$, for all $k = 1, 2, \ldots, K$, due to Assumption 4.4. Stacking over the spatial units, yields:

$$\boldsymbol{\Delta}_{k,t} = \left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right]^{-1}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \tag{93}$$

for all $k = 1, 2, \ldots, K$ and $t = 1, 2, \ldots, T$. The matrix $\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}$ is a $N \times N$ diagonal matrix, for each t, whose diagonal elements are given by $\phi(\eta_{it})$, with

$$\eta_{it} = \frac{\alpha\sum_{j\neq i} w_{ij}Y_{jt} + \sum_{k=0}^{K}\beta_k x_{it,k} + \sum_{k_1=1}^{K_1}\varphi_{k_1,1}\bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1}\varphi_{k_1,2}\sum_{j\neq i} w_{ij}\bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \tag{94}$$

The $N \times N$ matrix $\left\{\mathbf{I} - \left[\alpha/(1+\sigma_{\mathbf{e}}^2)^{1/2}\right]\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right\}^{-1}$ is the PFRSLPM link augmented spatial lag operator inverse, with $\mathbf{I}$ the $N \times N$ identity matrix, for each

t. Note that, under Assumption 4.5, the matrix $\left\{\mathbf{I} - \left[\alpha/\left(1+\sigma_{\mathbf{e}}^2\right)^{1/2}\right] \mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right\}$ is non-singular.

The time-varying matrix of partial effects, $\boldsymbol{\Delta}_{k,t}$, provides an interesting way to assess the effects of exogenous unitary changes over the responses, at different moments in time. In addition, it provides information on the way these effects are transmitted across different units in space and time. An exhaustive interpretation of each partial effect in $\boldsymbol{\Delta}_{k,t}$, for each t, can be done, but it is not common. Instead it is more appealing to summarize the time-varying partial effects into the five measures suggested by LeSage and Pace (2009). The first measure is the Average Direct Effect:

$$\begin{aligned} ADE_{k,t} &= \frac{1}{N}\sum_{i=1}^{N}\left\{\left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right]^{-1}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\right\}_{ii}\frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \\ &= \frac{1}{N}\operatorname{tr}\left(\left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right]^{-1}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\right)\frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \end{aligned} \tag{95}$$

given by the average of the diagonal elements of $\boldsymbol{\Delta}_{k,t}$, for each t. The function $\operatorname{tr}(\cdot)$ is the trace of a matrix. The second measure is the Average Total Effect To an observation:

$$ATET_{kj,t} = \frac{1}{N}\sum_{i=1}^{N}\left\{\left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right]^{-1}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\right\}_{ij}\frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \tag{96}$$

given by the average of the jth column of $\boldsymbol{\Delta}_{k,t}$, for each t. The third measure is the Average Total Effect From an observation:

$$ATEF_{ki,t} = \frac{1}{N}\sum_{j=1}^{N}\left\{\left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right]^{-1}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\right\}_{ij}\frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \tag{97}$$

given by the average of the ith row of $\mathbf{\Delta}_{k,t}$, for each t. The fourth measure is the Average Total Effect:

$$ATE_{k,t} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\mathbf{W}\right]^{-1}\mathbf{D}_{\phi(\boldsymbol{\eta}_t)}\right\}_{ij}\frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \tag{98}$$

given by the average of all elements of $\mathbf{\Delta}_{k,t}$, for each t. Finally, the fifth measure is the Average Indirect Effect:

$$AIE_{k,t} = ATE_{k,t} - ADE_{k,t} \tag{99}$$

given by the average of all the off-diagonal elements of $\mathbf{\Delta}_{k,t}$, for each t. These measures can also be evaluated at specific values of the explanatory variables, say means, medians, maximums, minimums or at different quantiles. Furthermore, global measures can be obtained by taking time averages of each of the five summary measures.

Nevertheless, there are two important limitations regarding the PFRSLPM with correlated random effects. First, obtaining the reduced form under a nonlinear simultaneous specification, is computationally complex. Even if Assumption 4.5 holds, complex numerical methods are required to implicitly compute the reduced form. However, no analytic formula can be deduced. As a result, the interpretation of the reduced form parameters becomes problematic. Second, due to the simultaneous nature of the PFRSLPM, the effects of unitary changes in a given explanatory variable over the responses, at any time t, depend on every element of $\mathbf{X}_t$, including the time averages and their spatial lags, and on the spatially lagged responses, as well. This poses a severe restriction for policy analysis, in the sense that policy makers cannot explicitly control the responses of each spatial unit. In practice, they can only control the exogenous determinants of the responses. See also Chapter 3 for further discussion.

In the next section, an alternative specification to model fractional responses with spatial panels data is presented. The proposed specification extends the aFRSLM specification, proposed in Chapter 3, to the panel data setting. Plus, it addresses the previously mentioned issues on the PFRSLPM in a simple way.

4.2.2. Approximate Panel Fractional Response Spatial Lag Probit Model

Consider that Assumption 4.1 to Assumption 4.5 are maintained. Consider the first-order series expansion of the PFRSLPM nonlinear function in (88), around $\alpha = 0$:

$$\begin{aligned}
&\Phi\left(\frac{\alpha\sum_{j\neq i}w_{ij}Y_{jt}+\sum_{k=0}^{K}\beta_k x_{it,k}+\sum_{k_1=1}^{K_1}\varphi_{k_1,1}\bar{x}_{i,k_1}+\sum_{k_1=1}^{K_1}\varphi_{k_1,2}\sum_{j\neq i}w_{ij}\bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right)\\
&\approx\Phi\left(\frac{\sum_{k=0}^{K}\beta_k x_{it,k}+\sum_{k_1=1}^{K_1}\varphi_{k_1,1}\bar{x}_{i,k_1}+\sum_{k_1=1}^{K_1}\varphi_{k_1,2}\sum_{j\neq i}w_{ij}\bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right)\\
&\quad+\frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\phi\left(\frac{\sum_{k=0}^{K}\beta_k x_{it,k}+\sum_{k_1=1}^{K_1}\varphi_{k_1,1}\bar{x}_{i,k_1}+\sum_{k_1=1}^{K_1}\varphi_{k_1,2}\sum_{j\neq i}w_{ij}\bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right)\sum_{j\neq i}w_{ij}Y_{jt}
\end{aligned} \tag{100}$$

then the approximate Panel Fractional Response Spatial Lag Model (aPFRSLPM) with correlated Random Effects follows as:

$$\begin{aligned}
Y_{it}\approx\Phi&\left(\frac{\sum_{k=0}^{K}\beta_k x_{it,k}+\sum_{k_1=1}^{K_1}\varphi_{k_1,1}\bar{x}_{i,k_1}+\sum_{k_1=1}^{K_1}\varphi_{k_1,2}\sum_{j\neq i}w_{ij}\bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right)+\frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\\
&\times\phi\left(\frac{\sum_{k=0}^{K}\beta_k x_{it,k}+\sum_{k_1=1}^{K_1}\varphi_{k_1,1}\bar{x}_{i,k_1}+\sum_{k_1=1}^{K_1}\varphi_{k_1,2}\sum_{j\neq i}w_{ij}\bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right)\sum_{j\neq i}w_{ij}Y_{jt}+u_{it}^*
\end{aligned} \tag{101}$$

with $i,j=1,2,\ldots,N$, and $t=1,2,\ldots,T$. Stacking over the spatial units yields:

$$\begin{aligned}
\mathbf{Y}_t&\approx\Phi\left(\frac{\mathbf{X}_t\boldsymbol{\beta}+\bar{\mathbf{X}}\boldsymbol{\varphi}_1+\mathbf{W}\bar{\mathbf{X}}\boldsymbol{\varphi}_2}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right)+\frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)}\mathbf{W}\mathbf{Y}+\mathbf{u}_t^*\\
\Leftrightarrow\mathbf{Y}_t&\approx\left[\mathbf{I}-\frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)}\mathbf{W}\right]^{-1}\Phi\left(\frac{\mathbf{X}_t\boldsymbol{\beta}+\bar{\mathbf{X}}\boldsymbol{\varphi}_1+\mathbf{W}\bar{\mathbf{X}}\boldsymbol{\varphi}_2}{\sqrt{1+\sigma_{\mathbf{e}}^2}}\right)+\mathbf{v}_t
\end{aligned} \tag{102}$$

where $\mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)}$ is an $N \times N$ diagonal matrix, for each t, whose diagonal elements are $\phi\left(\eta_{it}^a\right)$, with η_{it}^a equal to:

$$\eta_{it}^a = \frac{\sum_{k=0}^{K} \beta_k x_{it,k} + \sum_{k_1=1}^{K_1} \varphi_{k_1,1} \bar{x}_{i,k_1} + \sum_{k_1=1}^{K_1} \varphi_{k_1,2} \sum_{j \neq i} w_{ij} \bar{x}_{j,k_1}}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \tag{103}$$

The new error term is given by $\mathbf{v}_t = \left\{\mathbf{I} - \left[\alpha/\left(1+\sigma_{\mathbf{e}}^2\right)^{1/2}\right] \mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} \mathbf{W}\right\}^{-1} \mathbf{u}_t^*$, for each t. The $N \times N$ matrix $\left\{\mathbf{I} - \left[\alpha/\left(1+\sigma_{\mathbf{e}}^2\right)^{1/2}\right] \mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} \mathbf{W}\right\}^{-1}$ is the aPFRSLPM link augmented spatial lag operator inverse, for each t. As before, under Assumption 4.5, the matrix $\left\{\mathbf{I} - \left[\alpha/\left(1+\sigma_{\mathbf{e}}^2\right)^{1/2}\right] \mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} \mathbf{W}\right\}$ is non-singular.

This specification has two important advantages, as noted in Chapter 3. First, the PFRSLPM can be written as an approximate reduced form, with tractable analytic expression. In this way, the spatial units can be interpreted as resulting from an approximate steady-state equilibrium between the responses and the exogenous explanatory variables[13]. Second, it allows policy makers to consider the resulting approximate partial effects, as they rely exclusively on exogenous explanatory variables. In fact, the matrix of the approximated partial effects is given by:

$$\begin{aligned}
\boldsymbol{\Delta}_{k,t} &\approx \left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} \mathbf{W}\right]^{-1} \\
&\quad \times \left[\mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} + \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \mathbf{D}_{\boldsymbol{\phi}'(\boldsymbol{\eta}_t^a)} \mathbf{W}\, \boldsymbol{\Phi}^*(\boldsymbol{\eta}_t^a)\right] \frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \\
\Leftrightarrow \boldsymbol{\Delta}_{k,t} &\approx \left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} \mathbf{W}\right]^{-1} \\
&\quad \times \left\{\mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} \circ \left[\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \mathbf{D}_{\boldsymbol{\eta}_t^a} \mathbf{W}\, \boldsymbol{\Phi}^*(\boldsymbol{\eta}_t^a)\right]\right\} \frac{\beta_k}{\sqrt{1+\sigma_{\mathbf{e}}^2}}
\end{aligned} \tag{104}$$

for all $k = 1, 2, \ldots, K$ and $t = 1, 2, \ldots, T$. The function

$$\boldsymbol{\Phi}^*(\boldsymbol{\eta}_t^a) = \left\{\mathbf{I} - \frac{\alpha}{\sqrt{1+\sigma_{\mathbf{e}}^2}} \mathbf{D}_{\boldsymbol{\phi}(\boldsymbol{\eta}_t^a)} \mathbf{W}\right\}^{-1} \Phi\left(\boldsymbol{\eta}_t^a\right) \tag{105}$$

[13] See also LeSage and Pace (2010) for a discussion on this matter, considering a Linear Spatial Lag Model.

and $\mathbf{D}_{\boldsymbol{\eta}_t^a}$ is an $N \times N$ diagonal matrix, for each t, whose diagonal elements are η_{it}^a. The operator "$\circ$" is the Hadamard product. Note that the first derivative of the PDF of the Standard Normal distribution $\phi'(\mathfrak{u}) = -\mathfrak{u}\,\phi(\mathfrak{u})$ with $\mathfrak{u} \in \mathbb{R}^N$.

Analogous to the PFRSLPM specification, the matrix of partial effects, $\boldsymbol{\Delta}_{k,t}$, can be summarized according to the five measures proposed by LeSage and Pace (2009), the ADE, ATET, ATEF, ATE and AIE. These measures were already deduced in the previous section.

4.3. GMM estimation

The estimation of the proposed panel models for spatially lagged fractional responses considers the Generalized Method of Moments (GMM) approach presented on Chapter 2, based on the works of Pinkse and Slade (1998) and Klier and McMillen (2008). Simulations presented in this thesis show the adequacy of the GMM estimator, to a great variety of sampling scenarios, when estimating spatial nonlinear models. In addition, the GMM approach, proves to be greatly useful, due to its simplicity and its computational performance, especially when compared to other commonly used approaches, such as Maximum Likelihood (ML) or Markov Chains Monte Carlo (MCMC) methods (see Chapter 2).

Under the setting of the GMM, it is assumed that the unknown parameter vector $\boldsymbol{\Theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \sigma_{\mathbf{e}}^2)$ satisfy the following moment condition:

$$\mathrm{E}\begin{bmatrix} \mathbf{Z}^{\intercal}\mathbf{u}^* \\ \bar{\mathbf{X}}^{\intercal}\mathbf{u}^* \\ \left(\mathbf{W}\bar{\mathbf{X}}\right)^{\intercal}\mathbf{u}^* \end{bmatrix} = \mathbf{0} \Leftrightarrow \mathrm{E}\left(\mathbf{Z}_+\mathbf{u}^*\right) = \mathbf{0} \tag{106}$$

where $\mathbf{Z} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}]$, as suggested by Kelejian and Prucha (1998). The GMM estimates for the unknown parameter vector, $\boldsymbol{\Theta}$, are obtained by minimizing the objective function:

$$\mathcal{Q}\left(\boldsymbol{\Theta}\right) = \mathbf{u}^{*\intercal}\mathbf{Z}_+\left(\mathbf{Z}_+^{\intercal}\mathbf{Z}_+\right)^{-1}\mathbf{Z}_+^{\intercal}\mathbf{u}^* \tag{107}$$

and the GMM estimator reduces to nonlinear two stages least squares (N2SLS). As the minimization problem in (107) does not have a closed formula, the iterative procedure of Klier and McMillen (2008) is used. In addition, the spatial heteroskedasticity and spatial autocorrelation robust estimator of Kelejian and Prucha (2007) is considered, to overcome potential biases in the estimated asymptotic covariance matrix estimator of the (iterative) GMM estimator. See Section 2.2 in Chapter 2 for details.

The individual gradients, at time t, for the PFRSLPM are:

$$(\boldsymbol{\Gamma}_{\alpha})_{it} = \frac{\partial u_{it}^{*}}{\partial \alpha} = -\phi\left(\frac{\alpha\sum\limits_{j\neq i} w_{ij}Y_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_{i}\boldsymbol{\varphi}_{1} + \sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}\boldsymbol{\varphi}_{2}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}}\right)\frac{\sum\limits_{j\neq i} w_{ij}Y_{jt}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}} \tag{108}$$

$$(\boldsymbol{\Gamma}_{\boldsymbol{\beta}})_{it} = \frac{\partial u_{it}^{*}}{\partial \boldsymbol{\beta}^{\intercal}} = -\phi\left(\frac{\alpha\sum\limits_{j\neq i} w_{ij}Y_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_{i}\boldsymbol{\varphi}_{1} + \sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}\boldsymbol{\varphi}_{2}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}}\right)\frac{\mathbf{X}_{it}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}} \tag{109}$$

$$(\boldsymbol{\Gamma}_{\boldsymbol{\varphi}_{1}})_{it} = \frac{\partial u_{it}^{*}}{\partial \boldsymbol{\varphi}_{1}^{\intercal}} = -\phi\left(\frac{\alpha\sum\limits_{j\neq i} w_{ij}Y_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_{i}\boldsymbol{\varphi}_{1} + \sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}\boldsymbol{\varphi}_{2}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}}\right)\frac{\bar{\mathbf{X}}_{i}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}} \tag{110}$$

$$(\boldsymbol{\Gamma}_{\boldsymbol{\varphi}_{2}})_{it} = \frac{\partial u_{it}^{*}}{\partial \boldsymbol{\varphi}_{2}^{\intercal}} = -\phi\left(\frac{\alpha\sum\limits_{j\neq i} w_{ij}Y_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_{i}\boldsymbol{\varphi}_{1} + \sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}\boldsymbol{\varphi}_{2}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}}\right)\frac{\sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}} \tag{111}$$

$$\begin{aligned}(\boldsymbol{\Gamma}_{\sigma_{\mathbf{e}}})_{it} = \frac{\partial u_{it}^{*}}{\partial \sigma_{\mathbf{e}}} = \phi&\left(\frac{\alpha\sum\limits_{j\neq i} w_{ij}Y_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_{i}\boldsymbol{\varphi}_{1} + \sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}\boldsymbol{\varphi}_{2}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}}\right)\\ &\times\frac{\sigma_{\mathbf{e}}}{(1+\sigma_{\mathbf{e}}^{2})}\left[\frac{\alpha\sum\limits_{j\neq i} w_{ij}Y_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_{i}\boldsymbol{\varphi}_{1} + \sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}\boldsymbol{\varphi}_{2}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}}\right]\end{aligned} \tag{112}$$

with $i, j = 1, 2, \ldots, N$, $i \neq j$ and $t = 1, 2, \ldots, T$. The individual gradients, at time t, for the aPFRSLPM are:

$$(\boldsymbol{\Gamma}_{\alpha})_{it} = \frac{\partial u_{it}^{*}}{\partial \alpha} = -\phi\left(\frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_{i}\boldsymbol{\varphi}_{1} + \sum\limits_{j\neq i} w_{ij}\bar{\mathbf{X}}_{j}\boldsymbol{\varphi}_{2}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}}\right)\frac{\sum\limits_{j\neq i} w_{ij}Y_{jt}}{\sqrt{1+\sigma_{\mathbf{e}}^{2}}} \tag{113}$$

$$(\boldsymbol{\Gamma}_{\boldsymbol{\beta}})_{it} = \frac{\partial u^*_{it}}{\partial \boldsymbol{\beta}^{\intercal}} = -\phi\left(\frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}}\right)\frac{\mathbf{X}_{it}}{\sqrt{1+\sigma^2_{\mathbf{e}}}} \times\left(1 - \frac{\alpha\sum_{j\neq i} w_{ij}Y_{jt}}{\sqrt{1+\sigma^2_{\mathbf{e}}}} \times \frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}}\right) \tag{114}$$

$$\left(\boldsymbol{\Gamma}_{\boldsymbol{\varphi}_1}\right)_{it} = \frac{\partial u^*_{it}}{\partial \boldsymbol{\varphi}_1^{\intercal}} = -\phi\left(\frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}}\right)\frac{\bar{\mathbf{X}}_i}{\sqrt{1+\sigma^2_{\mathbf{e}}}} \times\left(1 - \frac{\alpha\sum_{j\neq i} w_{ij}Y_{jt}}{\sqrt{1+\sigma^2_{\mathbf{e}}}} \times \frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}}\right) \tag{115}$$

$$\left(\boldsymbol{\Gamma}_{\boldsymbol{\varphi}_2}\right)_{it} = \frac{\partial u^*_{it}}{\partial \boldsymbol{\varphi}_2^{\intercal}} = -\phi\left(\frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}}\right)\frac{\sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j}{\sqrt{1+\sigma^2_{\mathbf{e}}}} \times\left(1 - \frac{\alpha\sum_{j\neq i} w_{ij}Y_{jt}}{\sqrt{1+\sigma^2_{\mathbf{e}}}} \times \frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}}\right) \tag{116}$$

$$(\boldsymbol{\Gamma}_{\sigma_{\mathbf{e}}})_{it} = \frac{\partial u^*_{it}}{\partial \sigma_{\mathbf{e}}} = \phi\left(\frac{\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}}\right) \times \frac{\sigma_{\mathbf{e}}}{(1+\sigma^2_{\mathbf{e}})}\left[\frac{\alpha\sum_{j\neq i} w_{ij}Y_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2}{\sqrt{1+\sigma^2_{\mathbf{e}}}} - \frac{\left(\alpha\sum_{j\neq i} w_{ij}Y_{jt}\right)\left(\mathbf{X}_{it}\boldsymbol{\beta} + \bar{\mathbf{X}}_i\boldsymbol{\varphi}_1 + \sum_{j\neq i} w_{ij}\bar{\mathbf{X}}_j\boldsymbol{\varphi}_2\right)^2}{(1+\sigma^2_{\mathbf{e}})^{3/2}}\right] \tag{117}$$

with $i, j = 1, 2, \ldots, N$, $i \neq j$ and $t = 1, 2, \ldots, T$. Note that, having an explicit expression for the gradients significantly increases the performance of the iterative algorithms and/or numerical minimization methods.

Next, the performance of the iterative GMM (iGMM) procedure, presented in this section, is assessed through an insightful empirical application on the proportion of high-tech patents in the U.S. Metropolitan Statistical Areas (MSAs). The models

developed in Section 4.2 are estimated by iGMM and their results are compared with those from the Two-Stages Least Squares (2SLS) estimator for the pooled Linear Spatial Lag Model. The treatment of the unobserved heterogeneity and the estimation of partial effects are addressed as central issues.

4.4. Empirical Application

In this section an empirical application is presented to illustrate the adequacy of the previous iterative Generalized Method of Moments (iGMM) approach to estimate the proposed models for spatial panel fractional responses (the PFRSLPM and the aPFRSLPM). In addition, it addresses and compares the performance of the Two-Stages Least Squares (2SLS) estimator for the pooled Linear Spatial Lag Model (LSLM) and for the Linear Spatial Lag Model with fixed effects (LSLM-FE), as well.

The previous approaches are applied to study the proportion of U.S. origin high-tech patents in 201 U.S. Metropolitan Statistical Areas (MSAs), between 2010 and 2015. The expenditures in research and development (R&D) by source of funding, wages and number of employed individuals by educational level are of particular interest, as well as the corresponding spillover effects across the neighboring MSAs. Unobserved heterogeneity is added to the model specification. As a result, the treatment of the unobserved effects will be addressed as a central issue, once they are likely to be correlated with the included explanatory variables and the spatially lagged proportion. Estimation and inference of the partial effects will focus on the Average Direct Effects (ADEs) and Average Indirect Effects (AIEs).

4.4.1. Data

The dataset containing the relevant variables for this empirical application consists in a combination of datasets from three different data sources. First, the data on U.S. origin high-tech patents was collected from the U.S. Patent and Trademark Office (https://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm) and considers the Patent Technology Monitoring Team (PTMT) reports by

U.S. Metropolitan Statistical Area (MSA), for each patent technology class and for each year. The total number of U.S. origin patents and the total number of U.S. origin high-tech patents are computed, for each year and for each MSA. The U.S. origin high-tech patents include all the patent classes that are mentioned in the "Selected Technology Reports" and are related to "Biotechnology", "Telecommunications", "Electrical Computers, Digital Processing Systems, Information Security, Error/Fault Handling", "Medical Devices" and "Semiconductor Devices and Manufacturing". The proportion of U.S. origin high-tech patents results from the ratio between the total number of U.S. origin high-tech patents and the total number of U.S. origin patents, for each MSA and for each year.

Second, the data on the R&D expenditures at U.S. Colleges and Universities by source of funding – federal government, state and local government, business and institutional – was collected from the Higher Education Research and Development Survey (HERD) – National Science Foundation (https://www.nsf.gov/statistics/srvyherd), for each institution and for each year. This dataset only considers institutions that expended at least 150 thousand dollars in R&D, for a given fiscal year. All the R&D expenditures are adjusted for inflation. The data was aggregated to the MSA level, for each year, based on the information for the ZIP codes of each institution. Changes in the ZIP codes, across different years, were accommodated, as well.

Third, the data on the median real wages and the total number of employed individuals, both by educational level – Ph.D., graduate (except Ph.D.) and non-graduate –, was collected from the American Community Survey (ACS) – Integrated Public Use Microdata Series, IPUMS-USA (https://usa.ipums.org/usa/acs.shtml). According to Ruggles et al. (2019), the ACS replaces the decennial census and provides an annual snapshot of the population and housing characteristics in the U.S. In this empirical application, only the population characteristics are of interest. The ACS consists in the application of a questionnaire, similar to the "long form" census questionnaire, to a sample of individuals, rather than to the whole population. Each individual in the sample is given a personal weight, that indicates how many persons are represented by his/her characteristics. To

obtain the information for the whole population, the data has to be aggregated through the weighted sums of the individuals included in the sample. The same applies if one is interested in obtaining information for the whole population of a given region. However, one cannot directly aggregate the information to the MSA level, once the geographic information included in the ACS refers to the Public Use Microdata Areas (PUMAs), differing from the geographic definition of the MSAs. Hence, the data was initially aggregated to the PUMA level and afterwards converted to the MSA level, using an allocation factor from a crosswalk between the geographic delineation of the PUMAs and the geographic delineation of the MSAs (http://mcdc.missouri.edu/applications/geocorr2018.html). It is also important to note that the real wages by educational level, for each MSA and for each year, are summarized by the median, once they are top-coded.

4.4.2. Descriptive analysis

In this section, a descriptive analysis of the sample of MSAs and the variables included in this empirical application is presented. The relevant outputs are provided in Section A4. In Figure A4.1 the polygons of the MSAs are outlined and the centroids of the included MSAs are indicated by black dots. Figure A4.2 presents the histogram of the dependent variable, the proportion of U.S. origin high-tech patents in the included MSAs. Figure A4.3 displays the spatial distribution of the proportion of U.S. origin high-tech patents in the included MSAs, for each year t, with $t = 2010, 2011, \ldots, 2015$. Figure A4.4 and Figure A4.5 outline the time series of the U.S. origin aggregate patents (all patents and high-tech patents) and the time series of the U.S. origin aggregate patents change in the U.S. MSAs, between the years 2010 and 2015. Table A4.1 summarizes the panel descriptive statistics for the proportion of U.S. origin high-tech patents and for the explanatory variables considered in this empirical application.

The sample of 201 U.S. MSAs included in this analysis are typically concentrated in the eastern part of the U.S. The information loss is due to the combination of the datasets mentioned in Section 4.4.1. In fact, the majority of MSAs that were excluded did not report any information for the HERD dataset, during the years

2010 to 2015. Nevertheless, the sample can be shown to be representative, once the included MSAs comprise more than 75% of the total real GDP[14]. In addition, the included MSAs comprise more than 88% of the total U.S. origin patents and more than 90% of the total U.S. origin high-tech patents. Robustness checks show that the general conclusions did not change much, even if the HERD dataset is not used and all the 374 U.S. MSAs were included.

The sampling distribution of the proportion of U.S. origin high-tech patents in the U.S. MSAs, from 2010 to 2015, resembles an inverted "U"-shaped distribution, moderately skewed to the right. As results will show, this may indicate a small degree of spatial dependence. One particular characteristic of the proportion of U.S. origin high-tech patents is that there are observations at the corners, 0 and 1. Hence, there is a relatively small group of MSAs that focus their patenting activity on high-tech development. However, the value 0 must be interpreted with care. In fact, most of these observations may refer to MSAs with no high-tech development activity. This is because, the majority of the high-tech innovations tend to be patented, as pointed out by Fischer et al. (2009). Even so, some of these observations may correspond to a small group of MSAs that have a relevant high-tech development activity, but their innovations have not been patented yet. In addition, note that having observations at the corner 0 and 1 have important implications in terms of estimation. Considering approaches that apply *ad hoc* transformations to adjust the observations at 0 or 1, are not adequate and do not yield consistent estimates for the parameters of interest (see also section 18.6 of Wooldridge, 2010).

The spatial distribution of the proportion of U.S. origin high-tech patents in the U.S. MSAs is quite heterogeneous. Still, it tends to follow the aggregate pattern of the proportion of U.S. origin high-tech patents (see Figure A4.4 and Figure A4.5). Between 2010 and 2014, the proportion of U.S. origin high-tech patents appears to increase, for the majority of the MSAs. Not surprisingly, this coincides with the period of economic recovery in the U.S., after the turbulent 2000s, with successive economic crisis (energy, housing and financial). However, from 2014 to 2015, the

[14]Information on the real GDP was collected from the Bureau of Economic Analysis (BEA) – U.S. Department of Commerce (`https://usa.ipums.org/usa/acs.shtml`)

proportion of high-tech patents appears to decrease for most of the MSAs. In a 2016 technical report, Antonipillai et al. (2016) show that the share of workers with a bachelor degree or higher have declined for intellectual-property intensive industries, between 2010 and 2015. As the estimation results will show, employment has an important effect on the proportion of U.S. origin high-tech patents.

With regard to the panel descriptive statistics, it is possible to observe that there is no variable that is time-invariant, due to the positive within variation. However, all the R&D expenditures and employment variables have small within variation. In consequence, several complications may arise for the Fixed Effects estimator. See Hahn et al. (2011) for a discussion on this matter, under a non-spatial framework. Other three results are expected. First, the R&D expenditures at U.S. Colleges and Universities (adjusted for inflation) for the U.S. MSAs are, on average, mainly funded by the Federal Government. Two, on average, the Ph.D. workers earn about two times more than the graduate workers (except Ph.Ds) and about four times more than the non-graduate workers. Three, on average, the majority of workers in the MSAs are non-graduates. Though expected, this result proves that structural labor market policies have to be undertaken, in order to increase the potential of Human Capital in the U.S.

Next, the estimation results for the empirical application on the proportion of U.S. origin high-tech patents in the U.S. MSAs, between 2010 and 2015, are presented. The iterative GMM (iGMM) estimator is used to estimate the PFRSLPM and the aPFRSLPM. In addition, the Two-Stage Least Squares (2SLS) estimator is used to estimate the Pooled Linear Spatial Lag Model (Pooled LSLM) and the Linear Spatial Lag model with Fixed Effects (LSLM-FE). Three scenarios to model the unobserved heterogeneity are considered: no device, the Chamberlain-Mundlak device and the Debarsy (2012) device (see also Assumption 4.3). The estimates are compared and discussed in detail. The issues related to the estimation of spatially lagged fractional responses using a linear approach are stressed out.

4.4.3. Estimation results

In this section, the estimation results on empirical application are presented. The relevant outputs are provided in Section B4. Table B4.1 presents the Two-Stages Least Squares (2SLS) estimation results for the Pooled Linear Spatial Lag Model (Pooled LSLM)[15]. Table B4.2 presents the 2SLS estimation results for the Linear Spatial Lag Model with Fixed Effects (LSLM-FE)[16]. Table B4.3 presents the iterative Generalized Method of Moments (iGMM) estimation results for the Panel Fractional Response Spatial Lag Probit Model (PFRSLPM). Table B4.4 presents the iGMM estimation results for the approximate Panel Fractional Response Spatial Lag Probit Model (aPFRSLPM). For the Pooled LSLM, the PFRSLPM and the aPFRSLPM, the treatment of unobserved heterogeneity is approached in three ways. First, with no specific treatment of the unobserved heterogeneity (no device). Second, using the Chamberlain-Mundlak device (CM device). Third, using the Debarsy (2012) device (Debarsy device), as in Assumption 4.3. The PFRSLPM with Debarsy device is the benchmark model. All the estimated models use a row-standardized spatial weighting matrix, $\mathbf{W}$, given by the squared inverse distance between the MSAs, along the lines of Anselin et al. (1997), Acs et al. (2002) and Fischer and Varga (2003). In addition, all the estimation procedures consider the matrix of instruments $\mathbf{Z} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}]$. The time averages of the explanatory variables and their spatial lags were excluded as instruments to avoid perfect colinearity, resulting from small within variation. Table C4.1 presents the estimated time averages for the Average Direct Effects (ADEs) and for the Average Indirect Effects

[15]The Pooled LSLM considers the following model:

$$Y_{it} = \alpha \sum_{j \neq i} w_{ij} Y_{jt} + \sum_{k=1}^{K} \beta_k x_{it,k} + u_{it}, \qquad i, j = 1, 2, \ldots, N,\ t = 1, 2, \ldots, T \tag{118}$$

[16]The LSLM-FE considers the time-demeaned model:

$$\tilde{Y}_{it} = \alpha \sum_{j \neq i} w_{ij} \tilde{Y}_{jt} + \sum_{k=1}^{K} \beta_k \tilde{x}_{it,k} + \tilde{u}_{it}, \qquad i, j = 1, 2, \ldots, N,\ t = 1, 2, \ldots, T \tag{119}$$

with $\tilde{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, and $\tilde{x}_{it,k} = x_{it,k} - \frac{1}{T}\sum_{t=1}^{T} x_{it,k}$, for each $k = 1, 2, \ldots, K$, and $\tilde{u}_{it} = u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it}$.

(AIEs), based on the estimates obtained from the previous models (Pooled LSLM, LSLM-FE, PFRSLPM and aPFRSLPM), with the treatment of the unobserved heterogeneity approached by the Debarsy (2012) device. The estimated ADEs and AIEs are averaged across the temporal dimension, once they did not change much across different years. The standard errors of the estimated ADEs and AIEs for the PFRSLPM and the aPFRSLPM are obtained via simulation (see Bivand, 2019, for details).

The estimation results, across different approaches to unobserved heterogeneity, can be summarized in fourfold. First, in general, the estimates for each of the parameters of interest do not change sign. Exceptions are the variables *Business R & D*, *Median Real Wages Ph.Ds* and *Median Real Wages non-graduates*. Second, the estimated magnitudes significantly differ for the scaled and the unscaled estimates[17]. Third, the individual statistical significance of the parameter estimates tends to change more drastically for the Pooled LSLM. Here, most of the explanatory variables are statistically insignificant, especially when the unobserved heterogeneity is controlled for. To the contrary, under both the PFRSLPM and the aPFRSLPM, only the real wage variables change significance across different approaches to unobserved heterogeneity, becoming statistically insignificant when using the Debarsy (2012) device. Fourth, the iGMM estimates for the PFRSLPM and the iGMM estimates for the aPFRSLPM have similar magnitudes. This is consistent with the findings from the simulation study in Chapter 3. Furthermore, these estimates are identical in terms of sign and statistical significance.

In terms of model adequacy, the Wald tests of joint significance tend to reject the null and the Hansen tests for the validity of over-identifying moment conditions tend to not reject the null. Exceptions are the LSLM with CM device, for the Wald test (the null is not rejected), and the LSLM-FE, for the Hansen test (the null is rejected). Even so, note that, for the case where the unobserved heterogeneity is neglected ("no device"), the estimates are known to be biased and inconsistent.

[17]To ensure comparability with the iGMM estimates for both the PFRSLPM and the aPFRSLPM, the estimates for the Pooled LSLM and the estimates for the LSLM-FE were multiplied by $\left[(2\pi)\left(1+\hat{\sigma}^2_{\mathbf{e}}\right)\right]^{1/2}$, with $\hat{\sigma}^2_{\mathbf{e}}$ obtained from the estimation of the PFRSLPM with Debarsy device (the benchmark model).

Hence, statistical inference is incorrect. Similarly, considering the classic Chamberlain-Mundlak device, also leads to biased estimates and invalid inference. This is because, neighboring factors – spatial lags of the time averaged explanatory variables – are not included as controls for unobserved heterogeneity. These factors are particularly helpful to model the correlation between the unobserved heterogeneity and the spatially lagged responses. As a result, the new time-invariant component, e_i, cannot be independent from the set $(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, \mathbf{Y}_{-i,t})$, for all i and t (see Assumption 4.3). Therefore, the Debarsy (2012) approach to unobserved heterogeneity is preferred to the previous approaches, once it uses the time averages of the explanatory variables and their corresponding spatial lags.

It is also important to note that using a linear approach to model fractional responses comes at the cost of estimated effects and predictions falling outside the admissible interval. In addition, diminishing effects of the explanatory variables over the dependent variable are not accounted for. Consider, for example, the estimation results for the Pooled LSLM with Debarsy (2012) device. The estimate for the variable *Employed Ph.Ds* is negative and close to unity (in absolute value). This means that a unit increase in the number of employed Ph.Ds in the ith MSA, at time t, is predicted to always reduce the proportion of U.S. origin high-tech patents in the ith MSA (and eventually other MSAs, as well), at time t, by approximately one[18]. In this way, a continuous unit increase in the number of employed Ph.Ds in the ith MSA, implies that the predicted proportion of U.S. origin high-tech patents in the ith MSA (and eventually other MSAs, as well) would be less than zero. By construction, this cannot be true. See also Papke and Wooldridge (1996) and Chapter 3 for a discussion on this matter. Nevertheless, Papke and Wooldridge (2008) point out that the linear (non-spatial) model with fixed effects can provide good estimates for the average partial effects of the Fractional Probit model. The estimation results for the LSLM-FE, however, do not support the previous statement. This may be

[18]Considering a Linear Spatial Lag Model, the partial effects matrix for the kth explanatory variable, at time t, is given by:

$$\boldsymbol{\Delta}_{k,t} = (\mathbf{I} - \alpha\mathbf{W})^{-1}\beta_k, \quad \forall k = 1, 2, \ldots, K \text{ and } t = 1, 2, \ldots, T \tag{120}$$

.

due to small within variation of the explanatory variables, but also because the over-identifying moment conditions are not statistically correct.

With regard to the estimates for the spatial lag parameter, $\hat{\alpha}$, they range from moderate to low, when normalized to the closed interval $[-1, 1]$. In addition, they change much across different approaches to unobserved heterogeneity. The estimates for α increase as the classic time averaged explanatory variables are included (Chamberlain-Mundlak device), while decreasing (and becoming statistically insignificant) as the spatial lags of the time averaged explanatory variables are included (Debarsy device). The magnitude of these changes are larger for the iGMM estimator. On this matter, recall that the iGMM estimates for α are not normalized to the closed interval $[-1, 1]$. Their normalization yield:

$\frac{\hat{\alpha}}{\sqrt{(2\pi)\,(1+\hat{\sigma}_{\mathbf{e}}^2)}}$	PFRSLPM	aPFRSLPM
no device	0.046	0.043
CM device	0.166	0.197
Debarsy device	0.060	0.063

As both the PFRSLPM and the aPFRSLPM are more adequate than the linear approaches to model fractional responses and since the Debarsy device accommodates the correlation between the unobserved heterogeneity and the spatially lagged responses in a simple way, one may conclude that the degree of spatial dependence between different MSAs is, in fact, quite low.

In this context, estimating a small value for the spatial lag parameter can be a result of three main factors. One, due to spatial aggregation. Using firm-level data, results show that industries tend to form clusters. Examples are the semiconductor laboratory cluster in Silicon Valley and the biotechnology clusters in Massachusetts, New York, Pennsylvania and New Jersey (see Stuart and Sorenson, 2003). Two, due to strong monopolistic behavior of high-tech industries or high-tech industry clusters (see Gallini, 2002). Many examples can be found, where dominant firms or clusters tend to engage in antitrust practices, to promote their market position over their competitors (see, for example, the recent lawsuits on Intel, Google and Apple, to mention a few). Three, due to space-time lagged effects. The rapid dissemination

of knowledge is severely limited by the U.S. patent protection laws (see Scotchmer and Green, 1990, for a discussion on this matter).

In terms of the estimated partial effects, having estimated a small degree of spatial dependence, the spillovers for the whole spatial system tend to be substantially small. The Average Direct Effects (ADEs) only reflect MSA specific changes in the proportion of U.S. origin high-tech patents, resulting from changes in the values of their corresponding explanatory variables. This is because the feedback effects from the neighboring MSAs will be negligible. Hence, the Average Total Effects (ATEs) for all the explanatory variables will be close to zero. Plus, the Average Total Effects To an observation (ATETs), Average Total Effects From an observation (ATEFs) and the Average Indirect Effects (AIEs) should be interpreted with care, as they are contaminated by the Average Direct Effects (ADEs). Nonetheless, two interesting results can highlighted. First, the Pooled LSLM estimator and the LSLM-FE estimator can produce misleading estimates for the ADEs and AIEs (greater or approximately greater than one, in absolute value). Two, the iGMM estimates for both the PFRSLPM and the aPFRSLPM are quite similar, as expected.

Focusing on the iGMM estimates for the PFRSLPM, one can observe that the R&D expenditures at U.S. Colleges and Universities play an important role in the proportion of U.S. origin high-tech patents among the included MSAs. Results show that both Federal origin R&D and Institutional origin R&D have an ADE of 0.056 (0.054 for the aPFRSLPM), while State and Local origin R&D and Business origin R&D have an ADE of -0.043 and -0.162 (-0.042 and -0.157 for the aPFRSLPM), respectively. To the best of the author knowledge, there is no research that disentangles the different effects of the R&D expenditures at U.S. Colleges and Universities by source of funding. Most of the literature focus on the importance of federal funding (see, for example, Adams et al., 2003; Jaffe et al., 1998). Here, it seems to be useful to stress out the dichotomy between interest in high-tech innovative activity and investment capacity, for the four relevant sources. First, the federal government has interest in high-tech innovative activity and has the capacity to invest. This is because high-tech development increases the "stock" of knowledge and the propensity to innovative, promoting economic growth and welfare. Second, educational

institutions have the interest in high-tech innovative activity, but have limited investment capacity. Colleges and Universities are known to have the know-how and skilled labor. However, own funds are generally insufficient to cover the expenses, especially if they involve intermediate inputs to develop high-tech products. Third, state and local government may have interest in high-tech innovative activity, but also have limited budget. Fourth, businesses may have interest in high-tech innovative activity and investment capacity, but only if the research tends to be profitable for them (see Arora et al., 2017). Therefore, on one hand, one may expect that businesses and state and local R&D funding tends to be aimed at innovative activities in Colleges and Universities that are not related to high-tech development, thus, reducing the proportion of U.S. origin high-tech patents. On the other hand, one may expect that federal government and own institutional funding tends to be aimed at innovative activities that are related to high-tech development, thus, increasing the proportion of U.S. origin high-tech patents.

Employment also plays a significant role in the proportion of U.S. origin high-tech patents among the included MSAs. Results show that both the number of employees with a Ph.D. and the number of non-graduate employees have an ADE of -0.323 and -0.051 (-0.314 and -0.050 for the aPFRSLPM), respectively, while the number of graduate employees have an ADE of 0.105 (0.104 for the aPFRSLPM). The estimated signs for the ADEs of the number of non-graduate and graduate employees are expected. Non-graduate employees tend to be allocated to non-skilled jobs, whereas graduate employees tend to be allocated to skilled jobs or to be part of a research team. However, the sign for the estimated ADE of the number of employees with a Ph.D. is counterintuitive. In fact, one would expect a positive sign for the estimated ADE, once most of the applied works that assess the effects of high skilled labor (not necessarily workers with a Ph.D. degree) find a positive effect on the patent counts (see Acs et al., 2002; Acs and Audretsch, 1989; Anselin et al., 1997, to name a few). A recent work from Roach and Sauermann (2010) on the employment preferences of 400 U.S. Ph.D. students in science and engineering, stressed out two important results. One, there is a prevailing "taste for science" among the Ph.D. students, while being weaker for those that prefer industrial employment over

an academic career. Two, Ph.D. students that prefer industrial employment over an academic career tend to attribute special importance to the access to resources. Thus, regardless of the employment preferences, having a Ph.D. is intrinsically related with the interest to produce knowledge and to innovate. Nevertheless, it is important to point out that the development of high-tech products or services is time-consuming and research-intensive. Therefore, one should not expect a contemporaneous positive effect of increasing the number of employees with a Ph.D. on the high-tech patent counts. Hence, the contemporaneous ADE for the number of employees with a Ph.D. on the proportion of U.S. origin high-tech patents is likely to be negative, at time t, becoming positive in the subsequent periods, as high-tech patents are granted. A similar argument is valid for estimated ADE of the median real wage for the Ph.Ds on the proportion of U.S. origin high-tech patents. Increasing labor costs in high-tech intensive industries or universities does not necessarily result in a contemporaneous positive effect on the high-tech patent counts. Hence, the corresponding contemporaneous ADE is also likely to be negative, at time t, becoming positive in subsequent periods.

4.5. Conclusions

In this chapter the two specifications developed in Chapter 3 to model fractional responses with spatial dependence are extended to the panel data setting. In the same manner, the approach Papke and Wooldridge (2008) to model fractional responses with panel data is extended to the spatial framework. No transformations are applied to the responses and observations at the boundaries, zero and one, are admitted. The time-invariant individual effects are added and are allowed to be correlated with the explanatory variables. The setup for the proposed specifications rely on a set of assumptions that are commonly used in the literature. However, some assumptions may be too restrictive.

The first specification, the Panel Fractional Response Spatial Lag Probit Model (PFRSLPM), extends the approach of Papke and Wooldridge (2008) to the spatial framework and extends FRSLM specification, proposed in Chapter 3, to the panel

data setting. The PFRSLPM combines these approaches in a simple way, while introducing spatial dependence into the specification through a spatial lag of the fractional dependent variable and correlated unobserved heterogeneity. The second specification, the approximate Panel Fractional Response Spatial Lag Probit Model (aPFRSLPM) consists in a first order series approximation of the PFRSLPM around the spatial lag parameter equal to zero. The treatment of the unobserved heterogeneity is addressed by the spatial approach of Debarsy (2012), generalizing the classic approaches of Mundlak (1978) and Chamberlain (1980).

An empirical application on the proportion of U.S. origin high-tech patents in the U.S. Metropolitan Statistical Areas (MSAs), between 2010 and 2015, is also presented. The spatial spillovers of knowledge and innovation in the U.S. MSAs are of particular interest, as well as the MSA specific and neighbor specific effects of R&D expenditures by source of funding – federal government, state and local government, business and institutional – and wages and employment by educational level – Ph.D., graduate (except Ph.D.), non-graduate –.

The iterative Generalized Method of Moments (iGMM) estimator of Klier and McMillen (2008) is used to estimate both the PFRSLPM and the aPFRSLPM. In addition, the Two-Stages Least Squares estimator is also considered to estimate the Pooled Linear Spatial Lag Model (Pooled LSLM) and the Linear Spatial Lag Model with Fixed Effect (LSLM-FE). The estimates are compared, as well as the corresponding average direct effects (ADEs) and average indirect effects (AIEs). The drawbacks of considering a linear approach to model fractional responses are stressed out.

Results show the usefulness of the proposed specifications under the spatial panel setting. The iGMM estimates for both the FRSLPM and aPFRSLPM using the Debarsy (2012) device are more reliable than the estimates for other approaches that do not control for the correlation between the unobserved heterogeneity and the spatially lagged responses. In addition, the iGMM estimates for the aPFRSLPM are quite similar to those for the PFRSLPM. To the contrary, the estimates for the Pooled LSLM and the LSLM-FE do not reflect the true nature of the responses.

The estimated degree of spatial dependence for the proportion of U.S. origin

high-tech patents in the U.S. MSAs is significantly low. Factors such as spatial aggregation, regional monopolistic behavior of clustered high-tech industries or the absence of space-time effects, may help to explain this phenomenon. As a result, the feedback effects of the neighboring MSAs are extremely low and the usual measures for the partial effects are contaminated by the ADEs.

Federal government and institutional funding of the R&D expenditures at U.S. Colleges and Universities appear to have a direct positive effect on the proportion of U.S. origin high-tech patents in the U.S. MSAs. To the contrary, state and local government and businesses funding have a direct negative effect. This may be due to the lack of investment capacity or to the lack of interest from private companies on non-profitable development of high-tech innovations.

The number of employees with a Ph.D. and the number of non-graduate employees appear to have a direct negative effect on the proportion of U.S. origin high-tech patents in the U.S. MSAs. To the contrary, the number of graduate employees appear to have a direct positive effect. All the estimated signs are expected, with exception to the estimated sign for the number of employees with a Ph.D. While non-graduates tend to execute non-skilled jobs, graduates and doctorates tend to execute skilled jobs or be part of a research team. Introducing dynamic effects may resolve this issue.

It would be of interest to test these findings using more disaggregated data, at the County-level or even at the firm and/or the University level. In addition, it would be important to assess the properties of the iGMM estimator for the PFRSLPM and the iGMM estimator for the aPFRSLPM, through an extensive simulation study. The adequacy of Assumption 4.3 should be addressed as a central issue.

APPENDIXES

A4. Descriptive Analysis

FIGURE A4.1: Centroids of the U.S. Metropolitan Statistical Areas included in the empirical application

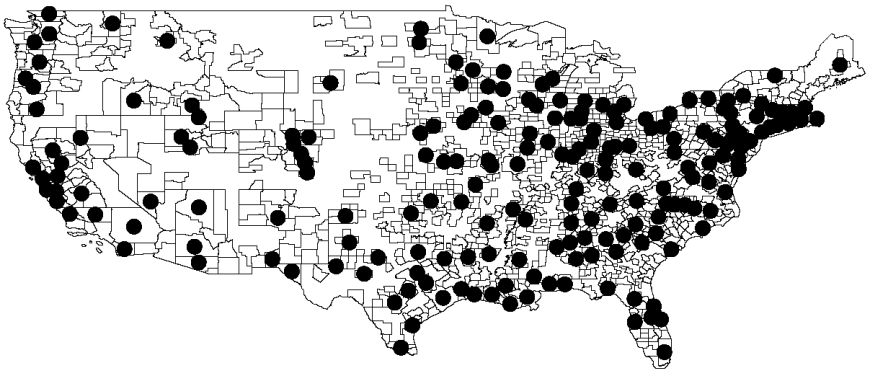


FIGURE A4.2: Empirical distribution of the proportion of High-Tech Patents in the U.S. Metropolitan Statistical Areas, between the years 2010 and 2015

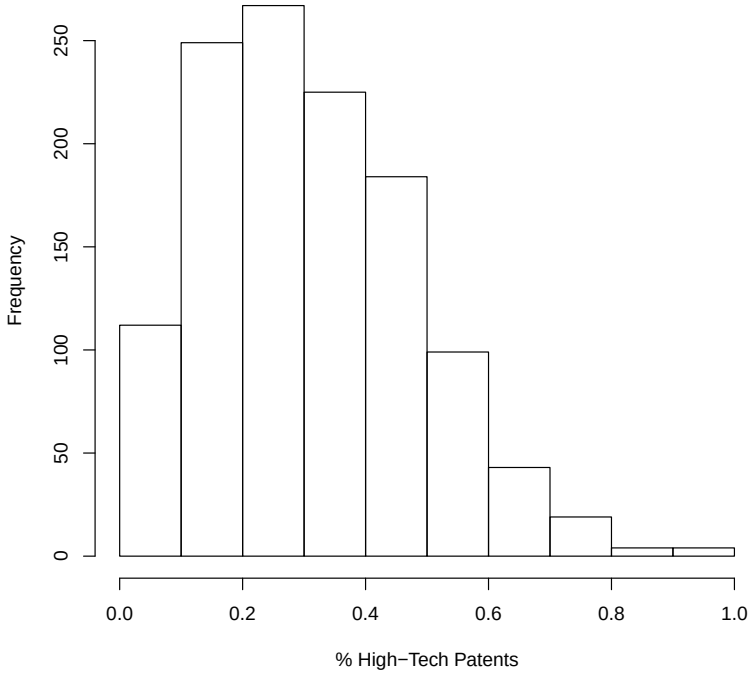


FIGURE A4.3: Spatial distribution of the proportion of High-Tech Patents in the U.S. Metropolitan Statistical Areas, for each year

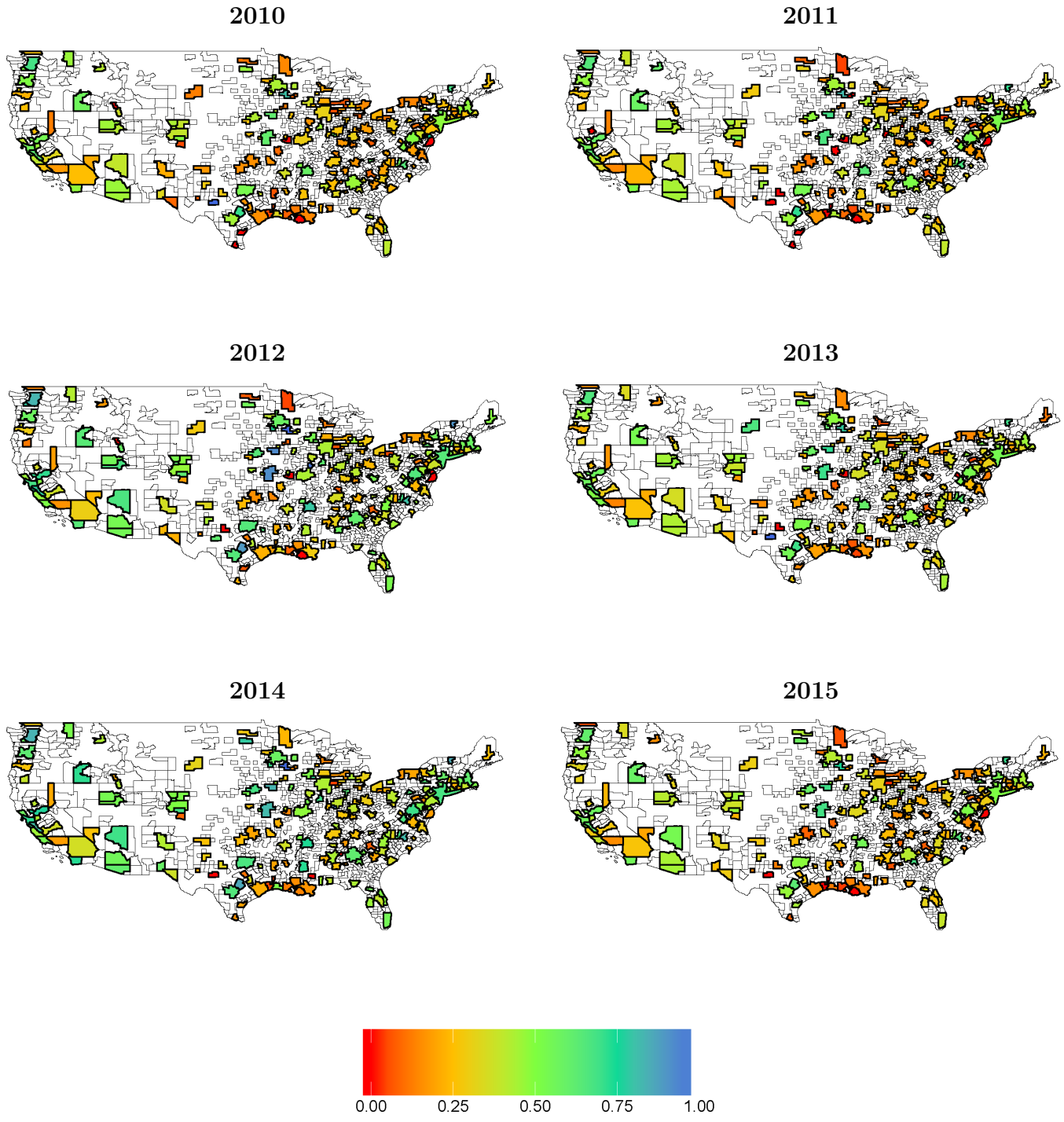


FIGURE A4.4: Time series of the aggregate patents (all patents and high-tech patents) in the U.S. Metropolitan Statistical Areas, between the years 2010 and 2015

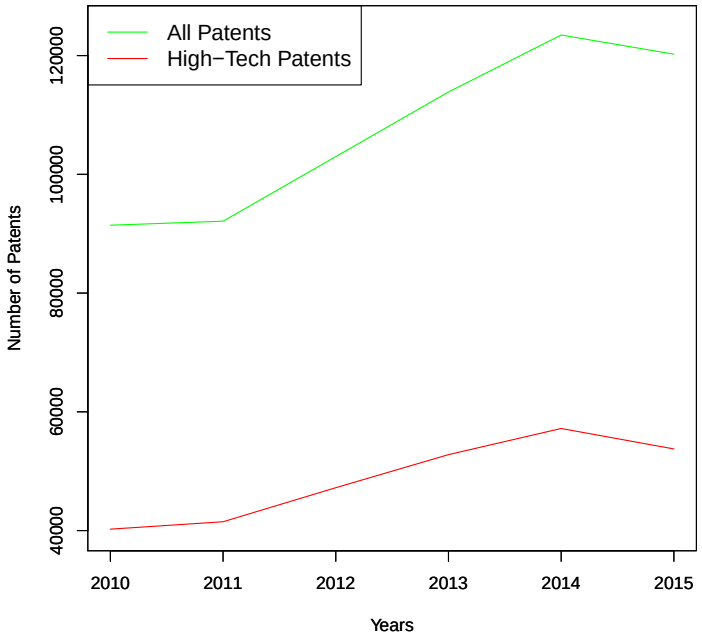


FIGURE A4.5: Time series of the aggregate change in patents (all patents and high-tech patents) in the U.S. Metropolitan Statistical Areas, between the years 2010 and 2015

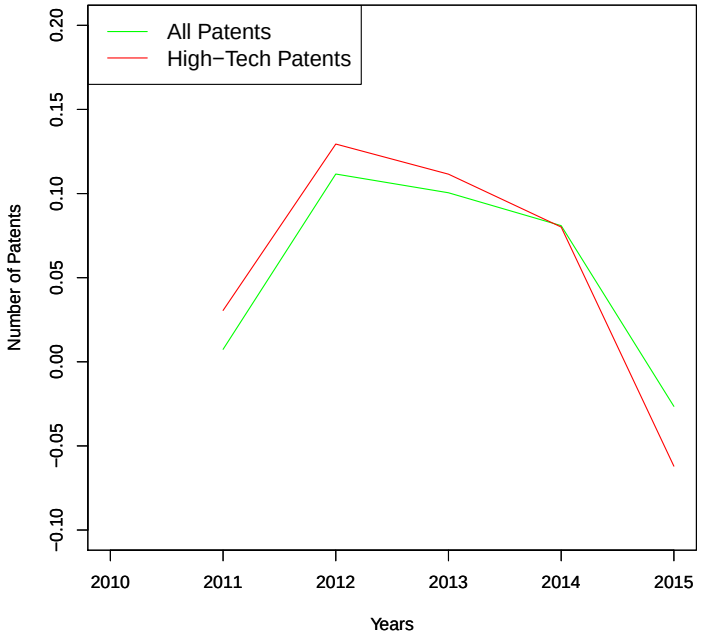


TABLE A4.1: Panel descriptive statistics for the variables included in the empirical application

Variable		**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs.**
% High-Tech Patents	overall	0.312	0.172	0.000	1.000	$NT = 1206$
	between		0.153	0.025	0.790	$N = 201$
	within		0.081	-0.105	1.070	$T = 6$
Federal R & D (inflation adj.)	overall	0.176	0.345	0.000003	2.588	$NT = 1{,}206$
(in millions of $)	between		0.345	0.00003	2.494	$N = 201$
	within		0.026	-0.135	0.386	$T = 6$
State and Local R & D (inflation adj.)	overall	0.017	0.036	0.000	0.393	$NT = 1{,}206$
(in millions of $)	between		0.035	0.0000005	0.338	$N = 201$
	within		0.008	-0.057	0.166	$T = 6$
Business R & D (inflation adj.)	overall	0.016	0.036	0.000	0.311	$NT = 1{,}206$
(in millions of $)	between		0.034	0.000	0.267	$N = 201$
	within		0.010	-0.059	0.158	$T = 6$
Institutional R & D (inflation adj.)	overall	0.066	0.112	0.000	1.138	$NT = 1{,}206$
(in millions of $)	between		0.109	0.000	0.917	$N = 201$
	within		0.029	-0.204	0.356	$T = 6$
Median Real Wages Ph.Ds	overall	6.315	2.384	0.000	18.120	$NT = 1{,}206$
(in millions of $)	between		1.915	0.971	11.371	$N = 201$
	within		1.425	-0.382	14.266	$T = 6$
Median Real Wages Graduates	overall	3.543	1.513	0.000	8.084	$NT = 1{,}206$
(in millions of $)	between		1.403	0.255	7.121	$N = 201$
	within		0.574	-0.017	5.640	$T = 6$
Median Real Wages Non-graduates	overall	1.740	0.865	0.000	4.056	$NT = 1{,}206$
(in millions of $)	between		0.804	0.085	3.690	$N = 201$
	within		0.324	-0.174	3.013	$T = 6$
Employed Ph.Ds	overall	0.019	0.042	0.0004	0.459	$NT = 1{,}206$
(in millions of individuals)	between		0.042	0.001	0.431	$N = 201$
	within		0.003	-0.016	0.047	$T = 6$
Employed Graduates	overall	0.155	0.321	0.006	3.603	$NT = 1{,}206$
(in millions of individuals)	between		0.321	0.007	3.345	$N = 201$
	within		0.020	-0.088	0.413	$T = 6$
Employed Non-graduates	overall	0.310	0.526	0.022	5.254	$NT = 1{,}206$
(in millions of individuals)	between		0.527	0.023	5.165	$N = 201$
	within		0.016	0.201	0.442	$T = 6$

B4. Estimation outputs

TABLE B4.1: Two-Stages Least Squares estimation results for the Pooled Linear Spatial Lag Model (Pooled LSLM) with no treatment of the unobserved heterogeneity (no device), with the Chamberlain-Mundlak device (CM device) and with the Debarsy (2012) device (Debarsy device)

	Dependent variable: % High-Tech Patents		
	(no device)	(CM device)	(Debarsy device)
Intercept	0.162***	0.124***	0.205***
	[0.413]	[0.313]	[0.515]
	(0.023)	(0.027)	(0.030)
Federal R & D (inflation adj.)	0.024	0.084	0.090
	[0.062]	[0.212]	[0.226]
	(0.031)	(0.207)	(0.194)
State & Local R & D (inflation adj.)	−0.255	−0.192	−0.180
	[−0.651]	[−0.485]	[−0.452]
	(0.171)	(0.662)	(0.281)
Business R & D (inflation adj.)	0.468**	−0.472	−0.485
	[1.194]	[−1.196]	[−1.220]
	(0.221)	(0.587)	(0.318)
Institutional R & D (inflation adj.)	0.067	0.029	0.047
	[0.172]	[0.072]	[0.119]
	(0.090)	(0.234)	(0.212)
Median Real Wages Ph.Ds	−0.001	−0.007*	−0.007*
	[−0.002]	[−0.017]	[−0.017]
	(0.003)	(0.004)	(0.004)
Median Real Wages Graduates	0.034***	0.001	0.001
	[0.086]	[0.003]	[0.003]
	(0.008)	(0.014)	(0.014)
Median Real Wages Non-graduates	−0.052***	0.003	0.002
	[−0.134]	[0.008]	[0.005]
	(0.012)	(0.025)	(0.024)
Employed Ph.Ds	−1.513*	−0.880*	−0.981***
	[−3.858]	[−2.228]	[−2.466]
	(0.876)	(0.483)	(0.013)
Employed Graduates	0.494***	0.034	0.019
	[1.259]	[0.087]	[0.049]
	(0.160)	(0.405)	(0.182)
Employed Non-graduates	−0.157***	−0.002	−0.004
	[−0.399]	[−0.006]	[−0.011]
	(0.046)	(0.446)	(0.182)
α	0.373***	0.381*	0.201
	[0.951]	[0.965]	[0.504]
	(0.053)	(0.209)	(0.197)
Observations	1,206	1,206	1,206
# Instruments	35	35	35
Wald test (overall sig.)	137.906	9.152	317.094×10^2
(p-value)	≈(0.000)	(0.608)	≈(0.000)
Hansen's J test	4.452	4.573	1.951
(p-value)	≈(1.000)	(0.971)	(0.377)

NOTE: Scaled estimates in brackets, to ensure comparability with the estimates from the PFRSLPM and aPFRSLPM. The estimates are multiplied by $\sqrt{2\pi}\sqrt{1+\hat{\sigma}^2_{\mathbf{e}}}$, with $\hat{\sigma}^2_{\mathbf{e}}$ obtained from the estimation of the PFRSLPM with Debarsy device. Standard errors in parentheses. Time effects are modeled by a time trend. Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively.

TABLE B4.2: Two-Stages Least Squares estimation results for the Linear Spatial Lag Model with Fixed Effects (LSLM-FE)

	Dependent variable: % High-Tech Patents
Federal R & D (inflation adj.)	0.097 [0.245] (0.106)
State & Local R & D (inflation adj.)	−0.167 [−0.419] (0.337)
Business R & D (inflation adj.)	−0.500* [−1.257] (0.297)
Institutional R & D (inflation adj.)	0.068 [0.172] (0.119)
Median Real Wages Ph.Ds	−0.006*** [−0.016] (0.002)
Median Real Wages Graduates	0.001 [0.004] (0.007)
Median Real Wages Non-graduates	0.001 [0.002] (0.013)
Employed Ph.Ds	−1.094 [−2.751] (1.485)
Employed Graduates	0.003 [0.006] (0.232)
Employed Non-graduates	−0.006 [−0.016] (0.226)
α	−0.002 [−0.005] (0.070)
Observations	1,206
# Instruments	35
Wald test (overall sig.) (*p*-value)	25.401 ≈(0.008)
Hansen's J test (*p*-value)	151.473 ≈(0.000)

NOTE: Scaled estimates in brackets, to ensure comparability with the estimates from the PFRSLPM and aPFRSLPM. The estimates are multiplied by $\sqrt{2\pi}\sqrt{1+\hat{\sigma}^2_{\mathbf{e}}}$, with $\hat{\sigma}^2_{\mathbf{e}}$ obtained from the estimation of the PFRSLPM with Debarsy device. Standard errors in parentheses. Time effects are modeled by a time trend. Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively.

TABLE B4.3: Iterative Generalized Method of Moments estimation results for the Panel Fractional Response Spatial Lag Probit Model (PFRSLPM) with no treatment of the unobserved heterogeneity (no device), with the Chamberlain-Mundlak device (CM device) and with the Debarsy (2012) device (Debarsy device)

	Dependent variable: % High-Tech Patents		
	(no device)	(CM device)	(Debarsy device)
Intercept	-0.649^{***}	-0.309^{**}	-0.496^{***}
	(0.048)	(0.130)	(0.131)
Federal R & D (inflation adj.)	0.158^{***}	0.072^{***}	0.165^{***}
	(0.036)	(0.024)	(0.023)
State & Local R & D (inflation adj.)	-0.253^{***}	-0.084	-0.126^{***}
	(0.001)	(0.175)	(0.034)
Business R & D (inflation adj.)	0.481^{***}	-0.405^{***}	-0.479^{***}
	(0.003)	(0.079)	(0.010)
Institutional R & D (inflation adj.)	0.107^{***}	0.328^{**}	0.164^{**}
	(0.007)	(0.135)	(0.065)
Median Real Wages Ph.Ds	-0.006	0.098^{***}	-0.053^{**}
	(0.011)	(0.028)	(0.024)
Median Real Wages Graduates	0.111^{***}	0.049	0.071
	(0.024)	(0.084)	(0.075)
Median Real Wages Non-graduates	-0.167^{***}	-0.183^{**}	0.040
	(0.033)	(0.090)	(0.053)
Employed Ph.Ds	-1.505^{***}	-0.816^{***}	-0.952^{***}
	(0.001)	(0.014)	(0.014)
Employed Graduates	0.543^{***}	0.376^{***}	0.311^{***}
	(0.010)	(0.074)	(0.080)
Employed Non-graduates	-0.162^{***}	-0.201^{***}	-0.150^{***}
	(0.023)	(0.044)	(0.043)
α	0.117^{***}	0.421^{***}	0.151^{**}
	(0.015)	(0.078)	(0.063)
$\sigma^2_{\mathbf{e}}$	0.035	0.020	0.006
	(–)	(–)	(–)
Observations	1,206	1,206	1,206
# Instruments	35	35	35
Wald test (overall sig.)	459.980×10^3	240.543×10^6	237.364×10^5
(p-value)	$\approx$(0.000)	$\approx$(0.000)	$\approx$(0.000)
Hansen's J test	4.694	3.884	1.858
(p-value)	$\approx$(1.000)	(0.973)	(0.173)

NOTE: Robust standard errors in parentheses, based on Kelejian and Prucha (2007). Time effects are modeled by a time trend. Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively.

TABLE B4.4: Iterative Generalized Method of Moments estimation results for the approximate Panel Fractional Response Spatial Lag Probit Model (aPFRSLPM) with no treatment of the unobserved heterogeneity (no device), with the Chamberlain-Mundlak device (CM device) and with the Debarsy (2012) device (Debarsy device)

	Dependent variable: % High-Tech Patents		
	(no device)	(CM device)	(Debarsy device)
Intercept	-0.650^{***}	-0.383^{***}	-0.509^{***}
	(0.048)	(0.133)	(0.133)
Federal R & D (inflation adj.)	0.160^{***}	0.083^{***}	0.164^{***}
	(0.036)	(0.023)	(0.024)
State & Local R & D (inflation adj.)	-0.252^{***}	-0.128	-0.127^{***}
	(0.001)	(0.170)	(0.034)
Business R & D (inflation adj.)	0.481^{***}	-0.432^{***}	-0.477^{***}
	(0.003)	(0.078)	(0.011)
Institutional R & D (inflation adj.)	0.107^{***}	0.324^{**}	0.167^{***}
	(0.007)	(0.134)	(0.065)
Median Real Wages Ph.Ds	-0.006	0.091^{***}	-0.055^{**}
	(0.011)	(0.027)	(0.024)
Median Real Wages Graduates	0.111^{***}	0.042	0.069
	(0.025)	(0.082)	(0.076)
Median Real Wages Non-graduates	-0.168^{***}	-0.192^{**}	0.040
	(0.034)	(0.089)	(0.053)
Employed Ph.Ds	-1.505^{***}	-0.822^{***}	-0.952^{***}
	(0.001)	(0.014)	(0.014)
Employed Graduates	0.542^{***}	0.367^{***}	0.315^{***}
	(0.010)	(0.072)	(0.080)
Employed Non-graduates	-0.162^{***}	-0.192^{***}	-0.152^{***}
	(0.023)	(0.042)	(0.043)
α	0.111^{***}	0.503^{***}	0.160^{**}
	(0.015)	(0.071)	(0.064)
$\sigma^2_{\mathbf{e}}$	0.048	0.039	0.019
	(–)	(–)	(–)
Observations	1,206	1,206	1,206
# Instruments	35	35	35
Wald test (overall sig.)	447.372×10^3	210.529×10^6	225.042×10^5
(p-value)	$\approx$(0.000)	$\approx$(0.000)	$\approx$(0.000)
Hansen's J test	4.712	3.817	1.861
(p-value)	$\approx$(1.000)	(0.975)	(0.172)

NOTE: Robust standard errors in parentheses, based on Kelejian and Prucha (2007). Time effects are modeled by a time trend. Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively.

C4. Partial effects

TABLE C4.1: Estimated time averages for the Average Direct Effects (ADEs) and Average Indirect Effects (AIEs), over the proportion of U.S. origin High-Tech patents in the U.S. Metropolitan Statistical Areas, based on the regression estimates for the Pooled Linear Spatial Lag Model (Pooled LSLM) with Debarsy device, the Linear Spatial Lag Model with Fixed Effects (LSLM-FE), the Panel Fractional Response Spatial Lag Probit Model (aPFRSLPM) with Debarsy device and the approximate Panel Fractional Response Spatial Lag Probit Model (aPFRSLPM) with the Debarsy device

	Pooled LSLM		**LSLM-FE**		**PFRSLPM**		**aPFRSLPM**	
	ADE	AIE	ADE	AIE	ADE	AIE	ADE	AIE
Federal R & D (inflation adj.)	0.090	−0.072	0.097	−0.081	0.056***	−0.046***	0.054***	−0.045***
	(–)	(–)	(–)	(–)	(0.008)	(0.006)	(0.008)	(0.006)
State & Local R & D (inflation adj.)	−0.181	0.143	−0.167	0.139	−0.043***	0.035***	−0.042***	0.034***
	(–)	(–)	(–)	(–)	(0.011)	(0.009)	(0.011)	(0.009)
Business R & D (inflation adj.)	−0.487	0.386	−0.500	0.417	−0.162***	0.134***	−0.157***	0.129***
	(–)	(–)	(–)	(–)	(0.004)	(0.003)	(0.004)	(0.003)
Institutional R & D (inflation adj.)	0.048	−0.038	0.068	−0.057	0.056***	−0.046***	0.055***	−0.045***
	(–)	(–)	(–)	(–)	(0.021)	(0.017)	(0.021)	(0.017)
Median Real Wages PhDs	−0.007	0.005	−0.006	0.005	−0.018**	0.015**	−0.018**	0.015**
	(–)	(–)	(–)	(–)	(0.008)	(0.007)	(0.008)	(0.007)
Median Real Wages Graduates	0.001	−0.001	0.001	−0.001	0.024	−0.020	0.023	−0.019
	(–)	(–)	(–)	(–)	(0.025)	(0.020)	(0.022)	(0.018)
Median Real Wages Non-graduates	0.002	−0.002	0.001	−0.001	0.014	−0.011	0.013	−0.011
	(–)	(–)	(–)	(–)	(0.019)	(0.015)	(0.017)	(0.014)
Employed PhDs	−0.985	0.780	−1.094	0.912	−0.323***	0.266***	−0.314***	0.258***
	(–)	(–)	(–)	(–)	(0.005)	(0.004)	(0.006)	(0.005)
Employed Graduates	0.020	−0.015	0.003	−0.002	0.105***	−0.087***	0.104***	−0.085***
	(–)	(–)	(–)	(–)	(0.025)	(0.021)	(0.028)	(0.023)
Employed Non-graduates	−0.004	0.003	−0.006	0.005	−0.051***	0.042***	−0.050***	0.041***
	(–)	(–)	(–)	(–)	(0.014)	(0.012)	(0.015)	(0.012)

NOTE: Unscaled estimates for Pooled LSLM and LSLM-FE. Simulated standard errors in parentheses. The simulated standard errors are based on 1000 replications and were separately generated for the PFRSLPM and the aPFRSLPM. The simulated values were drawn from a Multivariate Normal distribution with mean vector given by the vector of estimates and covariance matrix given by the estimated spatial HAC robust covariance matrix (Kelejian and Prucha, 2007). On each replication, both the ADE and the AIE were computed for each explanatory variable. The simulated standard errors are given by $\text{se}(\widehat{\overline{ADE}}) = \sqrt{(1/1000)\sum_{s=1}^{1000}(ADE_s^* - \overline{ADE}^*)^2}$ and $\text{se}(\widehat{\overline{AIE}}) = \sqrt{(1/1000)\sum_{s=1}^{1000}(AIE_s^* - \overline{AIE}^*)^2}$, where $\overline{ADE}^*$ and $\overline{AIE}^*$ are, respectively, the simulated mean for the ADE and the simulated mean for the AIE. This procedure is based on the function impacts() from the R package spdep (Bivand, 2019). Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively.

Chapter 5

Concluding remarks

This thesis contributes to the Spatial Econometrics literature by addressing the issues related to the estimation and interpretation of spatial models with dichotomous and fractional dependent variables. Three essays were presented. The first essay, *The inversion of the spatial lag operator in binary choice models: fast computation and a closed formula approximation*, proposes a new method to approximate the inverse of the spatial lag operator, used in the estimation of spatially lagged models for dichotomous outcomes. Considering the series expansion of the inverse, a "long run" spatial weighting matrix is used to approximate the second and higher order powers. The computational advantages of this approximation are demonstrated. The "long run" spatial weighting matrix can be written by known quantities, which can be obtained from the untransformed spatial weighting matrix. In this way, the inverse of the spatial lag operator can be approximated by sums of known quantities, requiring minimal computational burden. Estimation is addressed by a well-known iterative Generalized Method of Moments (GMM) procedure. A variant of this procedure is also presented. It considers the proposed approximation method to approximate the gradients of the original iterative GMM procedure. In an extensive Monte Carlo simulation study, the performance of the proposed approximation method is assessed and compared with that of other methods developed in the literature. Of interest is the computational time and accuracy to compute the inverse of the spatial lag operator and related matrix operations. The proposed approximation method is shown to be particularly useful, especially under scenarios with large and dense spatial weighting matrices. The statistical properties of the iterative GMM estimator with approximated gradients are also assessed

and compared with the original iterative GMM estimator and a GMM estimator for a linearized spatially lagged model for dichotomous outcomes. The proposed iterative GMM estimator with approximated gradients is shown to produce accurate estimates, especially for the case where the degree of spatial dependence is low or moderate. Also, it proves to be more precise than the original iterative GMM estimator for the majority of sampling scenarios. The adequacy of the proposed iterative GMM estimator is also assessed in an empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas (MSAs). Results found a moderately high degree of spatial dependence in regional competitiveness and significant effects of environmental variables on the probability of a given MSA to be competitive.

The second essay, *Fractional responses with spatial dependence*, proposes two specifications to model spatially dependent fractional responses. The first specification, the Fractional Response Spatial Lag Model (FRSLM), extends the seminal approach of Papke and Wooldridge (1996) to spatial frameworks and generalizes the recent approach of Xu and Lee (2015a) to accommodate responses defined in the closed interval $[0, 1]$. The second specification, the approximate Fractional Response Spatial Lag Model (aFRSLM), consists in a first order series approximation of the FRSLM, around the spatial lag parameter equal to zero. Also, it allows to write the FRSLM as an approximate reduced form. The partial effects for both the FRSLM and aFRSLM are deduced. The partial effects for the aFRSLM can be interpreted as approximate measures for policy changes, once they are given by nonlinear functions of the exogenous explanatory variables. Here, again, estimation is addressed by a well-known iterative Generalized Method of Moments (GMM) procedure. An extensive Monte Carlo simulation study showed the adequacy of both specifications in terms of accuracy and precision. The GMM estimates obtained for both the FRSLM and aFRSLM tend to be quite similar. Interestingly, the GMM estimator for the aFRSLM performed better for scenarios with a high degree of spatial dependence and denser spatial weighting matrices. It also performed better in the estimation of Average Indirect Effects (AIEs).

The third essay, *Unobserved heterogeneity in spatial panel data models for fractional responses: an application to the proportion of high-tech patents in the U.S. Metropolitan Statistical Areas*, extends the second essay to the panel data setting. Time invariant unobserved effects are added to the specifications and allowed to be correlated with the explanatory variables. The spatial approach of Debarsy (2012) to the unobserved heterogeneity is considered. The partial effects are deduced. The adequacy of the proposed extensions are assessed in an empirical application on the knowledge and innovation spillovers in the U.S. MSAs, between 2010 and 2015. Results found a low degree of spatial dependence in the proportion of U.S. origin high-tech patents and significant effects of both R&D expenditures at U.S. Colleges and Universities and high skilled employment on the regional high-tech patenting activity. The set of underlying assumptions may be too restrictive, but the proposed extensions tend to perform better when considering the Debarsy (2012) device.

The simple and intuitive approaches developed in this thesis can be used to address the specification and estimation of spatial models with ordered, multinomial, limited or count dependent variables. In addition, they can also be used in other frameworks, where the treatment of spatial dependence and spatial heterogeneity is a central issue.

The findings and brief discussions on the empirical applications presented in this thesis provide directions to future research in regional competitiveness and on the subject of knowledge and innovation.

Bibliography

Acs, Z. J., L. Anselin, and A. Varga (2002): "Patents and innovation counts as measures of regional production of new knowledge," *Research Policy*, 31, 1069–1085.

Acs, Z. J. and D. B. Audretsch (1988): "Innovation in large and small firms: an empirical analysis," *The American Economic Review*, 78, 678–690.

——— (1989): "Patents as a measure of innovative activity," *Kyklos*, 42, 171–180.

Acs, Z. J., D. B. Audretsch, and M. P. Feldman (1994): "R & D spillovers and recipient firm size," *The Review of Economics and Statistics*, 76, 336–340.

Adams, J. D., E. P. Chiang, and J. L. Jensen (2003): "The influence of federal laboratory R&D on industrial research," *Review of Economics and Statistics*, 85, 1003–1020.

Agarwal, D. K., A. E. Gelfand, and S. Citron-Pousty (2002): "Zero-inflated models with application to spatial count data," *Environmental and Ecological statistics*, 9, 341–355.

Anselin, L. (1988): *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic.

——— (2007): "Spatial Econometrics," in *A Companion to Theoretical Econometrics*, ed. by B. Baltagi, Blackwell Publishing Ltd, 310–330.

Anselin, L., A. Varga, and Z. Acs (1997): "Local geographic spillovers between university research and high technology innovations," *Journal of Urban Economics*, 42, 422–448.

Antonipillai, J., M. K. Lee, R. Rubinovitz, D. Langdon, F. Yu, W. Hawk, A. C. Marco, A. A. Toole, and A. Tesfayesus (2016): "Intellectual Property and the U.S. Economy: 2016 Update," Tech. rep., United States Patent and Trademark Office.

Arbia, G. (2014): *A Primer for Spatial Econometrics: With Applications in R*, Springer.

Arora, A., S. Belenzon, and L. Sheer (2017): "Back to basics: why do firms invest in research?" Tech. rep., NBER Working Paper No. 23187.

Arrow, K. (1962a): *Economic welfare and the allocation of resources for invention*, Princeton: Princeton University Press.

Arrow, K. J. (1962b): "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 29, 155–173.

Audretsch, D. B. and M. P. Feldman (1996): "R&D spillovers and the geography of innovation and production," *The American Economic Review*, 86, 630–640.

Audretsch, D. B. and M. Vivarelli (1996): "Firms size and R&D spillovers: Evidence from Italy," *Small Business Economics*, 8, 249–258.

Autant-Bernard, C. (2001): "Science and knowledge flows: evidence from the French case," *Research Policy*, 30, 1069–1078.

Baltagi, B. (2013): *Econometric Analysis of Panel Data*, Chichester: John Wiley & Sons, 5th ed.

Baltagi, B. H., J. P. LeSage, and R. K. Pace (2016): *Spatial econometrics: qualitative and limited dependent variables*, Emerald Group Publishing.

Becker, W. and J. Dietz (2004): "R&D cooperation and innovation activities of firms—evidence for the German manufacturing industry," *Research Policy*, 33, 209–223.

Bernstein, D. S. (2009): *Matrix Mathematics: theory, facts, and formulas*, Princeton University Press.

Beron, K. J., J. C. Murdoch, and W. P. M. Vijverberg (2003): "Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol," *The Review of Economics and Statistics*, 85, 286–297.

Beron, K. J. and W. P. M. Vijverberg (2004): "Probit in a Spatial Context: A Monte Carlo Analysis," in *Advances in Spatial Econometrics: Methodology, Tools and Applications*, ed. by R. J. G. M. F. Luc Anselin and S. J. Rey, Berlin, Heidelberg: Springer Berlin Heidelberg, 169–195.

Bhat, C. R. (2011): "The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models," *Transportation Research Part B: Methodological*, 45, 923–939.

Billé, A. G. (2013): "Computational issues in the estimation of the spatial probit model: A comparison of various estimators," *The Review of Regional Studies*, 43, 131–154.

Bivand, R. (2019): "Package 'spdep'. Spatial dependence: weighting schemes, statistics and models," *R package version*.

Blundell, R. and R. L. Matzkin (2014): "Control functions in nonseparable simultaneous equations models," *Quantitative Economics*, 5, 271–295.

Bolduc, D., B. Fortin, and S. Gordon (1997): "Multinomial probit estimation of spatially interdependent choices: an empirical comparison of two new techniques," *International Regional Science Review*, 20, 77–101.

Calabrese, R. and J. A. Elkink (2014): "Estimators of binary spatial autoregressive models: A Monte Carlo study," *Journal of Regional Science*, 54, 664–687.

Cameron, A. C. and P. K. Trivedi (2013): *Regression Analysis of Count Data*, Econometric Society Monographs, Cambridge University Press, 2 ed.

Case, A. (1992): "Neighborhood influence and technological change," *Regional Science and Urban Economics*, 22, 491–508.

Case, A. C., H. S. Rosen, and J. R. Hines (1993): "Budget spillovers and fiscal policy interdependence: Evidence from the states," *Journal of Public Economics*, 52, 285–307.

Cassiman, B. and R. Veugelers (2002): "R&D cooperation and spillovers: some empirical evidence from Belgium," *The American Economic Review*, 92, 1169–1184.

Chakir, R. and O. Parent (2009): "Determinants of land use changes: A spatial multinomial probit approach," *Papers in Regional Science*, 88, 327–344.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 47, 225–238.

Cincera, M. (1997): "Patents, R&D, and technological spillovers at the firm level: some evidence from econometric count models for panel data," *Journal of Applied Econometrics*, 12, 265–280.

Cliff, A. D. and J. K. Ord (1981): *Spatial Processes: Models & Applications*, London: Pion.

Conley, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 92, 1–45.

Cook, D. O., R. Kieschnick, and B. D. McCullough (2008): "Regression analysis of proportions in finance with self selection," *Journal of Empirical Finance*, 15, 860–867.

Cressie, N. A. C. (2015): *Statistics for Spatial Data*, New York: Wiley.

Debarsy, N. (2012): "The Mundlak approach in the spatial Durbin panel data model," *Spatial Economic Analysis*, 7, 109–131.

Duan, N. (1983): "Smearing estimate: a nonparametric retransformation method," *Journal of the American Statistical Association*, 78, 605–610.

Dudensing, R. M. and D. L. Barkley (2010): "Competitiveness of Southern Metropolitan Areas: The Role of New Economy Policies," *The Review of Regional Studies*, 40, 197–226.

Elhorst, J. P. (2014): *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer.

Fagerberg, J. (1988): "International Competitiveness," *The Economic Journal*, 98, 355–374.

——— (1996): "Technology and competitiveness," *Oxford Review of Economic Policy*, 12, 39–51.

Ferrari, S. and F. Cribari-Neto (2004): "Beta regression for modelling rates and proportions," *Journal of Applied Statistics*, 31, 799–815.

Fischer, M. M. and D. A. Griffith (2008): "Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union," *Journal of Regional Science*, 48, 969–989.

Fischer, M. M., T. Scherngell, and E. Jansenberger (2009): "Geographic localisation of knowledge spillovers: evidence from high-tech patent citations in Europe," *The Annals of Regional Science*, 43, 839–858.

Fischer, M. M. and A. Varga (2003): "Spatial knowledge spillovers and university research: Evidence from Austria," *The Annals of Regional Science*, 37, 303–322.

Fiva, J. H. and J. Rattsø (2007): "Local choice of property taxation: evidence from Norway," *Public Choice*, 132, 457–470.

Fontana, R., A. Nuvolari, H. Shimizu, and A. Vezzulli (2013): "Reassessing patent propensity: Evidence from a dataset of R&D awards, 1977–2004," *Research Policy*, 42, 1780–1792.

Fryges, H. (2009): "The export–growth relationship: Estimating a dose-response function," *Applied Economics Letters*, 16, 1855–1859.

Gallini, N. T. (2002): "The economics of patents: Lessons from recent US patent reform," *Journal of Economic Perspectives*, 16, 131–154.

Gamerman, D., A. R. B. Moreira, and H. Rue (2003): "Space-varying regression models: specifications and simulation," *Computational Statistics & Data Analysis*, 42, 513–533.

Garrido, R. A. and H. S. Mahmassani (2000): "Forecasting freight transportation demand with the space-time multinomial probit model," *Transportation Research Part B: Methodological*, 34, 403–418.

Gelfand, A. E., H.-J. Kim, C. F. Sirmans, and S. Banerjee (2003): "Spatial modeling with spatially varying coefficient processes," *Journal of the American Statistical Association*, 98, 387–396.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987): "Generalised residuals," *Journal of Econometrics*, 34, 5–32.

Griffith, D. A. (2000): "Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses," *Linear Algebra and its Applications*, 321, 95–112.

Griliches, Z. (1973): "Research expenditures and growth accounting," in *Science and Technology in Economic Growth*, Springer, 59–95.

——— (1979): "Issues in assessing the contribution of research and development to productivity growth," *Bell Journal of Economics*, 10, 92–116.

Grossman, G. M. and A. B. Krueger (1991): "Environmental Impacts of a North American Free Trade Agreement," Working Paper 3914, National Bureau of Economic Research.

Hahn, J., J. Ham, and H. R. Moon (2011): "Test of random versus fixed effects with small within variation," *Economics Letters*, 112, 293–297.

Härdle, W., M. Müller, S. Sperlich, and A. Werwatz (2004): *Nonparametric and semiparametric models*, Springer Science & Business Media.

Hayek, F. A. (1945): "The use of knowledge in society," *The American Economic Review*, 35, 519–530.

Holloway, G. and M. L. A. Lapar (2007): "How Big is Your Neighbourhood? Spatial Implications of Market Participation Among Filipino Smallholders," *Journal of Agricultural Economics*, 58, 37–60.

Holloway, G., B. Shankar, and S. Rahman (2002): "Bayesian spatial probit estimation: a primer and an application to HYV rice adoption," *Agricultural Economics*, 27, 383–402.

Horowitz, J. L. (2009): *Semiparametric and nonparametric methods in econometrics*, Springer.

Jaffe, A. B. (1989a): "Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers," *Research Policy*, 18, 87–97.

——— (1989b): "Real effects of academic research," *The American Economic Review*, 79, 957–970.

Jaffe, A. B., M. S. Fogarty, and B. A. Banks (1998): "Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation," *The Journal of Industrial Economics*, 46, 183–205.

Jenish, N. and I. R. Prucha (2009): "Central limit theorems and uniform laws of large numbers for arrays of random fields," *Journal of Econometrics*, 150, 86–98.

——— (2012): "On spatial processes and asymptotic inference under near-epoch dependence," *Journal of Econometrics*, 170, 178–190.

Kelejian, H. H. and I. R. Prucha (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, 99–121.

——— (2007): "HAC estimation in a spatial framework," *Journal of Econometrics*, 140, 131–154.

Kelejian, H. H. and D. P. Robinson (1995): "Spatial Correlation: A Suggested Alternative to the Autoregressive Model," in *New Directions in Spatial Econometrics*, ed. by L. Anselin and R. J. G. M. Florax, Berlin, Heidelberg: Springer Berlin Heidelberg, 75–95.

Kieschnick, R. and B. D. McCullough (2003): "Regression analysis of variates observed on (0, 1): percentages, proportions and fractions," *Statistical Modelling*, 3, 193–213.

Klier, T. and D. P. McMillen (2008): "Clustering of Auto Supplier Plants in the United States," *Journal of Business & Economic Statistics*, 26, 460–471.

Lahiri, S. N. (1996): "On Inconsistency of Estimators Based on Spatial Data under Infill Asymptotics," *Sankhyā: The Indian Journal of Statistics, Series A*, 58, 403–417.

Lambert, D. M., J. P. Brown, and R. J. Florax (2010): "A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application," *Regional Science and Urban Economics*, 40, 241–252.

Lee, L.-F. (2004): "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72, 1899–1925.

LeSage, J. P. (2000): "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models," *Geographical Analysis*, 32, 19–35.

LeSage, J. P., M. M. Fischer, and T. Scherngell (2007): "Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects," *Papers in Regional Science*, 86, 393–421.

LeSage, J. P., R. Kelley Pace, N. Lam, R. Campanella, and X. Liu (2011): "New Orleans business recovery in the aftermath of Hurricane Katrina," *Journal of the Royal Statistical Society: Series A*, 174, 1007–1027.

LeSage, J. P. and R. Pace (2009): *Introduction to Spatial Econometrics*, Statistics: A Series of Textbooks and Monographs, CRC Press.

LeSage, J. P. and R. K. Pace (2010): *Spatial Econometric Models*, Berlin, Heidelberg: Springer Berlin Heidelberg, 355–376.

Lin, X. and L.-f. Lee (2010): "GMM estimation of spatial autoregressive models with unknown heteroskedasticity," *Journal of Econometrics*, 157, 34–52.

Los, B. and B. Verspagen (2000): "R&D spillovers and productivity: evidence from US manufacturing microdata," *Empirical Economics*, 25, 127–148.

Manning, W. G., A. Basu, and J. Mullahy (2005): "Generalized modeling approaches to risk adjustment of skewed outcomes data," *Journal of Health Economics*, 24, 465–488.

Manning, W. G. and J. Mullahy (2001): "Estimating log models: to transform or not to transform?" *Journal of Health Economics*, 20, 461–494.

Martinetti, D. and G. Geniaux (2017): "Approximate likelihood estimation of spatial probit models," *Regional Science and Urban Economics*, 64, 30–45.

Matzkin, R. L. (2015): "Estimation of nonparametric models with simultaneity," *Econometrica*, 83, 1–66.

McDonald, J. (2009): "Using least squares and tobit in second stage DEA efficiency analyses," *European Journal of Operational Research*, 197, 792–798.

McMillen, D. P. (1992): "Probit with spatial autocorrelation," *Journal of Regional Science*, 32, 335–348.

——— (2013): "McSpatial: Nonparametric spatial data analysis," *R package version*, 2.

Millimet, D. L., J. A. List, and T. Stengos (2003): "The Environmental Kuznets Curve: Real Progress or Misspecified Models?" *The Review of Economics and Statistics*, 85, 1038–1047.

Miyamoto, K., V. Vichiensan, N. Shimomura, and A. Páez (2004): "Discrete Choice Model with Structuralized Spatial Effects for Location Analysis," *Transportation Research Record: Journal of the Transportation Research Board*, 1898, 183–190.

Monjon, S. and P. Waelbroeck (2003): "Assessing spillovers from universities to firms: evidence from French firm-level data," *International Journal of Industrial Organization*, 21, 1255–1270.

Mullahy, J. (1998): "Much ado about two: reconsidering retransformation and the two-part model in health econometrics," *Journal of Health Economics*, 17, 247–281.

Mundlak, Y. (1978): "On the pooling of time series and cross section data," *Econometrica*, 69–85.

Murdoch, J. C., T. Sandler, and W. P. Vijverberg (2003): "The participation decision versus the level of participation in an environmental treaty: a spatial probit analysis," *Journal of Public Economics*, 87, 337–362.

Ord, K. (1975): "Estimation methods for models of spatial interaction," *Journal of the American Statistical Association*, 70, 120–126.

Pace, R. K. and R. Barry (1997a): "Quick computation of spatial autoregressive estimators," *Geographical Analysis*, 29, 232–247.

——— (1997b): "Sparse spatial autoregressions," *Statistics & Probability Letters*, 33, 291–297.

Pace, R. K. and J. P. LeSage (2004): "Chebyshev approximation of log-determinants of spatial weight matrices," *Computational Statistics & Data Analysis*, 45, 179–196.

——— (2016): "Fast Simulated Maximum Likelihood Estimation of the Spatial Probit Model Capable of Handling Large Samples," in *Spatial Econometrics: Qualitative and Limited Dependent Variables*, ed. by T. B. Fomby, R. C. Hill, I. Jeliazkov, J. C. Escanciano, and E. Hillebrand, 3–34.

Pakes, A. and Z. Griliches (1980): "Patents and R&D at the firm level: A first report," *Economics Letters*, 5, 377–381.

Panayotou, T. (1993): "Empirical tests and policy analysis of environmental degradation at different stages of economic development," World Employment Programme research working paper 292778, Geneva.

Paolino, P. (2001): "Maximum likelihood estimation of models with beta-distributed dependent variables," *Political Analysis*, 9, 325–346.

Papke, L. E. and J. M. Wooldridge (1996): "Econometric methods for fractional response variables with an application to 401 (k) plan participation rates," *Journal of Applied Econometrics*, 11, 619–632.

——— (2008): "Panel data methods for fractional response variables with an application to test pass rates," *Journal of Econometrics*, 145, 121–133.

Piergiovanni, R. and E. Santarelli (2001): "Patents and the geographic localization of R&D spillovers in French manufacturing," *Regional Studies*, 35, 697–702.

Pinkse, J. and M. E. Slade (1998): "Contracting in space: An application of spatial statistics to discrete-choice models," *Journal of Econometrics*, 85, 125–154.

Porter, M. E. (1990): "The Competitive Advantage of Nations," *Harvard Business Review*, 68, 73–93.

Porter, M. E., D. S. Gee, and G. J. Pope (2015): "America's Unconventional Energy Opportunity: A win-win plan for the economy, the environment, and a lower-carbon, cleaner-energy future," Tech. rep., Harvard Business School and The Boston Consulting Group.

Proença, I. and H. C. Faustino (2015): "Modelling bilateral intra-industry trade indexes with panel data: a semiparametric approach," *Computational Statistics*, 30, 865–884.

Qu, X. and L.-f. Lee (2012): "LM tests for spatial correlation in spatial models with limited dependent variables," *Regional Science and Urban Economics*, 42, 430–445.

——— (2013): "Locally most powerful tests for spatial interactions in the simultaneous SAR Tobit model," *Regional Science and Urban Economics*, 43, 307–321.

Ramalho, E. A. and J. J. Ramalho (2017): "Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses," *Econometric Reviews*, 36, 397–420.

Ramalho, E. A., J. J. Ramalho, and J. M. Murteira (2011): "Alternative estimating and testing empirical strategies for fractional regression models," *Journal of Economic Surveys*, 25, 19–68.

——— (2014): "A Generalized Goodness-of-functional Form Test for Binary and Fractional Regression Models," *The Manchester School*, 82, 488–507.

Ramalho, J. J. and J. V. da Silva (2009): "A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms," *Quantitative Financ*, 9, 621–636.

Rice, P., A. J. Venables, and E. Patacchini (2006): "Spatial determinants of productivity: Analysis for the regions of Great Britain," *Regional Science and Urban Economics*, 36, 727–752.

Roach, M. and H. Sauermann (2010): "A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry," *Research Policy*, 39, 422–434.

Roorda, M. J., A. Páez, C. Morency, R. Mercado, and S. Farber (2010): "Trip generation of vulnerable populations in three Canadian cities: a spatial ordered probit approach," *Transportation*, 37, 525–548.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2019): "IPUMS USA: Version 9.0 [dataset]," *Minneapolis, MN: IPUMS*.

Rupasingha, A., S. J. Goetz, D. L. Debertin, and A. Pagoulatos (2004): "The environmental Kuznets curve for US counties: A spatial econometric analysis with extensions," *Papers in Regional Science*, 83, 407–424.

Sanyal, P. (2003): "Understanding patents: The role of R&D funding sources and the patent office," *Economics of Innovation and New Technology*, 12, 507–529.

Schumpeter, J. A. (1942): *Capitalism, Socialism and Democracy*.

Schwab, K. and X. Sala-i-Martin (2010): "The Global Competitiveness Report 2010-2011," Geneva: World Economic Forum.

Scotchmer, S. and J. Green (1990): "Novelty and disclosure in patent law," *The RAND Journal of Economics*, 131–146.

Shafik, N. and S. Bandyopadhyay (1992): "Economic growth and environmental quality:time series and cross-country evidence," World development report WPS 904, World Bank.

Sigrist, F. and W. A. Stahel (2011): "Using the censored gamma distribution for modeling fractional response variables with an application to loss given default," *ASTIN Bulletin: The Journal of the IAA*, 41, 673–710.

Smirnov, O. A. (2005): "Computation of the information matrix for models with spatial interaction on a lattice," *Journal of Computational and Graphical Statistics*, 14, 910–927.

——— (2010): "Modeling spatial discrete choice," *Regional Science and Urban Economics*, 40, 292–298.

Smith, T. E. and J. P. LeSage (2004): "A Bayesian probit model with spatial dependencies," in *Spatial and Spatiotemporal Econometrics*, ed. by T. B. Fomby, R. C. Hill, I. Jeliazkov, J. C. Escanciano, and E. Hillebrand, 127–160.

Stuart, T. and O. Sorenson (2003): "The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms," *Research Policy*, 32, 229–253.

van Meijl, H. (1997): "Measuring intersectoral spillovers: French evidence," *Economic Systems Research*, 9, 25–46.

Wagner, J. (2001): "A note on the firm size–export relationship," *Small Business Economics*, 17, 229–237.

Wang, H., E. M. Iglesias, and J. M. Wooldridge (2013): "Partial maximum likelihood estimation of spatial probit models," *Journal of Econometrics*, 172, 77–89.

Wang, P., l. M. Cockburn, and M. L. Puterman (1998): "Analysis of patent data–A mixed-Poisson-regression-model approach," *Journal of Business & Economic Statistics*, 16, 27–41.

Wang, X. and K. M. Kockelman (2009): "Application of the dynamic spatial ordered probit model: Patterns of land development change in Austin, Texas," *Papers in Regional Science*, 88, 345–365.

Winkelmann, R. (2008): *Econometric Analysis of Count Data*, Springer-Verlag Berlin Heidelberg, 5 ed.

Wollni, M. and C. Andersson (2014): "Spatial patterns of organic agriculture adoption: Evidence from Honduras," *Ecological Economics*, 97, 120–128.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2nd ed.

Xu, X. and L.-f. Lee (2015a): "A spatial autoregressive model with a nonlinear transformation of the dependent variable," *Journal of Econometrics*, 186, 1–18.

——— (2015b): "Maximum likelihood estimation of a spatial autoregressive Tobit model," *Journal of Econometrics*, 188, 264–280.