Three Essays In Spatial Econometrics

Koch, Matthias

Published: 01/08/2012

Document Version
Unknown

THREE ESSAYS IN SPATIAL ECONOMETRICS

Matthias Koch

Dissertation

Erster Betreuer:
o. Univ. Prof. Dr. Manfred Fischer

Zweiter Betreuer:
o. Univ. Prof. Dr. Michael Pfaffermayr

Matthias.Koch@wu.ac.at

Abstract: In the last 20 years spatial econometric models, methods and techniques have been applied to a great variety of empirical problems. The essence of a spatial econometric model is the incorporation of a spatial autoregressive lag, which is scaled by the so called spatial autocorrelation parameter. From a mathematical perspective introducing a spatial autoregressive term into the linear regression model yields a system of equations, which may or may not be solvable for the dependent variable. Furthermore even if the system of equations is solvable, the dependent variable may be diverging if the number of observations approaches infinity. One can show that the solvability and boundedness of the dependent variable in spatial autoregessive models are crucially dependent on the (pre-) specified parameter space of the spatial autocorrelation parameter. Since almost all theoretical work in spatial econometrics assumes both model properties, the validity of spatial econometric methods and techniques is also crucially dependent on the (pre-) specified parameter space.

This thesis investigates parameter spaces for spatial autoregressive models and finds that the concepts used in the literature are inadequate, and provides an alternative one. Given this new parameter space likelihood based methods, like maximum likelihood or Bayesian spatial econometrics face severe numerical difficiulities. Since Pace et al. (2010) show that the instrumental variable estimator developed by Kelejian and Prucha (1997) should not be used for the so-called Spatial Durbin Model (SDM) and data often encountered in applied spatial economics, it seems that the most "popular" estimation approaches in applied spatial economics for one of the most "popular" models faces serious problems. These problems motivate a new estimation method for the SDM which is in line with the alternative parameter space concept and is not based on a distributional assumption about the error term. Finally the thesis discusses potential consequences regarding the interpretation of the SDM if a spatial autocorrelation parameter would be found outside the parameter space concepts used in the literature. One consequence is that increasing the explanatory variable would - independent of the associated regression coefficient's sign - for some observations increase the dependent variable while decreasing it for others, if spillover effects or so-called indirect effects are neglected.

Kurzfassung: In den letzten 20 Jahren wurden räumlich ökonometrische Modelle, Methoden und Techniken auf eine Vielzahl von empirischen Problemen angewandt. Die Essenz räumlich ökonometrischer Modelle ist das Vorhandensein räumlich autoregressiver Terme, welche über den sogenannten räumlich autoregressiven Term skaliert werden. Die Verwendung räumlich autoregressiver Terme in einem linearen Regressionsmodell führt zu einem linearen Gleichungssystem, welches nicht unbedingt auf die abhängige Variable aufgelöst werden kann. Sogar wenn das System lösbar ist, kann es sein, dass die abhängige Variabel divergiert, wenn die Anzahl der Observationen gegen undendlich geht. Man kann zeigen, dass die Auflösbarkeit und Beschränktheit der abhängigen Variabel, in räumlich autoregressiven Modellen, sehr stark vom vorher festgelegten Parameterraum des räumlich autoregressiven Korrelationsparameters abhängt. Da fast alle theoretischen Arbeiten bez üglich räumlich ökonometrischer Schätzungen, beide Modelleigenschaften voraussetzen, ist für die Validierung räumlich autoregressiver Modelle und Techniken der Parameterraum existenziell.

Diese Dissertation untersucht Parameterräume für räumlich autoregressive Modelle und findet, dass die in der Literatur üblichen Konzepte inadäquat sind und bietet hierfür Alternativen an. Dieser neue Parameterraum verursacht jedoch numerische Probleme bei sog. Liklihood basierten Schätzmethoden, wie Maximum Likelihood und räumlich- bayesianischen Schätzern. Da Pace et al. (2010) zeigen, dass die Schä tzung von sogenannten Spatial Durbin Modellen (SDM) mittels des Instrumentenvariablenschätzers von Kelejian und Prucha (1997) mit Problemen verbunden ist, scheint es, dass die "populärsten" Schätzer für eines der "populärsten" räumlich ökonometrischen Modelle nicht mehr verwendet werden können. Diese Probleme motivieren ein neues Schätzver-

fahren für SDM, welches kompatibel ist mit dem alternativen Parameterraumkonzept, beziehungsweise verteilungsunabhängig vom Fehlerterm angewendet werden kann. Ausserdem diskutiert diese Arbeit potentielle Folgen für die Interpretation von SDM falls ein rä umlicher Autokorrelationsparameter ausserhalb des traditionellen Parameterraums gefunden wird. Eine Konsequenz wäre, dass das Erhöhen einer erklärenden Variable - unabhängig von dem dazugehörigem Vorzeichen des Regressionskoeffizienten - für manche Beobachtungen eine Erhöhung der zu erklärenden Variable bei gleichzeitiger Verminderung für die übrigen Beobachtungen zur Folge hätte, gegeben dass indirekte Spillovereffekte (sog. indirect effects) vernachlässigt werden.

Table of Contents

1 **Introduction** 1

2 **Paper I: A research note on asymptotic parameter spaces: A proof for the spatial one forward one behind pattern** 7
- 2.1 Introduction 7
- 2.2 Spatial econometrics and the spatial one forward one behind pattern 8
- 2.3 Density proof for the one forward one behind pattern 9
- 2.4 Concluding remarks 12
- 2.5 Comments from the *Linear Algebra and its Applications'* referees and some notes 13

3 **Paper II: Stable parameter spaces in spatial econometric modelling** 15
- 3.1 Introduction 15
- 3.2 Spatial DGPs and problems of the Kelejian & Prucha and the Lee & Liu parameter space 18
- 3.3 A formal parameter space definition 23
- 3.4 Advantages of the new parameter space concept 25
- 3.5 Concluding remarks 31
- 3.6 Change in model interpretation 32
- 3.7 Comments from the *Econometric Theory* referee and some notes 35

4 **Paper III: Spatial filtering and model interpretation for Spatial Durbin Models** 37
- 4.1 Introduction 37
- 4.2 The spatial Durbin Model and "classical" spatial filtering 40
- 4.3 Estimation procedure and model interpretation 42
- 4.4 Monte Carlo design and used performance measures 46
- 4.5 Monte Carlo results 49
- 4.6 Concluding remarks 53
- 4.7 Adapting the estimation method to the second paper's parameter space 54

5 **Conclusions** 57

References 62

A. **Appendix to Paper I** 66

B. **Appendix to Paper II** 68

C. **Appendix to Paper III** 71

List of Tables

Table 3.1 Stability of a certain $(\mathbf{I}_n - 0.95\mathbf{W}_n)^{-1}$ 21

Table 4.1 Monte Carlo Results: Maximum Likelihood for the total effect of x where γ_0=1 and λ_0=0.00 50

Table 4.2 Monte Carlo Results: New estimator for the total effect of x where γ_0=1 and λ_0=0.00 51

Table 4.3 Monte Carlo Results: New estimator for the total effect of x where γ_0=1 and λ_0=0.99 51

Table 4.4 Monte Carlo Results: Maximum Likelihood for the total effect of x where γ_0=1 and λ_0=0.99 52

Table 4.5 Monte Carlo Results: New estimator for ρ_0 where γ_0=1 and λ_0=0.0 55

Table 4.6 Monte Carlo Results: New estimator for the total effect of x where γ_0=1 and λ_0=0.0 55

Table C.1 Monte Carlo Results: Maximum Likelihood for ρ_0 where γ_0=1 and λ_0=0.99 75

Table C.2 Monte Carlo Results: New estimator for ρ_0 where γ_0=1 and λ_0=0.99 75

Table C.3 Monte Carlo Results: Maximum Likelihood for ρ_0 where γ_0=1 and λ_0=0.00 76

Table C.4 Monte Carlo Results: New estimator for ρ_0 where γ_0=1 and λ_0=0.00 76

Table C.5 Monte Carlo Results: Maximum Likelihood for ρ_0 where γ_0=-$\rho_0\beta_0$ and λ_0=0.00 76

Table C.6 Monte Carlo Results: Maximum Likelihood for the total effect of x where γ_0=-$\rho_0\beta_0$ and λ_0=0.00 77

Table C.7 Monte Carlo Results: New estimator for ρ_0 where γ_0=-$\rho_0\beta_0$ and λ_0=0.00 77

Table C.8 Monte Carlo Results: New estimator for the total effect of x where γ_0=-$\rho_0\beta_0$ and λ_0=0.00 77

List of Figures

Figure 3.1 Visualization of the different parameter spaces suitable to describe a two group interaction model, where $|\rho_{11}| = |\rho_{22}| = 1/3$ 29

Figure 4.1 Eigenvalue-histogram of different spatial weight matrices 47

Chapter 1 Introduction

Spatial econometric models, methods and techniques are essential for dealing with data generating processes (DGP) where the dependent variable or error term is spatially autorcorrelated. The key element to deal with spatial autocorrelation is to incorporate spatial autoregressive lags into the regression model. The influence of the spatial lag is scaled by the spatial autocorrelation parameter. In order to derive estimators for spatially autocorrelated DGPs we first have to solve the DGP for the dependent variable. Solving the DGP is mathematically equivalent to solving a linear system of equations. However this first step is only possible if the spatial lag structure and autocorrelation parameter meet some criteria. To be more precise the inverse of the so called "spatial filter" has to exist. Additionally the spatial autocorrelation parameter must only take values such that the first and second moments of the dependent variable are defined. For a given spatial lag we call this set of possible spatial autocorrelation parameters "parameter space". Hence spatially autocorrelated DGPs with a spatial autocorrelation parameter which lies outside such a parameter space, cannot be estimated. Therefore empirical results which find such DGPs have to be treated as inconsistent with the underlying econometric theory. Furthermore, some parameter space properties could be used in order to optimize numerical procedures necessary for an efficient implementation of certain estimation methods, like maximum likelihood (ML).

However what would happen if the parameter space recognized by the literature is inappropriately defined? First we might interpret results as reasonable, although they are inconsistent with the underlying estimation theory and therefore are actually wrong. Second we might treat reasonable results as inconsistent. Furthermore, we might not even find these seemingly inconsistent results, since the numerical approximations, necessary for the estimation, fail. We investigate this train of thought and argue that this may be the case in spatial econometrics. While the first two papers discuss spatial econometric parameter spaces, the third paper proposes a new estimation method which theoretically could be used to find spatial autocorrelation parameters outside the traditional boundaries without distributional assumption about the error term.

The first paper discusses the parameter space proposed by Kelejian and Prucha (2010). These authors consider the solvability of spatial DGPs as sufficient for the existence of the first and second moment of the dependent variable. Since no general analytical proof can be found, the first paper draws attention to the solvability of spatial DGPs given the spatial structure called one forward one behind pattern. The paper finds that for this particular case, asymptotical solvability implies the boundedness of the dependent variable's first and second moment.

The second paper shows that the permissible parameter spaces generally assumed in the literature are inadequate. This paper proves that the parameter space concept proposed by Kelejian and Prucha (2010) can result in nonstationary DGPs, while the parameter space proposed by Lee and Liu (2010) can be too restrictive in applied cases. Therefore the practice of row standardizing lacks a mathematical foundation, since its mathematical proof is based on the Kelejian and Prucha parameter space. These observations are the motivation to supply a new definition for the spatial econometric parameter space. In particular it shows which assumptions are necessary to give row standardizing the needed mathematical foundation. Furthermore the parameter space concept is applied to models concerning group interaction and panels with fixed cross sectional size and yields substantially larger parameter spaces than the ones the literature had so far considered to be stable. Finally a section added to the paper discusses the implications of nontraditional parameter spaces for the model interpretation.

Therefore the first two thesis papers establish that the traditional parameter spaces used in the literature are insufficient: While the parameter space proposed by Kelejian and Prucha (2010) only works for certain weight matrices like the one discussed in the first paper, the parameter space proposed by Lee and Liu (2010) does not cover row standardization. Furthermore there exists the possibility for much larger parameter spaces. If the last observation is correct there still remains the question: Why do applied researchers, in the field of spatial economics not report spatial autocorrelation parameters outside the traditional boundaries? We argue that two currently applied and very popular among practitioners estimation techniques are inadequate: The Instrumental Variable (IV) estimator proposed by Kelejian and Prucha (1998) and likelihood based estimation methods based on normality, like the ML estimator (see for example Anselin (1988)) or Bayesian methods (see LeSage and Pace (2009)).

The virtue of the IV- estimator is that it is very simple to implement since it is in its application very similar to a two stage least squares estimator. However, LeSage et al. (2010) point out that the IV estimator is heavily biased if the explanatory variables were generated by a spatially autocorrelated process. Furthermore they show that housing data in the U.S. has these kind of properties and that these properties might be observable in many applied cases. Additionally, for the among practitioners popular SDM, the IV estimator is not identified if the SDM collapses in the so called spatial error model (see LeSage et al. (2010) for details) and hence is biased in this particular case as well. Therefore many IV- estimation results might be biased for real world data. As a result a verification of spatial autocorrelation parameters outside of traditional boundaries seems unlikely. Another reason why applied researchers might only report spatial autocorrelation parameters inside the traditional $(-1, 1)$ bounds is that, simply spoken, the current literature treats them as unreasonable and therefore publishing such results would be very "tricky".

The main reason for the popularity of the maximum likelihood estimator is its estimation efficiency. Nevertheless maximum likelihood estimation faces some numerical issues: In order to maximize the log likelihood efficiently one needs efficient numerical algorithms for evaluating the log- determinant of the variance covariance matrix given different spatial autocorrelation parameters. Current state of the art routines are implemented in the popular spatial econometrics toolbox of James P. LeSage. The LeSage toolbox basically provides three options regarding the calculation of the log determinant: the medium fast approximation, the very fast approximation and the full calculation:

1 *The very fast approximation* is based on Pace and Barry (1999) where they use the Monte Carlo simulation method for approximating the log determinant. The analytical proof for this simulation method however is based on traditional parameter spaces. To be more precise the proof for Theorem 1 on page 44 in Pace and Barry (1999) requires the existence of a Neumann Series which only exists for traditional parameter spaces[1].

2 *The medium fast approximation* is based on Pace and Barry (1998) where they use a spline approximation and the grid points are calculated via the LU- factorization. This numerical approximation faces one conceptual problem: The Spline approximation is only feasible for continuous functions. However the log-determinant function has as many poles as real eigenvalues. Hence for symmetric spatial weight matrices we would need as many spline approximations as observations. Furthermore reasonable grid points for nontraditional parameter spaces are a function of the weight matrix eigenvalues[2], which impose numerical problems for large sample sizes. We have to conclude that the medium fast approximation is not feasible if the spatial autocorrelation parameter lies outside the traditional parameter space.

3 *The full calculation* is very similar to the medium fast approximation. The only difference is that it is based on an equicontinuous grid. Therefore this method shares the same problems as the medium fast approximation. Furthermore, since the distances between each of the poles implied the log determinant function are different, a continuous grid is inefficient from

---

1 The approximation method suggested by Martin (1993) faces a similar problem. Martin (1993) uses a Taylorseries approximation for the log derterminant which only exists for traditional parameter spaces.

2 The Chebyshev approximation technique suggested by Pace and LeSage (2004) would face similar problems, since the Chebyshev approximation has to be calculated separately between each of the poles. Therefore we would need to know all real eigenvalues of the spatial weigth matrix.

a numerical perspective.

Furthermore the LeSage toolbox uses for the optimization the routine *fminbnd*. This routine assumes a fixed interval, which is unknown beforehand. Given these numerical issues we have to conclude that the LeSage toolbox, in its current form, can not be used for the verification of DPGs where the spatial autocorrelation parameter maybe outside the traditional parameter space. Note that even the in Ord (1975) suggested log determinant computation via the eigenvalues of the spatial weight matrix can not be applied, since it is based on traditional parameter spaces[3].

Although these previously described numerical issues regarding the maximum likelihood estimation can be fixed at the cost of computational efficiency, there still remain spatially autocorrelated DGPs where maximum likelihood can not be used. Pace et. al. (2011) for example name models where the dependent variables represent binary, discrete choice outcomes or Poisson distributed counts and the observed autoregressive structure is in the dependent variable.

These kind of problems motivate the third paper where a new estimation method, based on spatial filtering is proposed. These spatial filters for spatial autoregressive models like the SDM have seen great interest in the recent literature. Pace et al. (2011) show that the spatial filtering methods developed by Griffith (2000) have desirable estimation properties for some parameters associated with spatial autoregressive models. However, spatial filtering faces two conceptual weaknesses: First if the spatial autocorrelation enters via the dependent variable the estimated parameters lack in general, and especially for the Spatial Durbin Model a proper interpretation. Second, there exists an inherent trade-off between the estimator bias and its efficiency, depending on the spectrum of the used spatial weight matrix. The third thesis paper tackles both problems by introducing a new four step estimation procedure based on the eigenvectors of the spatial weight

[3] Ord (1975) suggests to use for the numerical implementation of the loglikelihood given a spatial autoregressive model: $\log(\mathbf{\Omega}^{-1})^{1/2} = \log(\sigma^{-2}\,(\mathbf{I}_n - \rho\mathbf{W}_n)\,(\mathbf{I}_n - \rho\mathbf{W}'_n))^{1/2} = \frac{-n}{2}\log(\sigma^2) + \log\left(\prod_{i=1}^{n}(1-\rho\lambda_i)\right)$, where $\mathbf{\Omega}$ is the variance covariance matrix of the dependent variable, σ^2 the variance of the error term, $\mathbf{W}_n$ the spatial weigth matrix, ρ the spatial autocorrelation parameter, n the number of observations and λ_i are the eigenvalues of $\mathbf{W}_n$. Note however that the correct formula if $|\rho| > 1$ is $\log(\mathbf{\Omega}^{-1})^{1/2} = \frac{-n}{2}\log(\sigma^2) + \log\left(\prod_{i=1}^{n}|1-\rho\lambda_i|\right)$

matrix. This new estimation procedure estimates all parameters of interest in a Spatial Durbin model and thus allows for a proper model interpretation. Additionally the estimation procedure's efficiency is only marginally influenced by the number of added eigenvectors, which allows us to use $\sim$95% of the available eigenvectors. By using Monte Carlo Simulations we observe that the estimation procedure has a lower (or equal) bias and smaller (or equal) sample variance as the corresponding Maximum Likelihood estimator based on normality. Although only discussed for a spatially autocorrelated DGPs with traditional parameter space, this method could potentially be extended to cover parameter spaces suggested in second paper. Such an extension is discussed in an added subchapter of the third paper.

Chapter 2 Paper I: A research note on asymptotic parameter spaces: A proof for the spatial one forward one behind pattern

Abstract[4]**:** In spatial econometrics the asymptotic parameter space informs applied researchers whether certain values of the spatial lag can be properly interpreted and if a spatial model asymptotically exists. Unlike the parameter space form the time series literature, it is only possible to postulate sufficient conditions for the stability of spatial econometric models. Therefore, we develop an explicit proof for the parameter space of the spatial one forward one behind pattern. We will show that the inverse eigenvalues of this normalized W- matrix are dense outside the interval (-1,1)

Keywords: spatial econometrics, parameter space, inverse eigenvalues, spatial one forward one behind pattern

JEL-Classification Codes: C13, C18, C31

2.1 Introduction

Common econometric sense requires that the data generating process represented by a spatial econometric model is well defined. Therefore, the spatial econometric literature stipulates for the parameter space of ρ to be so that $\frac{1}{\rho} \notin Sp\left(\mathbf{W}_n\right)$ where n denotes the sample size and $Sp(\mathbf{W}_n)$ the set containing the eigenvalues of $\mathbf{W}_n$ (see Kelejian and Prucha (1998) or Assumption 3 from Lee and Yu (2010)). Otherwise $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ is not defined. If $\mathbf{W}_n$ has a matrix norm smaller than or equal one ($\|\mathbf{W}_n\|_1 \vee \|\mathbf{W}_n\|_\infty \leq 1$) it follows due to the Gershgorin theorem that the process is well defined for every $\rho \in (-1, 1)$ (see Kelejian and Prucha (1998), Note 8). On account of these "standard" assumptions it is sufficient to use as a possible parameter space $\triangle = (-1, 1)$. However, it is not clear whether this is too restrictive for spatial models. As a consequence applied

[4] This paper was submitted to *Linear Algebra and its Applications* in Oct. 2010. Furthermore it was presented at the V World Conference of the Spatial Econometrics Association 2011 in Tolouse and research seminars at the Vienna University of Economics and Business and at the University of Innsbruck

researchers do not interpret results like $|\rho| > 1$ because they think that their models are not stable or asymptotically defined. This fact motivates analyzing the parameter space for the often used and simple spatial one forward one behind pattern in more detail. We will show that the inverse eigenvalues of this specific pattern, represented by $\mathbf{W}_n$ are dense[5] in the set $\mathbb{R}\backslash(-1, 1)$.

Let $\mathbf{Y}_n, \mathbf{S}_n \in \mathbb{R}^{n\times 1}$. $\mathbf{I}_n$ denotes the identity matrix of dimension n. One can write $\mathbf{Y}_n = (y_1, y_2, ..., y_n)'$. $|\mathbf{A}|$ denotes the determinant of the matrix $\mathbf{A}$. We define the set Γ_n so that $\forall \rho \in \Gamma_n : |\mathbf{I}_n - \rho \mathbf{W}_n| = 0$. In addition we define $\Gamma := \lim_{n\to\infty} \Gamma_n$. Therefore, the parameter space $\triangle_{FB}$ for the one forward one behind pattern will be $\triangle_{FB} = \mathbb{R}\backslash\Gamma$.

2.2 Spatial econometrics and the spatial one forward one behind pattern

In this section we will briefly characterize one general formulation for a spatial econometric data generating process (DGP) and then describe the spatial one forward one behind pattern[6].

One general formulation for a spatial econometric DGP is represented by the Manski model given in Eq. (2.1).

$$\begin{aligned} &\mathbf{Y}_n = \widetilde{\rho}\mathbf{W}_n\mathbf{Y}_n + \mathbf{X}_n\widetilde{\beta} + \mathbf{W}_n\mathbf{X}_n\widetilde{\theta} + \mathbf{u} \\ &\text{where } \mathbf{u} = \widetilde{\lambda}\mathbf{W}_n\,\mathbf{u} + \boldsymbol{\epsilon},\ \epsilon_i \sim i.i.d(0, \sigma^2) \end{aligned} \tag{2.1}$$

In (2.1) $\mathbf{X}_n$ represents a n by k matrix of explanatory variables, $\widetilde{\rho}$, $\widetilde{\beta}$, $\widetilde{\theta}$ and $\widetilde{\lambda}$ are the parameters to be estimated and the ϵ_is are independently and identically distributed with zero mean and finite variance σ^2. The Manski model incorporates various representations of spatial DGPs like the Spatial Autoregressive Model, the Spatial Error Model and the Spatial Durbin Model [7]. Obviously, the DGP stated in Eq. (2.1) can be solved for $\mathbf{Y}_n = (\mathbf{I}_n - \widetilde{\rho}\mathbf{W}_n)^{-1}\left(\mathbf{X}_n\widetilde{\beta} + \mathbf{W}_n\mathbf{X}_n\widetilde{\theta} + \left(\mathbf{I}_n - \widetilde{\lambda}\mathbf{W}_n\right)^{-1}\boldsymbol{\epsilon}\right)$ if $\frac{1}{\widetilde{\rho}} \notin Sp(\mathbf{W}_n)$ and $\frac{1}{\widetilde{\lambda}} \notin Sp(\mathbf{W}_n)$. Therefore, the inverse eigenvalues of $\mathbf{W}_n$

5 A subset M is called dense in X if every neighbourhood in X contains a point in M, where a neighbourhood $U_\varepsilon(x)$ is definied by
$U_\varepsilon(x) := \{y \in M \text{ where } |x - y| < \varepsilon\}$
and $\varepsilon > 0$

6 We will use henceforth F-B as an abbreviation for the spatial one forward one behind pattern.

7 For more details to the asuumptions and properties of the data generating process stated in Eq. (2.1), see Elhorst (2010)

are of great interest.

The $\mathbf{W}_n$-matrix describes the spatial neighborhood structure. Apparently we need to specify it in order to give an explicit proof for the parameter space. The $\mathbf{W}_n$-matrix of the F-B pattern describes that the first observation is the neighbor of the second, the second observation is neighbor of the first and third observation and so on. At last the n$^{\text{th}}$ observation is neighbor just of the $(n-1)^{\text{th}}$ observation. It is assumed that the neighbors are affecting each other. The $\mathbf{W}_n$-matrix has one entries for the status "neighbor" and zeros else[8]. As a consequence, $\mathbf{W}_n$ is of size n by n and is symmetric. $\mathbf{W}_n$ is bounded by row and column sums and has all its eigenvalues in the set[9] $[-2, 2]$. It is notationally convenient not to standardize this matrix. The typical element $\overline{w}_{i,j}$ of $\mathbf{W}_n$ is defined by (2.2)

$$\overline{w}_{i,j} = \begin{cases} \overline{w}_{j,i} = 1 \text{ if } j = i+1 \text{ and } i \in \{1, 2, ..., n-1\} \\ 0 \text{ otherwise} \end{cases} \tag{2.2}$$

2.3 Density proof for the one forward one behind pattern

In this section we will prove that the inverse eigenvalues of the F-B pattern are dense in $\mathbb{R}\backslash\left(-\frac{1}{2}, \frac{1}{2}\right)$. The proof contains different theorems and lemmas. While theorems themselves will be stated and verified in this section lemmas will be stated and verified in the Appendix.

Before we start with the first theorem, we lay out the structure of the proof: First we will solve the system of equations $G_n : \mathbf{Y}_n = \rho \mathbf{W}_n \mathbf{Y}_n + \mathbf{S}_n$ where $\mathbf{S}_n$ can be any vector out of $\mathbb{R}^{n\times 1}$. Then Theorem 2.1 and 2.2 will show that for $|\rho| > \frac{1}{2}$ it is possible to derive an analytical solution for y_1 by using trigonometric functions. It is clear that $\lim_{\rho \to \rho_0} y_1$ is diverging if $\frac{1}{\rho_0} \in Sp\left(\mathbf{W}_n\right)$. Theorem 2.3 will show that a set of angles Φ that cause the diverging of $\lim_{\rho \to \rho_0} y_1$ is dense in $\mathbb{R}$ if $n \to \infty$ and that there exists a continuous function $\overline{g}(\rho) = g(f(\rho)) = \varphi_\alpha(\rho)$ on $\left(-\infty, -\frac{1}{2}\right) \cup \left(\frac{1}{2}, \infty\right)$ to Φ and thus showing that the inverse eigenvalues of $\mathbf{W}_n$ are dense in $\mathbb{R}\backslash\left(-\frac{1}{2}, \frac{1}{2}\right)$. If $\mathbf{W}_n$ is normalized by

8 It is a commen convention to set the diagonal entries of the $\mathbf{W}_n$ matrix to zero.

9 This follows directly by applying the Gershgorin theorem

the factor $\frac{1}{2}$ it follows that the set $\lim_{n\to\infty}\left\{\frac{1}{\rho}\text{ where }\rho\in Sp\left(\frac{1}{2}\mathbf{W}_n\right)\right\}$ is dense in $\mathbb{R}\setminus(-1,1)$.

Observation 1.1: *The following set is dense in* $\mathbb{R}$: $\left\{\frac{a+\frac{1}{2}}{b+1}|a\in\mathbb{Z},b\in\mathbb{N}\right\}$

Theorem 1 *Finding the parameter space of (2.1) is like solving the following linear system of equations* $G_n:\mathbf{Y}_n=\rho\mathbf{W}_n\mathbf{Y}_n+\mathbf{S}_n$. *One can solve the system for* y_1 *with iterative insertion and get the following recursion for* y_1: $y_1=\rho^{n-1}s_n\prod_{j=1}^{n-1}(z_j)^{-1}+\sum_{i=1}^{n-1}\rho^{i-1}s_i\prod_{j=n-i}^{n-1}(z_j)^{-1}$ *with* $z_{i+1}=1-\frac{\rho^2}{z_i}, z_1=1-\rho^2$. *Through applying lemma A.1.1 we can write instead of the recursion:*

$$z_j=\frac{\left(\frac{-1-w}{\rho^2}\right)^j\left(1-2\rho^2+w\right)-\left(\frac{-1+w}{\rho^2}\right)^j\left(1-2\rho^2-w\right)}{\left(\frac{-1-w}{\rho^2}\right)^j(1+w)-\left(\frac{-1+w}{\rho^2}\right)^j(1-w)}\text{ where } w=\sqrt{1-4\rho^2}$$

Proof. by using iterative insertion it is trivial, but time consuming. ■

Theorem 1.1 and Lemma A.1.1 help us to find an analytical solution for y_1 under the assumption of the linear system of equations G_n. Although it is an analytical solution it is unpractical to work with. Therefore, we are using theorem 2.2. Note that the parameter $w=\sqrt{1-4\rho^2}$ is a complex number if $|\rho|>\frac{1}{2}$. As we will see in theorem 1.2, the assumption of $|\rho|>\frac{1}{2}$ enables us to write the z_j with trigonometric functions. Furthermore, we can derive conditions for the divergence of $\lim_{n\to\infty}y_1$.

Theorem 2 (1.2) *Let* $|\rho|>\frac{1}{2}$. *Therefore,* z_j *can be written as:* $z_j=-\rho\frac{\sin\left(j\varphi_\alpha+\varphi_\beta\right)}{\sin(j\varphi_\alpha+\varphi_\alpha)}$ *with* $\varphi_\beta=\frac{\pi}{2}-\arcsin\left(\frac{1-2\rho^2}{2\rho^2}\right)$, $\varphi_\alpha=\pi+\arcsin\left(\frac{\sqrt{4\rho^2-1}}{|2\rho|}\right)$, $\varphi_\alpha\in\left(\pi,\frac{\pi}{2}\right)$ *and* $\varphi_\beta\in(0,\pi)$. *Lemma A.1.2 shows that* $\varphi_\beta=2\varphi_\alpha-2\pi$. *As a result* y_1 *can now be written as:* $y_1=\prod_{j=1}^{n-1}\frac{-\sin(j\varphi_\alpha+\varphi_\alpha)}{\sin((j+2)\varphi_\alpha)}s_n+\sum_{i=1}^{n-1}s_j\prod_{j=n-i}^{n-1}\frac{-\sin(j\varphi_\alpha+\varphi_\alpha)}{\sin((j+2)\varphi_\alpha)}=\prod_{j=1}^{n-1}\frac{-s_n}{\cos(\varphi_\alpha)}+\frac{-\tan((j+1)\varphi_\alpha)s_n}{\sin(\varphi_\alpha)}+\sum_{i=1}^{n-1}\prod_{j=n-i}^{n-1}\frac{-s_i}{\cos(\varphi_\alpha)}+\frac{-\tan((j+1)\varphi_\alpha)s_i}{\sin(\varphi_\alpha)}$. *In addition we see by using lemma A.1.3 that* $\lim_{n\to\infty}y_1$ *diverges for example if* $(j+1)\varphi_\alpha=\left(m+\frac{1}{2}\right)\pi$ *with* $m\in\mathbb{Z}$, $j\in\mathbb{N}$.

Proof. Define $\alpha := -1 - w = 2\rho e^{i\varphi_\alpha}$ with $\varphi_\alpha = \pi + \arcsin\left(\frac{\sqrt{4\rho^2-1}}{|2\rho|}\right)$, $\beta := 1 - 2\rho^2 + w = 2\rho^2 e^{i\varphi_\beta}$ with $\varphi_\beta = \frac{\pi}{2} + \arcsin\left(\frac{\left|1-2\rho^2\right|}{2\rho^2}\right) = \frac{\pi}{2} - \arcsin\left(\frac{1-2\rho^2}{2\rho^2}\right)$, $\gamma := -1 + w = 2\rho e^{i\varphi_\gamma}$ with $\varphi_\gamma = 2\pi - \varphi_\alpha$ and $\delta := 1 - 2\rho^2 - w = 2\rho^2 e^{i\varphi_\delta}$ with $\varphi_\beta = 2\varphi_\alpha - 2\pi$. Now we can rewrite z_j: $z_j = \frac{\left(\frac{2\rho e^{i\varphi_\alpha}}{\rho^2}\right)^j\left(2\rho^2 e^{i\varphi_\beta}\right)-\left(\frac{2\rho e^{i\varphi_\gamma}}{\rho^2}\right)^j\left(2\rho^2 e^{i\varphi_\delta}\right)}{\left(\frac{2\rho e^{i\varphi_\alpha}}{\rho^2}\right)^j(-1)\left(2\rho e^{i\varphi_\alpha}\right)-\left(\frac{2\rho e^{i\varphi_\gamma}}{\rho^2}\right)^j\left(2\rho e^{i\varphi_\gamma}\right)(-1)} = -\rho\,\frac{e^{ji\varphi_\alpha}e^{i\varphi_\beta}-e^{ji\varphi_\gamma}e^{i\varphi_\delta}}{e^{ji\varphi_\alpha}e^{i\varphi_\alpha}-e^{ji\varphi_\gamma}e^{i\varphi_\gamma}} = -\rho\,\frac{e^{ji\varphi_\alpha}e^{i\varphi_\beta}-e^{-ji\varphi_\alpha}e^{-i\varphi_\beta}}{e^{ji\varphi_\alpha}e^{i\varphi_\alpha}-e^{-ji\varphi_\alpha}e^{-i\varphi_\alpha}}$ $= -\rho\,\frac{\frac{e^{i\left(j\varphi_\alpha+\varphi_\beta\right)}-e^{-i\left(j\varphi_\alpha+\varphi_\beta\right)}}{2i}}{\frac{e^{i(j\varphi_\alpha+\varphi_\alpha)}-e^{-i(j\varphi_\alpha+\varphi_\alpha)}}{2i}} = -\rho\,\frac{\sin\left(j\varphi_\alpha+\varphi_\beta\right)}{\sin(j\varphi_\alpha+\varphi_\alpha)}$ where $\varphi_\alpha \in \left(\pi, \frac{\pi}{2}\right)$ and $\varphi_\beta \in (0,\pi)$ Under the assumption that $\varphi_\beta = 2\varphi_\alpha - 2\pi$ follows: $\frac{\sin(j\varphi_\alpha+\varphi_\alpha)}{\sin\left(j\varphi_\alpha+\varphi_\beta\right)} = \frac{1}{\cos(\varphi_\alpha)} + \frac{\tan((j+1)\varphi_\alpha)}{\sin(\varphi_\alpha)}$ since $\frac{\sin(j\varphi_\alpha+\varphi_\alpha)}{\sin\left(j\varphi_\alpha+\varphi_\beta\right)} = \frac{\sin((j+1)\varphi_\alpha)}{\sin((j+1)\varphi_\alpha+\varphi_\alpha)}$ and $\sin(\alpha+\beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)$ ■

We now have a condition for finding values, or in this case angles, which lead to a divergence of $\lim_{n\to\infty} y_1$. Thus we are able to further analyze this set Φ of angles. Theorem 1.3 (a) concludes that we can construct, due to observation 1.1, dense sets of angles for diverging $\lim_{n\to\infty} y_1$ and that these angles lead to a dense set of ρs in $\left(-\infty, -\frac{1}{2}\right) \cup \left(\frac{1}{2}, \infty\right)$. Since these "diverging" angles represent (at least some) of the inverse eigenvalues of $\mathbf{W}_n$, theorem 1.3 (b) completes the proof.

Theorem 3 (2.3) *(a) Let* $\Phi = \left\{\frac{a+\frac{1}{2}}{b+1}\pi | a \in \mathbb{Z}, b \in \mathbb{N}\right\}$. *Due to observation 1.1 it follows that* Φ *is dense in the real numbers. Furthermore lemma A.1.3 concludes that if we observe* $\lim_{n\to\infty} y_1$ *for every* $\varphi_\alpha \in \Phi$ $\lim_{n\to\infty} y_1$ *is diverging. (b) We observe that* φ_α *can be seen as a continuous, monotone and partly bijective function of* ρ: $\mathbb{R}\setminus\left(-\frac{1}{2}, \frac{1}{2}\right) \to \left(\pi, \frac{3\pi}{2}\right)$ *with* $\varphi_\alpha = \pi + \arcsin\left(\frac{\sqrt{4\rho^2-1}}{|2\rho|}\right)$

Proof. *(a) trivial*

(b) Let $f : \mathbb{R}\setminus\left[-\frac{1}{2}, \frac{1}{2}\right] \to (0,1)$ and $f(\rho) = \frac{\sqrt{4\rho^2-1}}{|2\rho|}$. Obviously, $f(\rho)$ is a continuous and monotonic function. The function $f(\rho)$ is on the domain $\left(-\infty, -\frac{1}{2}\right)$ and $\left(\frac{1}{2}, \infty\right)$ bijective. Furthermore let $g(x)$: $[-1,1] \to \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ with $g(x) = \pi + \arcsin(x)$. $g(x)$ is a continuous, monotonic and bijective function. Therefore, the composition of $g(f(\rho)) = \varphi_\alpha(\rho)$ is also monotonic, continuous and partly bijective. ■

Remark 1 *Note that φ_α is on $\left(-\infty, -\frac{1}{2}\right) \to \left(\pi, \frac{3\pi}{2}\right)$ and $\left(\infty, \frac{1}{2}\right) \to \left(\pi, \frac{3\pi}{2}\right)$ bijective. Let $\widetilde{\Phi} = \Phi \cap \left(\pi, \frac{3\pi}{2}\right)$. It is easy to see that the preimage $\widetilde{\Gamma}$ created by $\varphi_\alpha(\rho)$ and the set $\widetilde{\Phi}$ is part of $\mathbb{R}\backslash\left(-\frac{1}{2}, \frac{1}{2}\right)$. Since $\widetilde{\Phi}$ is dense in the real numbers, it follows that this preimage $\widetilde{\Gamma}$ is dense in $\mathbb{R}\backslash\left(-\frac{1}{2}, \frac{1}{2}\right)$. Theorem 1.3 shows under the assumption of $\lim_{n\to\infty} y_1$ the existence of Γ, where Γ is such that $\forall\rho \in \Gamma : \lim_{n\to\infty} |\mathbf{I}_n - \rho\mathbf{W}_n| = 0$, which is dense in $\mathbb{R}\backslash\left(-\frac{1}{2}, \frac{1}{2}\right)$, because of $\widetilde{\Gamma} \subseteq \Gamma$. Therefore, the asymptotic parameter space for ρ for the one forward one behind pattern is $\triangle = \mathbb{R}\backslash\Gamma = \left(-\frac{1}{2}, \frac{1}{2}\right)$ or when $\mathbf{W}_n$ is normalized with the factor $\frac{1}{2}$ the asymptotic parameter space is $\triangle = (-1, 1)$.*

2.4 Concluding remarks

This proof showed that the inverse eigenvalues of $\mathbf{W}_n$ are dense in $\mathbb{R}\backslash(-1, 1)$ for the one forward one behind pattern if $\mathbf{W}_n$ is normalized with $\frac{1}{2}$ and $n \to \infty$. Hence, we have outside of $(-1, 1)$ no "islands of stability" and thus spatial econometric models with the F-B $\mathbf{W}_n$- matrix are only asymptotically stable if the spatial parameter is smaller than one in absolute value. For the case where the spatial parameter is greater than one in absolute value, it would make no sense to use asymptotics in order find estimator-properties. This case is similar to non-stationarity in the time series literature.

Additionally, the proof helps us to understand the asymptotic behavior of spatial models, because so far, the spatial econometric literature only considered very general assumptions for constructing the parameter space. It showed that for one of the simplest spatial patterns the parameter space considered in the literature is sufficiently large. This is a priori not clear and as a result some authors like Lee, Liu (2010) point out that $(-1, 1)$ is only "a possible parameter space" for the spatial parameters[10]. But the proof strengthens the assumption that it does not make sense to consider parameter spaces greater than $(-1, 1)$ for normalized $\mathbf{W}_n$-matrices in spatial DGPs. As a final consequence it seems more likely that the current best practice of applied researchers, not to interpret empirical results where the spatial parameter lies outside the bound $(-1, 1)$, is correct.

[10] One has to point out that Lee and Yu (2010), Lee and Liu (2010) or Kelejian and Prucha (1998) require addional properties for the parameter space. Addional to $\frac{1}{\rho} \notin Sp(\mathbf{W}_n)$ for all $n \in \mathbb{N} \cup \{\infty\}$ they require that $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ and $(\mathbf{I}_n - \rho\mathbf{W}_n)$ are both bounded in row and column sums in absolute value as $n \to \infty$.

2.5 Comments from the *Linear Algebra and its Applications'* referees and some notes

The first paper was submitted to *Linear Algebra and its Applications* and rejected.

2.5.1 First referee report and comments

This research note shows that the inverse eigenvalues of the spatial one forward one behind pattern are dense in the set $\mathbb{R}\backslash(-1,1)$*, and, hence, the corresponding spatial econometric model are only asymptotically stable if the parameter space is* $(-1,1)$*. The proof is no doubt professionally executed. But I am a little concerned that the usefulness of this result is of limited scope. The spatial one forward one behind pattern is quite ad hoc. Due to its simplicity, such a specific pattern is mostly used in Monte Carlo simulation experiments to study finite-sample performance of estimators for spatial econometric models. It is seldom, if ever, observed in real world applications of spatial econometric models. Thus, the parameter space for the one forward one behind pattern might be of little interest to applied researchers. I think the result of this note would have a much broader impact if the author could generalize the proof to the following cases: (i) a circular one forward one behind pattern, where the first observation is also the neighbor of the last observation; (ii) a spatial* k *forward* k *behind pattern, for some* $k \leq n-1$*, where* n *is the number of spatial units; (iii) an (asymmetric) spatial* k_j *forward* k_j *behind pattern, for* $j = 1, ..., n$*, where the number of neighbors may vary across spatial units; or some other more general spatial patterns.*

Sketches for such proofs were presented at Spatial econometrics Conference 2011 in Tolousse. However a rigorous and complete proof is still work in progress and therefore not part of this thesis. The third generalization should be very tricky as the example in the second thesis paper indicates, since the Kelejian and Prucha (2008) parameter space concept is no longer sufficient.

The note is quite clean except a couple of typos. In the text of Section 3, Observation 2.1 and Theorem 2.1, 2.2, 2.2 should be Observation 3.1 and Theorem 3.1, 3.2, 3.2.

Overall the comments of this referee are very insightful for the first thesis paper.

2.5.2 Second referee report

If there was some special reason for doing so, an applied research could use a spatial parameter of greater than 1 in absolute value. In fact, sometimes they have for negative eigenvalues. However, everyone wants a compact space because optimization doesn't handle other spaces well and maximum likelihood inference has problems with boundaries. The thesis' author is confident that (i) and (ii) can be shown. Sketches for such proofs where presented at Spatial econometrics Conference 2011 in Tolousse. However a rigorous and complete proof is still work in progress and therefore not part of this thesis. The third generalization should be very tricky as the example in the second thesis paper indicates, since the Kelejian and Prucha (2008) parameter space concept is no longer sufficient. Overall the comments of the second referee are very insightful for the first thesis paper.

The general Manski model is not identified when using a single W as discussed in the Elhorst article in the references. In the third sentence in the abstract, it should be space from and not form. In the fifth sentence in the abstract, it should be explicit proof and not prove. On page two, it should be "the stated DGP in (2.1) and not the in (2.1) stated DGP. On page three, has "one entries" should be entries of one. Some footnotes have periods, others do not. Is like is too informal in proof labeled Theorem 3.1. Try not to end sentences with prepositions (with). Theorem 3.1 mislabeled.

Chapter 3 Paper II: Stable parameter spaces in spatial econometric modelling

Abstract[11]**:** Unlike the time series literature, the spatial econometrics literature has not really dealt with the issue of stationary/ stable parameter spaces. This paper shows that current parameter space concepts for spatial econometric DGPs are inadequate. It proves that the parameter space concept proposed by Kelejian and Prucha (2010) can result in nonstationary DGPs, while the parameter space proposed by Lee and Liu (2010) can be too restrictive in applied cases. Furthermore it is discussed that the practice of row standardizing lacks a mathematical foundation.

The paper provides a new definition for the spatial econometric parameter space. In particular it shows which assumptions are necessary to give row standardizing the needed mathematical foundation. Two additional applications for the new parameter space definition concerning models with group interaction and panels with fixed cross sectional size are provided. Both applications result in parameter spaces that are substantially larger than the ones the literature had so far considered to be stable.

*Keywords***:** spatial econometrics, parameter space, row standardizing

JEL-Classification Codes: C13, C18, C31

3.1 Introduction

Spatial econometric models are widely used for data with spatial autocorrelation. The key element to deal with spatial autocorrelation is to incorporate spatial lags into the regression model. This is similar to the case of time series models where time lags are added to deal with serial correlation. For the instance of linear problems in time series it is well known which parameter space configurations yield stationary data generating processes (DGPs) (see for example Hamilton (1994),

[11] This paper was submitted to Econometric Theory in Sep. 2011. Furthermore it was presented at the V World Conference of the Spatial Econometrics Association 2011 in Tolousse, the European Regional Science Association congress 2011 in Barcelona and Dissertationseminars at the Vienna University of Economics and Business and at the University of Innsbruck.

chapter 1). There is no doubt that the (admissible) parameter space for the time lag given a stationary first order autoregressive process is $(-1, 1)$. The spatial econometric literature, so far, has not dealt with the issue of the spatial parameter space in detail. That is even reflected by the nomenclature. In the time series literature it is for example common to distinguish between admissible, stationary and other kind of parameter spaces, while in spatial econometrics virtually no distinction takes place. In this respect the paper can address the following three central issues: *First* it shows that the commonly used parameter space concepts are inadequate. *Second* it proposes a new mathematical parameter space definition and *third* it shows some applications for this new parameter space. In order to keep the nomenclature simple this paper will refer to sets containing the spatial autocorrelation parameters which result into stable DGPs simply as spatial parameter spaces or, if it is clear from the context, parameter space. Since the term "stationary" has a fixed meaning, this paper uses the term "stability" for DGPs that fulfill certain properties, which are given in Section 3. However for the case of a simple time series first order autoregressive process both terms share common properties.

In applied spatial econometrics the following spatial parameter space concept is quite common and this paper will refer to it as the "practitioners" approach: The approach is characterized by normalizing the spatial structure represented by the spatial weight matrix $\mathbf{W}_n$ either by dividing $\mathbf{W}_n$ with its maximum absolute row sum (maximum row standardizing) or dividing each row of $\mathbf{W}_n$ by its absolute row sum (row standardizing) and assumes that the process will be stable for spatial lag parameters in the set $(-1, 1)$. This procedure seems plausible due to the similarity to the time series approach. The "practitioners" approach of row standardizing can be seen as an approximation of the parameter space concept proposed by Kelejian and Prucha. Kelejian and Prucha argue that if $\mathbf{W}_n$ has only real eigenvalues the parameter space must be in a subset of $(1/\tau_{\min}, 1/\tau_{\max})$, where $\tau_{\max}$ and $\tau_{\min}$ represent the maximum and minimum eigenvalue[12] of the

12 Since the simple normalzing procedures like maximum row standardizing can be seen as approximation for the

spatial weight matrix $\mathbf{W}_n$. On the other hand, Lee and Liu (2010) motivate their spatial parameter space concept differently. It is based on the Neumann Series, which states that if a matrix norm of $\rho\mathbf{W}_n$ is smaller than one then the inverse of $(\mathbf{I}_n - \rho\mathbf{W}_n)$ exists, where $\mathbf{I}_n$ is a identity matrix of size n. Additionally if the Manhattan and Infinity norms of $\rho\mathbf{W}_n$ are assumed to be smaller than 1, due to the Neumann Series[13] and the subadditivity of matrix norms, the inverse of $(\mathbf{I}_n - \rho\mathbf{W}_n)$ will be bounded in row and column sums in absolute value[14]. These concepts and their background will be introduced in the next sections in more detail.

This paper shows that three problems exist concerning the previously mentioned parameter spaces. *First* the parameter space concept of Kelejian and Prucha (2010), which is only considering the eigenvalues of the spatial matrix can result in nonstationary DGPs. *Second* applying the Lee and Liu (2010) parameter space and the "practitioners" approach of row standardizing on the same $\mathbf{W}_n$ can result in different parameter spaces[15]. While Lee and Liu (2010), given a spatial one-forward/one-behind structure of $\mathbf{W}_n$ would only consider the spatial parameters to be in $(-2/3, 2/3)$, the "practitioners" approach of row standardizing would consider the spatial parameters to be in $(-1, 1)$. Since neither the Lee and Liu (2010) nor the Kelejian and Prucha (2010) parameter space concept cover or are useful to deduce the "practitioners" approach of row standardizing, it lacks a mathematical foundation. *Third*, this paper shows that in many cases the Lee and Liu (2010) parameter space is too restrictive.

These three problems of current parameter space concepts are the main motivation to define the spatial parameter space indirectly via its desired mathematical properties. If a parameter space

---

biggest absolute eigenvalue, some authors like LeSage and Pace (2009) correctly pointed out that this may result in a too restrictive parameter space, since $(-1, 1)$ may only a subset of $1/\max_i \left\{\sum_{j=1}^{n} |w_{i,j}|\right\} (-|\tau_{\min}|, |\tau_{\max}|)$ where $w_{i,j}$ is the typical element of $\mathbf{W}_n$

13 The Literature (see for example LeSage and Pace (2009)) is often incorrectly referring to the series expansion of $(\mathbf{I}_n - \rho\mathbf{W}_n)$ as Taylor series. Since the series is much more general (see Appendix) and was published by Carl Neumann in 1877 and is referred to as Neumann series in functional analysis, this paper uses the term "Neumann series" if $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ can be written as $\sum_{k=0}^{\infty} \rho^k \mathbf{W}_n^k$

14 See Lemma 1 in the appendix

15 The second practitioners approach, namely maximum row standardizing is covered by the parameter space concept proposed by Lee and Liu (2010)

fulfills these properties we refer to it as stable. The properties are chosen so that the new spatial parameter space can be applied to draw inferences for example with mean distance estimators, like Generalized Method of Moments or Maximum Likelihood. Additionally, these properties ensure for a spatial autoregressive model that the model's total and indirect effects[16] will be finite as the sample size tends to infinity.

Due to this new parameter space definition the paper provides examples where accounting for the inner structure of the spatial weight matrix results in spatial parameter spaces that the literature would have treated as unstable or exploding. Additionally, it is possible to show that the $(-1, 1)$ parameter space for a spatial lag after applying the "practitioners" approach of row standardizing, will result in almost all applied cases in stable DPGs.

The outline of the paper is the following: The next section briefly describes some spatial econometric DGPs, introduces the spatial parameter spaces of Kelejian and Prucha (2010) and Lee and Liu (2010) and then shows some fundamental problems of them. The problems of Section 2 motivate a different definition for spatial parameter spaces which is presented in Section 3. Section 4 uses the definitions to show the advantages of the new parameter space concept. The last section briefly concludes and summarizes this work. The appendix provides some useful theorems, lemmas and proofs like the Neumann series and Gerschgorin theorem.

3.2 Spatial DGPs and problems of the Kelejian & Prucha and the Lee & Liu parameter space

This section provides two general spatial econometric data generating processes, introduces the spatial parameter spaces of Kelejian and Prucha (2010) and Lee and Liu (2010) and describes their problems.

Notation: Let $\mathbf{Y}_n \in \mathbb{R}^{n\times 1}$ and $\mathbf{S}_n \in \mathbb{R}^{n\times n}$. One can write $\mathbf{Y}_n$ also as $\mathbf{Y} = (y_1, y_2, ..., y_n)'$. $|\chi|$

16 LeSage and Pace (2009) suggest to use direct, indirect and total effects for model interpretation.

denotes the determinant of the matrix χ or the absolute value of the scalar χ, while $\|\|_\infty$ denotes the maximum absolute row sum and $\|\|_1$ the maximum absolute column sum of a matrix[17]. Due to notational convenience, not all indices will always be written. It should be clear from the context. $Sp(\mathbf{W}_n)$ denotes the set containing the eigenvalues of the matrix $\mathbf{W}_n$.

3.2.1 Problems of Kelejian & Prucha parameter space

One general formulation for a spatial econometric DGP is represented by the Manski model given in Eq. (3.1).

$$\begin{array}{c}\mathbf{Y}_n = \rho\mathbf{W}_n\mathbf{Y}_n + \mathbf{X}_n\beta + \mathbf{W}_n\mathbf{X}_n\theta + \mathbf{u}_n \\ \text{where } \mathbf{u}_n{=}\lambda\mathbf{W}_n\mathbf{u}_n{+}\boldsymbol{\epsilon}_n,\ \epsilon_i \sim i.i.d(0,\sigma^2)\end{array} \tag{3.1}$$

In Eq. (3.1) $\mathbf{X}_n$ represents the n by k matrix of explanatory variables, ρ, β, θ and λ are the parameters to be estimated where ρ and λ are scalars and β and θ are k by 1 vectors. The ϵ_i are independently and identically distributed with zero mean and finite variance σ^2. $\mathbf{W}_n$ represents the n by n spatial weights matrix of known constants. The diagonal entries of $\mathbf{W}_n$ are assumed to be zero[18]. The Manski model incorporates various representations of spatial DGPs like the Spatial Autoregressive Model, the Spatial Error Model and the Spatial Durbin Model[19]. The DGP stated in Eq. (3.1) can be solved for $\mathbf{Y}_n = (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\left(\mathbf{X}_n\beta + \mathbf{W}_n\mathbf{X}_n\theta + (\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}\boldsymbol{\epsilon}\right)$ if $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ and $(\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}$ exist.

The parameter space proposed by Kelejian and Prucha (2010) for the Manski model stated in Eq. (3.1) can be sketched in the following manner: They argue correctly that if $\mathbf{W}_n$ is not normalized $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ might not be defined for some values of $\rho \in (-1,1)$. Therefore, they suggest that the parameter space of the spatial parameter should be $(-1/\tau_n, 1/\tau_n)$, where $\tau_n = \max_{1\leq i\leq n}\{|\nu_i| \text{ where } \nu_i \in Sp(\mathbf{W}_n)\}$. Since evaluating the eigenvalues of $\mathbf{W}_n$ can be numerically difficult, they suggest to use the Gershgorin theorem to get an upper bound for τ_n, what they

17 Note that these matrix norms satisfy the following useful inequality: $\|\mathbf{A}_n\mathbf{B}_n\|_\vartheta \leq \|\mathbf{A}_n\|_\vartheta \|\mathbf{B}_n\|_\vartheta$ where $\vartheta \in \{1,\infty\}$ and $\mathbf{A}_n$ and $\mathbf{B}_n$ are n by n matrices. For more detais see Johnsen and Horn (1985)

18 Although it is possible to derive parameter spaces for $\mathbf{W}_n$ matrices where the diagonal elements are not zero, such matrices are not common in applications.

19 For more details to the assumptions and properties of the data generating process stated in (3.1), see Elhorst (2010)

call $\tau_n^* = \min\left\{\max\limits_{1\le i\le n}\left\{\sum\limits_{j=1}^{n}|w_{ij,n}|\right\}, \max\limits_{1\le i\le n}\left\{\sum\limits_{j=1}^{n}|w_{ji,n}|\right\}\right\}$ so that $|\tau_n| \le |\tau_n^*|$. As a result they recommend to use $(-1/\tau_n^*, 1/\tau_n^*)$ as a parameter space for the spatial parameter[20].

Although there are $\mathbf{W}_n$-matrices like the one-forward/one-behind pattern (see Eq. (3.4))[21], where the set defined by $\vee n \in \mathbb{N} \cup \{\infty\} : 1/\rho \notin Sp(\mathbf{W}_n)$ is a well defined parameter space, this result is generally not the case. Consider for example $\mathbf{W}_n = \overline{\mathbf{W}}_n$ with the typical element $\overline{w}_{ij}$ defined by Eq. (3.2)

$$\overline{w}_{i,j} = \begin{cases} 1 \text{ if } i = j+1 \text{ and } j \in \{1,2,...,n-1\} \\ 0 \text{ otherwise} \end{cases} \tag{3.2}$$

One could interpret the spatial lag $\overline{\mathbf{W}}_n\mathbf{Y}_n$ in a model like $\mathbf{Y}_n = \rho\overline{\mathbf{W}}_n\mathbf{Y}_n + \varepsilon_n$ in analog to the time lag of a first order autoregressive process form the time series literature. Obviously the (admissible) parameter space for this "spatial" lag is $(-1, 1)$ given the DPG has to be stationary. As the Appendix A.2 shows $Sp\left(\overline{\mathbf{W}}_n\right) = \{0\}$. Therefore, the parameter space in Kelejian and Prucha (2010) would be $(-1/\tau_n, 1/\tau_n) = (-\infty, \infty) = \mathbb{R}$ and hence can result in nonstationary DGPs[22]. This is not surprising, since there is a difference between the solvability and the boundedness of a spatial DGP. For the $\overline{\mathbf{W}}_n$ stated in Eq. (3.2), one can write down an analytical solution for the inverse: $\vee n \in \mathbb{N}, \vee\rho \in \mathbb{R} : \left(\mathbf{I}_n - \rho\overline{\mathbf{W}}_n\right)^{-1} = \sum_{k=0}^{n} \rho^k \overline{\mathbf{W}}_n^k$, since $\left(\overline{\mathbf{W}}_n\right)^{n+1} = \mathbf{0}_{n\times n}$ where $\mathbf{0}_{n\times n}$ denotes a matrix of size n by n only containing zeros. Note that the series $\sum_{k=0}^{n} \rho^k \overline{\mathbf{W}}_n^k$ converges for every $n \in \mathbb{N}$. This example shows that if the DGP stated in Eq. (3.1) is not solvable for a $n \in \mathbb{N}$, it is also not stable, but if the DGP is solvable it does not mean that it is stable as $n \to \infty$. Therefore, one has to use additional assumptions for the parameter space like the boundedness[23] of $\left(\mathbf{I}_n - \rho\overline{\mathbf{W}}_n\right)^{-1}$ in absolute row and column sums as $n \to \infty$. Since $\sum_{k=0}^{\infty} \rho^k \overline{\mathbf{W}}_n^k$ converges only

20 Please note that this parameter space is not the same Lee and Liu (2010) proposed.

21 see Chapter 2

22 Note when talking about parameter spaces Kelejian and Prucha (2010) only discuss sufficient conditions for their Assumption 1. Kelejian and Prucha (2010) never claim that their suggested parameter space will satisfy the Assumption 3 in Kelejian and Prucha (2010). Therefore the weight matrix given by Eq. (3.2) is an example where fulfilling Assumption 1 does not necessarily mean that Assumption 3 in Kelejian and Prucha (2010) will hold.

23 This is a standard assumption for drawing inferences on spatial econometric models, see for example Kelejian and Prucha (2010). Additionally this assumption ensures that the total effects for a spatial autoregressive model are asymptotically bounded, see LeSage and Pace (2009).

if $|\rho| < 1$, the boundedness condition constrains the spatial parameter space in a useful manner.

An anonymous referee pointed out that $\mathbf{W}_n$ matrices like given in Eq. (3.2) do not represent spatial patterns due to their triangular structure. Therefore it was argued that matrices with $Sp\left(\overline{\mathbf{W}}_n\right) = \{0\}$ are no essential concern of spatial econometrics. We contradict these views by providing the following numerical example: Let $\mathbf{W}_n = \frac{1}{\max\left\{|Sp(\overline{\mathbf{W}}_n + 1/8\overline{\mathbf{W}}'_n)|\right\}}\left(\overline{\mathbf{W}}_n + 1/8\overline{\mathbf{W}}'_n\right)$ where the typical element of $\overline{w}_{ij}$ is defined by Eq. (3.2). Kelejian and Prucha (2010) would argue that $(-1, 1)$ can be used as a parameter space for ρ. Hence we set ρ to 0.95 and report in Table (3.1) ξ_n for different n where ξ_n is defined as $\xi_n = \max_i \{\max_j \{|s_{i,j}|\}\}$ and $s_{i,j}$ is the typical element of $\mathbf{S}_n = (\mathbf{I}_n - 0.95\mathbf{W}_n)^{-1}$. For calculating ξ_n in Table (3.1) we used MATLAB[24].

Table 3.1: Stability of a certain $(\mathbf{I}_n - 0.95\mathbf{W}_n)^{-1}$

n	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*
ξ_n	1.5E+03	1.1E+06	1.2E+09	1.2E+12	2.6E+13	6.5E+13	4.0E+13	3.6E+18
n	*90*	*100*	*110*	*120*	*130*	*140*	*150*	*160*
ξ_n	2.2E+12	1.2E+12	6.4E+11	3.4E+11	1.8E+11	9.7E+10	4.7E+24	1.5E+25

Table (3.1) shows that the matrix $\mathbf{S}_n$ seems to produce numerical difficulties. We find that for $n = 150$ the absolute largest absolute entry in $\mathbf{S}_n$ is $4.7 \cdot 10^{24}$. Additionally MATLAB reports numerical difficulties for $n \in \{150, 160, ..., 270\}$. This numerical example raises the question if the inverse $\mathbf{S}_n$ is really bounded by row and column sums for $\rho \in (-1, 1)$ just by restricting the maximum absolute eigenvalue of $\mathbf{W}_n$ to one. The boundedness of $\mathbf{S}_n$ however is an essential assumption for the proofs given in Kelejian and Prucha (2010) or Lee and Liu (2010).

3.2.2 Problems of the Lee & Liu parameter space

Another general spatial econometric DGP[25] is represented by the higher order spatial autoregres-

24 For calculating the inverse we used the MATLAB- function *inv()*
25 see Lee and Liu (2010) for model details

sive model with autoregressive disturbances.

$$\begin{aligned} \mathbf{Y}_n &= \sum_{j=1}^{p} \rho_j \mathbf{W}_{jn} \mathbf{Y}_n + \mathbf{X}_n \beta + \mathbf{u}_n \\ \text{where } \mathbf{u}_n &= \sum_{k=1}^{q} \lambda_k \mathbf{M}_{kn} \mathbf{u}_n + \boldsymbol{\epsilon}_n,\ \epsilon_i \sim i.i.d(0, \sigma^2) \end{aligned} \tag{3.3}$$

In Eq. (3.3) $\rho_1, ..., \rho_p$ and $\lambda_1, ..., \lambda_q$ represent the different spatial lag parameters, $\mathbf{W}_{1n}$, ..., $\mathbf{W}_{pn}$ and $\mathbf{M}_{1n}$, ..., $\mathbf{M}_{qn}$ are n by n dimensional spatial weights matrices. Like in Eq. (3.1) it is assumed that the diagonal elements of $\mathbf{W}_{jn}$ and $\mathbf{M}_{kn}$ are set to zero. The DGP stated in Eq. (3.3) can be solved for $\mathbf{Y}_n = (\mathbf{I}_n - \sum_{j=1}^{p} \rho_j \mathbf{W}_{jn})^{-1} (\mathbf{X}_n\beta + (\mathbf{I}_n - \sum_{k=1}^{q} \lambda_k \mathbf{M}_{kn})^{-1}\boldsymbol{\epsilon})$ if $\mathbf{I}_n - \sum_{j=1}^{p} \rho_j \mathbf{W}_{jn}$ and $\mathbf{I}_n - \sum_{k=1}^{q} \lambda_k \mathbf{M}_{kn}$ are invertible.

While Kelejian and Prucha (2010) motivate their parameter space via the eigenvalues of $\mathbf{W}_n$, Lee and Liu (2010) motivate the spatial parameter space for the DGP stated in Eq. (3.3) via the following reasoning: They argue that if $\sum_{j=1}^{p} \left|\rho_j\right| < 1/\max_{j=1...p} \{\|\mathbf{W}_{jn}\|_1, \|\mathbf{W}_{jn}\|_\infty\}$ and $\sum_{k=1}^{q} |\lambda_k| < 1/\max_{k=1...q} \{\|\mathbf{M}_{kn}\|_1, \|\mathbf{M}_{kn}\|_\infty\}$ then the inverses of $\mathbf{I}_n - \sum_{j=1}^{p} \rho_j \mathbf{W}_{jn}$ and $\mathbf{I}_n - \sum_{k=1}^{q} \lambda_k \mathbf{M}_{kn}$ exist due to the existence of the Neumann series and both are bounded in row and column sums in absolute value. This condition represents their spatial parameter space. Although it can be applied on neighboring patterns like that given by Eq. (3.2) the parameter space can be restrictive if $p = 1, q = 0$ and $\mathbf{W}_n$ is row normalized.

As an example to show the constraints of the Lee and Liu (2010) spatial parameter space, the spatial one-forward/ one-behind pattern represented by $\overline{\mathbf{W}}_n$ is used where the typical element $\overline{w}_{i,j}$ is defined by Eq. (3.4)

$$\overline{w}_{i,j} = \begin{cases} \overline{w}_{j,i} = 1 \text{ if } j = i+1 \text{ and } i \in \{1, 2, ..., n-1\} \\ 0 \text{ otherwise} \end{cases} \tag{3.4}$$

Row-normalizing $\mathbf{W}_n$ yields $\widetilde{\mathbf{W}}_n$. Obviously $\left\|\widetilde{\mathbf{W}}_n\right\|_\infty = 1$ holds, but $\left\|\widetilde{\mathbf{W}}_n\right\|_1 = 1.5$. Therefore, it is not clear, whether $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_1 = \left\|\sum_{k=0}^{\infty} \rho_n^k \widetilde{\mathbf{W}}_n^k\right\|_1$ converges, because $\left\|\widetilde{\mathbf{W}}_n\right\|_1 > 1$. Hence, the parameter space stated in Lee and Liu (2010) for $\widetilde{\mathbf{W}}_n$ would be $\rho \in \left(-\frac{1}{1.5}, \frac{1}{1.5}\right) =$

$\left(-\frac{2}{3}, \frac{2}{3}\right)$ in order to ensure convergence of $\left\|\sum_{k=0}^{\infty} \rho_n^k \widetilde{\mathbf{W}}_n^k\right\|_1$.

The previous reasoning does not imply that the spatial parameter space of Eq. (3.4) has to be $\left(-\frac{2}{3}, \frac{2}{3}\right)$. With the help of the new parameter space definition presented in the next section and and Theorem 2.1 in section 3.4.1, it can be shown that for row standardized $\mathbf{W}_n$-matrices like $\widetilde{\mathbf{W}}_n$ given in Eq. (3.4) the parameter space is still $(-1, 1)$. Therefore, the Lee and Liu (2010) parameter space results in too restrictive parameter spaces for all $\mathbf{W}_n$ where a row standardization was applied and the row sums of $\mathbf{W}_n$ before the normalization were different and the original weight matrix was symmetric.

3.3 A formal parameter space definition

In order to derive estimator properties for DGPs like Eq. (3.1) or Eq. (3.3), stability is required. The stability assumption is reflected via boundedness conditions of the DGP. Hence, this chapter proposes to use these conditions to construct the spatial parameter space. Consequently the proposed stable spatial parameter space has to satisfy the following properties (1), (2), and (3):

Definition 1 *Let $\rho_i \in \Theta_i \subset \mathbb{R}$, $\mathbf{\Psi}_{n,p} = \sum_{j=1}^{p} \rho_j \mathbf{W}_{jn}$ if $p > 1$, $\mathbf{\Psi}_{n,1} = \rho \mathbf{W}_n$ and the diagonal entries of $\mathbf{W}_n$ be zero. Θ_i where $i \in \{1, ..., p\}$ is labeled a stable spatial parameter space, if the following properties are met:*

(1) $\forall \rho_i \in \Theta_i, n \in \mathbb{N} \cup \{\infty\} : |\mathbf{I}_n - \mathbf{\Psi}_{n,p}| \neq 0$

(2) $\forall \rho_i \in \Theta_i, n \in \mathbb{N} \cup \{\infty\} : \left\|(\mathbf{I}_n - \mathbf{\Psi}_{n,p})^{-1}\right\|_1 < \infty$, $\left\|(\mathbf{I}_n - \mathbf{\Psi}_{n,p})^{-1}\right\|_\infty < \infty$, $\|\mathbf{\Psi}_{n,p}\|_\infty < \infty$ and $\|\mathbf{\Psi}_{n,p}\|_1 < \infty$

(3) $\Theta_i \subset \mathbb{R} \backslash \bigcup_{j \in \mathbb{N}} \{\alpha_j\}$ where $\alpha_j \in \mathbb{R}$ and $\bigcup_{j \in \mathbb{N}} \{\alpha_j\}$ is not dense while Θ_i is an interval

The first property simply states that for every sample size, even if it approaches infinity, the inverse of $\mathbf{I}_n - \rho \mathbf{W}_n$ or $\mathbf{I}_n - \sum_{j=1}^{p} \rho_j \mathbf{W}_{jn}$ must always exist. If one only uses the the first condition in order to find parameter spaces, one would use the Kelejian and Prucha (2010) parameter space. The first condition simply ensures that the spatial DGP exists for every $n \in \mathbb{N} \cup \{\infty\}$.

The second property ensures the boundedness of the inverse in absolute row and column sums. In a statistical sense this property ensures finite moments of $\mathbf{Y}_n$. This property guarantees for $\mathbf{W}_n$ as defined in Eq. (3.2) that the parameter space is only $(-1,1)$. Additionally, Property (2) ensures that for example a spatial autoregressive model can be properly interpreted in a way that was suggested by LeSage and Pace (2009). LeSage and Pace recommend to use direct, indirect and total effects for the model interpretation, where as the j-th total effect for the DGP given in Eq. (3.3) is given by: $\frac{1}{n}\boldsymbol{\iota}_n'\left(\mathbf{I}_n - \sum_{j=1}^{p}\rho_j\mathbf{W}_{jn}\right)^{-1}\boldsymbol{\iota}_n\beta_j$. Hence the boundedness of the inverse ensures a proper model interpretation. The properties $\|\boldsymbol{\Psi}_{n,p}\|_\infty < \infty$ and $\|\boldsymbol{\Psi}_{n,p}\|_1 < \infty$ imply that the spatial spillover has to be bounded as well. Hence the nomenclature stability.

The third property reflects the idea that only intervals are used as a parameter space and not a countable set of points. This is necessary for example if mean value theorems for deriving estimator properties are used. We point out that for the Lebesgue measure of $\bigcup_{j\in\mathbb{N}}\{\alpha_j\}$ is 0 and therefore the set $\bigcup_{j\in\mathbb{N}}\{\alpha_j\}$ does not influence the consistency proofs.

Properties (1) and (2) clearly show the difference between solvability and stability of a spatial DGP. Although there exist examples, like the spatial one forward one behind pattern, where property 1 is sufficient for Property (2), this is not generally the case like in example given by Eq. (3.2).

Equipped with the properties (1)-(3) the parameter space proposed by Lee and Liu (2010) becomes clearer. If there is no additional knowledge about $\mathbf{W}_{jn}$ and an explicit spatial parameter space for ρ_i is desired one could use the following reasoning: If any matrix norm of $\boldsymbol{\Psi}_{n,p}$ is smaller than one the Neumann series can be applied to find the inverse of $\mathbf{I}_n - \boldsymbol{\Psi}_{n,p}$. Additionally Lemma 2.1 shows for $\varrho \in \{1,\infty\}$ if $\|\boldsymbol{\Psi}_{n,p}\|_\varrho < 1$ that $\left\|(\mathbf{I}_n - \boldsymbol{\Psi}_{n,p})^{-1}\right\|_\varrho \leq \frac{1}{1-\|\boldsymbol{\Psi}_{n,p}\|_\varrho} < \infty$. This is equivalent to the spatial parameter space suggested by Lee and Liu (2010).

The next section shows how the Properties (1)-(3) can help to construct spatial parameter spaces that are larger than the ones currently considered by the literature. Additionally the next section shows which conditions are necessary in order for the "practitioners" approach of row normalizing to fulfill the Properties (1)-(3).

3.4 Advantages of the new parameter space concept

This section provides three applications of the new parameter space concept in order to show its advantages. Since the concept is defined indirectly via mathematical properties it is possible to use the inner structure of the spatial lag(s) to derive the corresponding spatial parameter spaces. The examples of this section suggest that the more inner structure of the spatial lag(s) is present the more accurate and in these examples larger the spatial parameter space gets. *First* subsection 3.4.1 delivers the mathematical foundation for the "practitioners" approach of row standardizing. The key assumption for this proof is that the $\mathbf{W}_n$ was symmetric before the row standardization took place. A *second* application is a group interaction model where the within and between group interaction is modelled with partitioned $\mathbf{W}$- matrices. This particular inner structure of the spatial lags results in a significantly larger spatial parameter space than the one proposed by Lee and Liu (2010). Subsection 3.4.3 provides the *third* application, where it turns out that in some panel settings it is possible to have the whole real line except a finite number of points as the spatial parameter space. Additionally this section suggests one possible interpretation for the inevitable cross section sample size dependence of the parameter space. In the context of repeated sampling it can be seen for example as a consequence of the geographic scale.

3.4.1 Row standardizing and stability

This subsection shows that the "practitioners" approach of row standardizing yields stable parameter spaces. Let $\mathbf{W}_n$ be the spatial weight matrix with the typical element $w_{i,j}$ and $w_{i,i} = 0$. The row standardization for $\mathbf{W}_n$ is represented by the product with the diagonal matrix $\mathbf{\Lambda}_n$ with the typical element[26] $\lambda_{i,i} = 1/\sum_{j=1}^{n}|w_{i,j}|$. Hence the normalized weight matrix is given by $\widetilde{\mathbf{W}}_n = \mathbf{\Lambda}_n\mathbf{W}_n$. The aim is to find whether $\rho \in (-1,1)$ fulfills the parameter space properties (1)-(3). Property (3) is obviously fulfilled. In order to find the inverse of $\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n$ a Neumann series can be applied $(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n)^{-1} = \sum\limits_{k=0}^{\infty}\rho^k\widetilde{\mathbf{W}}_n^k$ and hence, property (1) is fulfilled. To show property (2) via Theorem 1 additional assumptions are necessary: $\mathbf{W}_n$ has to be symmetric and $\|\mathbf{W}_n\|_\infty < \kappa \in \mathbb{R}$ is bounded.

The first assumption of a symmetric $\mathbf{W}_n$ is generally fulfilled if for example, $\mathbf{W}_n$ represents a spatial neighboring structure. If observation A is neighbor of observation B, the reverse must also be true. The second assumption requests that $\mathbf{W}_n$ has bounded absolute row sums. In the context of a neighboring structure that is equivalent to limiting the number of neighbors for each observation to a finite constant.

Theorem 2.1 *Let W_n be a symmetric weight matrix with finite weights $w_{i,j} \in \mathbb{R}$. The dependence structure is limited such that $\|\mathbf{W}_n\|_\infty < \kappa \in \mathbb{R}$. Let $\widetilde{\mathbf{W}}_n$ be the row standardized version of $\mathbf{W}_n$. If $|\rho| < 1$ it follows that:* $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_\infty < \infty$, $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_1 < \infty$, $\left\|\widetilde{\mathbf{W}}_n\right\|_\infty < \infty$, $\left\|\widetilde{\mathbf{W}}_n\right\|_1 < \infty$

Proof. Four properties have to be shown:

(I) $\left\|\widetilde{\mathbf{W}}_n\right\|_\infty < \infty$: Due to the construction of $\mathbf{\Lambda}_n$: $\left\|\widetilde{\mathbf{W}}_n\right\|_\infty = \|\mathbf{\Lambda}_n\mathbf{W}_n\|_\infty = 1 < \infty$.

(II) $\left\|\widetilde{\mathbf{W}}_n\right\|_1 < \infty$: Observe that $\left\|\widetilde{\mathbf{W}}_n\right\|_1 = \|\mathbf{\Lambda}_n\mathbf{W}_n\|_1 = \|\mathbf{W}_n'\mathbf{\Lambda}_n\|_\infty$ holds. Since $\mathbf{W}_n$ is symmetric and the property of the sub- multiplicatively of matrix norms $\|\mathbf{W}_n'\mathbf{\Lambda}_n\|_\infty = \|\mathbf{W}_n\mathbf{\Lambda}_n\|_\infty \leq \|\mathbf{W}_n\|_\infty \|\mathbf{\Lambda}_n\|_\infty < \kappa\lambda_{\max} < \infty$ where $\lambda_{\max} = \max_{1\leqslant i\leqslant n}\left\{1/\sum_{j=1}^{n}|w_{i,j}|\right\}$.

(III) $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_\infty < \infty$: Due to Lemma 2.1 (applying the Neumann series and using the subadditivity/submultiplicativity of matrix norms) it follows that: $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_\infty = \frac{1}{1-|\rho|} \leq c \in \mathbb{R}$.

[26] It is assumed that $\sum_{j=1}^{n}|w_{i,j}| > 0$. If $w_{i,j}$ represents a neighboring structure, $\sum_{j=1}^{n}|w_{i,j}| > 0$ is fullfilled if each observation has at least one neighbor.

(IV) $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_1 < \infty$: First a Neumann series and the subadditivity of matrix norms is used: $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_1 \leq \sum_{k=0}^{\infty} |\rho|^k \left\|\left(\widetilde{\mathbf{W}}_n\right)^k\right\|_1$. Note that if $k > 2$: $\left\|\widetilde{\mathbf{W}}_n^k\right\|_1 = \|\boldsymbol{\Lambda}_n (\prod_{l=1}^{k-1} \mathbf{W}_n\boldsymbol{\Lambda}_n) \mathbf{W}\|_1 \leq \|\boldsymbol{\Lambda}_n\|_1 \left\|\prod_{l=1}^{k-1} \mathbf{W}_n\boldsymbol{\Lambda}_n\right\|_1 \|\mathbf{W}_n\|_1 = \|\boldsymbol{\Lambda}_n\|_1 \left\|\prod_{l=1}^{k-1} \boldsymbol{\Lambda}_n\mathbf{W}_n\right\|_\infty \|\mathbf{W}_n\|_1 = \lambda_{\max}\kappa$ holds. Hence $\left\|\left(\mathbf{I}_n - \rho\widetilde{\mathbf{W}}_n\right)^{-1}\right\|_1 \leq 1 + \sum_{k=1}^{\infty} |\rho|^k \kappa\lambda_{\max} < \infty$ ■

3.4.2 Group interactions and spatial econometric modelling

This subsection considers the following empirical problem: The DGP not only contains spatial autocorrelation but also has to account for different spatial interaction parameters, namely within and between groups. The DGP could for example model a housing market with two distinct geographical markets, namely an east-market and a west-market. These two groups [East and West] have the sample sizes n_1 [West- market] and n_2 [East- market]. The overall sample size is denoted by $n = n_1 + n_2$. For simplicity it is assumed that the data are ordered by these particular groups. To account for the within and between group effects the following model specification given in Eq. (3.5) could be used[27].

$$\mathbf{Y}_n = \left(\rho_{11}\hat{\mathbf{W}}_{11} + \rho_{12}\hat{\mathbf{W}}_{12} + \rho_{21}\hat{\mathbf{W}}_{21} + \rho_{22}\hat{\mathbf{W}}_{22}\right)\mathbf{Y}_n + \mathbf{X}_n\beta + \boldsymbol{\epsilon}_n \qquad (3.5)$$
$$\text{where } \epsilon_i \sim i.i.d(0, \sigma^2)$$

Since the data is ordered, the spatial weight matrices have the following simple form: $\hat{\mathbf{W}}_{11} = \begin{pmatrix} \mathbf{W}_{n1,n1} & \mathbf{0}_{n1,n2} \\ \mathbf{0}_{n2,n1} & \mathbf{0}_{n2,n2} \end{pmatrix}$, $\hat{\mathbf{W}}_{12} = \begin{pmatrix} \mathbf{0}_{n1,n1} & \mathbf{W}_{n1,n2} \\ \mathbf{0}_{n2,n1} & \mathbf{0}_{n2,n2} \end{pmatrix}$, $\hat{\mathbf{W}}_{21} = \begin{pmatrix} \mathbf{0}_{n1,n1} & \mathbf{0}_{n1,n2} \\ \mathbf{W}_{n2,n1} & \mathbf{0}_{n2,n2} \end{pmatrix}$, $\hat{\mathbf{W}}_{22} = \begin{pmatrix} \mathbf{0}_{n1,n1} & \mathbf{0}_{n1,n2} \\ \mathbf{0}_{n2,n1} & \mathbf{W}_{n2,n2} \end{pmatrix}$. Note that $\hat{\mathbf{W}}_{11}$, $\hat{\mathbf{W}}_{12}$,$\hat{\mathbf{W}}_{21}$ and $\hat{\mathbf{W}}_{22}$ are partitioned.

The within group effects, given by the terms $\rho_{11}\hat{\mathbf{W}}_{11}\mathbf{Y}$ and $\rho_{22}\hat{\mathbf{W}}_{22}\mathbf{Y}$ would answer the question of how are the Western- ($\rho_{11}\hat{\mathbf{W}}_{11}\mathbf{Y}$) and Eastern- market ($\rho_{22}\hat{\mathbf{W}}_{22}\mathbf{Y}$) affected by themselves. On the other hand the between groups effects, given by the terms $\rho_{12}\hat{\mathbf{W}}_{12}\mathbf{Y}$ and $\rho_{21}\hat{\mathbf{W}}_{21}\mathbf{Y}$ would answer the question of how is the Western market influenced by the Eastern ($\rho_{12}\hat{\mathbf{W}}_{12}\mathbf{Y}$) market and vice versa ($\rho_{21}\hat{\mathbf{W}}_{21}\mathbf{Y}$).

To find the spatial parameter space for the model given in Eq. (3.5) let $\boldsymbol{\Psi}_{n,4}$ be defined by

[27] Of course in order to find asymptotic properties for the spatial parameters it has to be assumed that n_1 and $n_2 \rightarrow \infty$.

$\mathbf{\Psi}_{n,4} := \rho_{11}\hat{\mathbf{W}}_{11} + \rho_{12}\hat{\mathbf{W}}_{12} + \rho_{21}\hat{\mathbf{W}}_{21} + \rho_{22}\hat{\mathbf{W}}_{22}$. Additionally it is assumed[28] that $\max\{||\hat{\mathbf{W}}_{i,j}||_{\varrho}$ where $\varrho \in \{1, \infty\}$ and $i, j \in \{1, 2\}\} \leq 1$. Although not necessary for the following proof, in most applied cases $\mathbf{W}_{n1,n1}$ and $\mathbf{W}_{n2,n2}$ will be symmetric and $\mathbf{W}_{n1,n2} = (\mathbf{W}_{n2,n1})'$.

The following paragraph will sketch the search and proof for the parameter space, the actual proof is given in the appendix: The first step is to use the equation system $\mathbf{y}_1 = \rho_{11}\mathbf{W}_{n1,n1}\mathbf{y}_1 + \rho_{12}\mathbf{W}_{n1,n2}\mathbf{y}_2 + \mathbf{s}_1$ (West) and $\mathbf{y}_2 = \rho_{21}\mathbf{W}_{n2,n1}\mathbf{y}_1 + \rho_{22}\mathbf{W}_{n2,n2}\mathbf{y}_2 + \mathbf{s}_2$ (East) and solve for $\mathbf{y}_1$ and $\mathbf{y}_2$. One condition for solving these equations is that $\mathbf{A}_1 := (\mathbf{I}_{n1} - \rho_{11}\mathbf{W}_{n1,n1})^{-1}$ and $\mathbf{A}_2 := (\mathbf{I}_{n2} - \rho_{22}\mathbf{W}_{n2,n2})^{-1}$ are defined. Therefore, two conditions for the parameter space are: $|\rho_{11}| < 1$ and $|\rho_{22}| < 1$. The second step is to find the restrictions for ρ_{21} and ρ_{12}. Since it is assumed that $|\rho_{11}| < 1$ and $|\rho_{22}| < 1$ one can solve the two equations $\mathbf{y}_1 = (\mathbf{I}_{n1} - \rho_{11}\mathbf{W}_{n1,n1})^{-1}$ $(\rho_{12}\mathbf{W}_{n1,n2}\mathbf{y}_2 + \mathbf{s}_1)$ and $\mathbf{y}_2 = (\mathbf{I}_{n2} - \rho_{22}\mathbf{W}_{n2,n2})^{-1}$ $(\rho_{21}\mathbf{W}_{n2,n1}\mathbf{y}_1 + \mathbf{s}_2)$. Inserting $\mathbf{y}_2$ into $\mathbf{y}_1$ yields: $\mathbf{y}_1 = \mathbf{A}_1\ (\rho_{12}\mathbf{W}_{n1,n2}\mathbf{A}_2\ (\rho_{21}\mathbf{W}_{n2,n1}\mathbf{y}_1 + \mathbf{s}_2) + \mathbf{s}_1)$. It is possible to solve this equation for $\mathbf{y}_1$ if $\|\mathbf{A}_1\rho_{12}\mathbf{W}_{n1,n2}\mathbf{A}_2\rho_{21}\mathbf{W}_{n2,n1}\|_1 < 1$. This inequality is satisfied if $\frac{|\rho_{12}|}{|1-\rho_{22}|}\frac{|\rho_{21}|}{|1-\rho_{11}|} < 1$. Hence it is shown that the following two conditions fulfill the parameter space properties (1)-(3) for the DGP given in Eq. (3.5).

(1) $|\rho_{11}| < 1$ and $|\rho_{22}| < 1$

(2) $\frac{|\rho_{12}|}{|1-\rho_{22}|}\frac{|\rho_{21}|}{|1-\rho_{11}|} < 1$

This parameter space is significantly larger than the one proposed by Lee and Liu (2010). For example if $|\rho_{11}| = |\rho_{22}| = 1/3$ the area bounded by the dashed line in figure (1) represents the Lee and Liu parameter space and the area bounded by the solid lines represents the parameter space due to inequality $\frac{|\rho_{12}|}{|1-\rho_{22}|}\frac{|\rho_{21}|}{|1-\rho_{11}|} < 1$.

[28] The appendix shows that if $\mathbf{W}_{n1,n1}$, $\mathbf{W}_{n2,n2}$, $\mathbf{W}_{n2,n1}$ and $\mathbf{W}_{n1,n2}$ are row standardized, $\mathbf{W}_{n1,n1}$ and $\mathbf{W}_{n2,n2}$ can be written as $\mathbf{W}_{n1,n1} = \Lambda_{n1}\overline{\mathbf{W}}_{n1,n1}$ and $\mathbf{W}_{n2,n2} = \Lambda_{n2}\overline{\mathbf{W}}_{n2,n2}$ where Λ represents the row- standardizing and both $\overline{\mathbf{W}}$ are symmetric and $\|\overline{\mathbf{W}}\|_{\infty}$, $\|\mathbf{W}_{n1,n2}\|_{\infty}$, $\|\mathbf{W}_{n2,n1}\|_{\infty} < \infty$ hold the same spatial parameter space can be applied for the DGP given in (3.6)

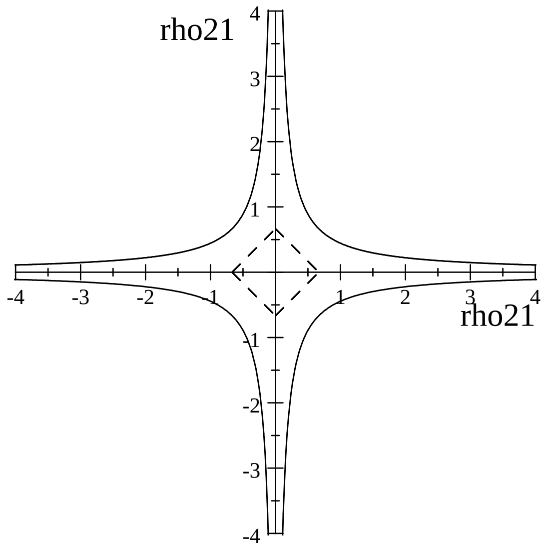

Figure 3.1: Visualization of the different parameter spaces suitable to describe a two group interaction model, where $|\rho_{11}| = |\rho_{22}| = 1/3$

Figure (3.1) shows quite dramatically that one should take the inner structure of the $\mathbf{W}_n$- matrices into account in order to find parameter spaces. This is to some extent similar to the "practitioners" approach of row standardizing. The more knowledge is present about the inner structure of the spatial lags, like in this example where the weight matrices are partitioned, the more precise and in this particular example larger the resulting spatial parameter space gets. This can also be seen in the next application of the new parameter space definition.

3.4.3 Geographic scale and the spatial parameter space

This subsection explores panels where the number of observation units (n) is fixed. Since n is fixed, the number T of time periods has to go to infinity in order to derive the asymptotic properties of possible estimators. Let N denote the overall sample size, namely $N = nT$. The vector $\mathbf{Y}_N$ has the elements $y_{1,1}, y_{2,1}, ..., y_{n,1}, ..., y_{n,T}$. A possible spatial DGP reflecting these properties is given by Eq. (3.6)

$$\mathbf{Y}_N = \rho \left(\mathbf{I}_T \otimes \mathbf{W}_n \right) \mathbf{Y}_N + \mathbf{X}_N \beta + \boldsymbol{\epsilon}_N \text{ where } \epsilon_i \sim i.i.d(0, \sigma^2) \tag{3.6}$$

where $\otimes$ denotes the Kronecker product. Finding a spatial parameter space for Eq. (3.6) can be achieved by the following reasoning: Obviously it has to be ensured that the inverse of $\mathbf{I}_N - \rho\left(\mathbf{I}_T \otimes \mathbf{W}_n\right)$ exists. Since $Sp\left(\mathbf{I}_T \otimes \mathbf{W}_n\right) = Sp\left(\mathbf{W}_n\right)$ the condition $1/\rho \notin Sp(\mathbf{W}_n)$ follows directly. Note that the number of elements in $Sp(\mathbf{W}_n)$ is always smaller or equal n. Since n is fixed $\|\mathbf{W}_n\|_\infty < \infty$, $\|\mathbf{W}_n\|_1 < \infty$ and $\left\|(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\right\|_{1,\infty} < \infty$ have to hold for $1/\rho \notin Sp(\mathbf{W}_n)$. Observe that $\left(\mathbf{I}_N - \rho\left(\mathbf{I}_T \otimes \mathbf{W}_n\right)\right)^{-1} = \mathbf{I}_T \otimes (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ and therefore $\left\|\left(\mathbf{I}_N - \rho\left(\mathbf{I}_T \otimes \mathbf{W}_n\right)\right)^{-1}\right\|_{1,\infty}$ $< \infty$ for all $T \in \mathbb{N} \cup \{\infty\}$ as long $1/\rho \notin Sp(\mathbf{W}_n)$. Thus the spatial parameter space for Eq. (3.6) is given by $\mathbb{R}\backslash\{1/\widetilde{\tau}_n$ where $\widetilde{\tau}_n \in Sp(\mathbf{W}_n)\}$.

The previous paragraph showed above all two noteworthy characteristics of the spatial parameter space for the DGP given in Eq. (3.6): First, $\mathbb{R}\backslash\{1/\widetilde{\tau}_n$ where $\widetilde{\tau}_n \in Sp(\mathbf{W}_n)\}$ is tremendously larger than the parameter spaces considered by the literature so far. Second, the spatial parameter space is a function of n which is fixed by assumption and thus can offer an interesting interpretation: The implicit n dependence of the DGP dynamics can for example be seen as a consequence of different geographic scales. To illustrate this point let $n = 27$ represent the countries in the European Union and $y_{i,t}$ their corresponding GDP- growth rate. If n (and hence $\mathbf{W}_n$) is now changed to $\widetilde{n} = 271$ ($\widetilde{n}$ now represents the NUTS-2 regions of the European Union) this simply would reflect a change in the geographical scale. Since a change in the geographical scale is often associated with a change in the model dynamics a parameter space as function dependent on n seems plausible and would simply reflect some influences of the geographical scale on the GDP/ gross regional product.

It is important that the previous reasoning is not suggesting to use $\mathbb{R}\backslash\{1/\widetilde{\tau}_n$ where $\widetilde{\tau}_n \in Sp(\mathbf{W}_n)\}$ for every spatial panel data DGP. Whether it makes sense depends on what the DGP should describe. Consider for example house prices in a real estate market. Economic theory would suggest that for perfect markets the observed prices are independent of the market size n. As a result the DGP- dynamics should be independent of n as well and consequently the parameter space must be independent of n too, like for example the classical $(-1, 1)$- spatial parameter space, after normalizing the weight matrix.

3.5 Concluding remarks

Unlike the time series literature, there has been not much effort in the spatial econometric literature to substantially examine the parameter spaces for spatial econometric models. This paper raises three important issues concerning spatial parameter spaces:

First, current parameter space concepts and practical approaches are inadequate. This point is supported by the three following observations:

Since the Kelejian and Prucha (2010) parameter space only considers the eigenvalues of the spatial weight matrix, it is only concerned about the existence of the DGP and not its stability. Hence it can result in nonstationary DGPs if for example the spatial weight matrix mimic a process from the time series literature.

The Lee and Liu (2010) parameter space can result in too small parameter spaces, especially if it is confronted with row standardized weight matrices. Although it will always result in stable/stationary DGPs it has to be seen as too restrictive.

Since neither the Kelejian and Prucha (2010) nor the Lee and Liu (2010) parameter space can be used as a mathematical foundation for the "practitioners" approach of row standardizing, this approach lacks a theoretical basis.

Second, a useful spatial parameter space can be defined indirectly via desired mathematical properties. These properties show clearly the difference between the necessary conditions for the existence of the DGP and its stability/stationary. Additionally it shows that the Lee and Liu (2010) parameter space can be seen as a special case of this new parameter space definition.

Third, the advantages of the new parameter space concept lies in its ability to account for the inner structure of the spatial lag(s). Hence it is possible to derive more precise and in some cases *larger spatial parameter spaces*. This can be verified with the help of three practical examples:

Section 3.4.1 shows that the "practitioners" approach of row standardizing under the practical assumption that if the weight matrix before row normalizing was symmetric, the approach always yields stable DGPs. Hence, it is possible to give the "practitioners" approach of row standardizing a mathematical foundation.

Section 3.4.2 handles models with different spatial interactions reflecting the assumed group structure. It is possible to find substantially larger spatial parameter spaces than the ones previously considered by the literature.

Finally, the example of spatial panels with fixed n raises two interesting issues: First it shows that under certain assumptions about the DGP, it is possible to use almost the whole real line as the spatial parameter space. Second, it suggested that the implicit n dependence of the DGP dynamics can for example be seen as a consequence of different geographic scales.

These results highlight the importance of the spatial parameter space. Therefore, applied researchers should be encouraged to deal with their parameter space in more detail since it could be larger and reveal some interesting dynamics of the spatial DGP.

3.6 Change in model interpretation

We investigate in this subsection how the model interpretation[29] would change if we observe for real world data that $\hat{\rho} \notin (-1, 1)$ for the following Spatial Durbin Model

$$\mathbf{Y}_n = \rho \mathbf{W}_n \mathbf{Y}_n + \mathbf{X}_n \beta + \mathbf{W}_n \mathbf{X}_n \gamma + \epsilon_n \text{ where } \epsilon_i \sim i.i.d(0, \sigma^2) \tag{3.7}$$

where $\hat{\rho}$ is the estimate of ρ. In order to reduce the analytical burden we assume that $\mathbf{W}_n$ is row standardized. LeSage and Pace (2009) suggest to use average direct, indirect and total effects, which are given in the Eqs. (3.8) to (3.10) respectively for a Spatial Durbin model:

[29] This subchapter was added in order to provide one of this thesis major arguments. However from a publication perspective it was not added to the paper, since such consequences for the model interpretation would have needed a discussion why we actually do not observe such dynamics in real world data. Hence the short discussion of the numerical algorithms used for maximum likelihood estimation in this thesis's Introduction.

$$\textit{direct effect}_k = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial y_i}{\partial x_{i,k}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\beta_k + \gamma_k}{1-\rho\tau_i} \tag{3.8}$$

$$\textit{total effect}_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y_i}{\partial x_{j,k}} = \frac{\beta_k + \gamma_k}{1-\rho} \tag{3.9}$$

$$\textit{indirect effect}_k = \textit{total effect}_k - \textit{direct effect}_k \tag{3.10}$$

Note that per construction $|\tau_i| \leq 1$. Investigating Eq. (3.8) we find for $\hat{\rho} \notin (-1, 1)$ that the elements in the sum may no longer have the same sign. Therefore averaging over these effects, like it is suggested in Eq. (3.8) may results in a loss of information, since there exists the possibility that the partial derivative with respect to an explanatory variable has different signs among the observations. Hence we have to split up each effect into an average for each sign class. It proves to be convenient to define the Matrix $\mathbf{S}(\rho) := (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ with the typical element $s_{i,j}(\rho)$. Let N denote the set of observed observations $N = \{1, ..., n\}$. We split up the observation set according to the following formula:

$$N = N_1 \cup N_2 \text{ where } N_1 = \{i \in N | s_{i,i}(\rho) < 0\} \text{ and } N_2 = \{i \in N | s_{i,i}(\rho) > 0\} \tag{3.11}$$

For the following illustration let $n = 10$, β_k, $\gamma_k > 0$ and $\mathbf{W}_n$ be a row standardized one forward one behind pattern: If $\rho = 3$ then $N_1 = \{2, 4, 7, 9\}$ and $N_2 = \{1, 3, 5, 6, 8, 10\}$ and therefore four observations would have a negative direct effect. If $\rho = -8$ then $N_1 = \{1, .., N\}$ and $N_2 = \{\oslash\}$ and therefore all observations would have negative direct effects, although the overall total effects are still positive and β_k, $\gamma_k > 0$. We can define for each group ($g \in \{1, 2\}$) their corresponding average direct[30], indirect and total effects if $|N_g| > 0$:

$$\textit{direct effect}_{k,g} = \frac{1}{|N_g|}\sum_{i \in N_g}\frac{\partial y_i}{\partial x_{i,k}} = \frac{1}{|N_g|}\sum_{i \in N_g} s_{i,i}(\rho)\,(\beta_k + \gamma_k) \tag{3.12}$$

[30] Another possibility would be to define the average indirect and total effects in more granular structure, like: How the average observation in group 1 is influenced by the observations in group 2 or by the (other) observations of group 1.

$$\textit{total effect}_{k,g} = \frac{1}{|N_g|} \sum_{i \in N_g} \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_{j,k}} = \frac{n}{|N_g|} \frac{\beta_k + \gamma_k}{1 - \rho} \tag{3.13}$$

$$\textit{indirect effect}_{k,g} = \textit{total effect}_{k,g} - \textit{direct effect}_{k,g} \tag{3.14}$$

In Eq. (3.13) the total effects are defined as the influence of a change in the k^{th} explanatory variable for all observations on the group average. It would also be possible to define the total effects as the summation of the average influences over the group where the average influence is defined by the change in the k^{th} explanatory variable in an observation. Note that these differences in the definition do not effect the following interpretation.

Let us consider the example from Section 3.4.3 "geographic scale and the spatial parameter space", where the observations represented the NUTS-2 regions. Furthermore we are interested in influence of growth determinants and use "openness" as explanatory variable. Openness measures how well a region is integrated into the overall trade-network. Additionally let $\rho < -1$ and the to "openness" corresponding explanatory coefficients $\beta_k, \gamma_k > 0$. Therefore there could exist two groups of regions, one where $s_{i,i}(\rho) < 0$ (labelled group 1) and the other where $s_{i,i}(\rho) > 0$ (labelled group 2). Note that the due to $\rho < -1$ the total effects for each group are positive. However if we increase openness for the average region in group 1 then the effect on its own growth rate would be negative while the reverse is true for the average region in group 2. If due to the assumption $|\rho| < 1$ such effects are ignored we could face the following situation: Due to their experience, policy makers from group 2 are arguing that their counterparts form group 1 should also increase their openness although the opposite is the case.

Note that in this particular example the geography of the regions is responsible for the splitting up of the regions into two groups. This example illustrates that by "expanding" the parameter space beyond $(-1, 1)$ we are naturally defining geographical groups that react similar to changes in the explanatory variables. Since such changes are group specific, such findings would have strong policy implications.

3.7 Comments from the *Econometric Theory* referee and some notes

The second paper was submitted to *Econometric Theory* and rejected. One referee did not write a report.

3.7.1 First referee‘s report and comments

Main issues

1. At p. 2 the author states that "The spatial econometric literature, so far has not dealt with the issue of the spatial parameter space in detail". Quite similar or even equivalent to Condition 2 in Definition 1 have been imposed on the parameter space of spatial models in a lot of technical papers studying asymptotic properties of estimators in spatial econometric models. Also, for an example of a paper discussing the parameter space of spatial models specifically, and outside an inferential framework, see Elhorst, Lacombe, Piras (2012)".

The thesis' author argues against this remark as follows: The whole point of this paper was that Condition 2 in Definition 1 (the solvability of spatial DGP) is not sufficient. Hence parameter spaces purely based on Condition 2 in Definition 1 can not, in general, be stable. As a result the work of Elhorst, Lacombe and Piras (2012), as it is based on insufficient assumptions, has to be considered as wrong.

2. Kelejian and Prucha (2010) formulation requires the spectral radius of W to be nonzero. It should not be used for triangular W's, such as in Eq. (2) in the paper or more generally those of time series models, because such matrices have zero spectral radius.

The thesis' author argues against this remark as follows: First Kelejian and Prucha (2010) never assume that the spectral radius has to be different form zero. Hence the claim made by the second paper of the thesis has to be seen as valid. As a reaction to this particular comment, an example was added to the second paper of the thesis where a nontriangular matrix is normalized as it is suggested in Kelejian and Prucha (2010), the spatial autocorrelation parameter is set to 0.95 and then largest absolute value of the spatial multiplier is given as a function of the sample size. Values as high as $\sim 10^{25}$ are observed (for a sample size of 150) and thus this spatially autocorrelated DGP can not be seen as stationary or stable.

Suggestions

a.)*Stationary and stability are, surprisingly never defined in the paper. This causes ambiguity because such terms have been given different meanings in different contexts.*

b.)*I did not understand what the author means by "it is discussed that the practice of row standardizing lacks a mathematical foundation".*

c.)*p.7. "In order to derive estimator properties for DGPs like (1) or (3), stability is required." Stability is certainly not required, but it is usually concenient. Also non-stable processes do not seem to be very useful empirically.*

d.)*Proofs of well known results, e.g. the second an third proofs in the appendix, can be obmitted.*

Chapter 4 Paper III: Spatial filtering and model interpretation for Spatial Durbin Models

Abstract[31]**:** Spatial filter for spatial autoregressive models like the spatial Durbin Model have seen a great interest in the recent literature. Pace et al. (2011) show that the spatial filtering methods developed by Griffith (2000) have desirable estimation properties for some parameters associated with spatial autoregessive models. However, spatial filtering faces two conceptual weaknesses: First the estimated parameters lack in general, and especially for the Spatial Durbin Model a proper interpretation. Second, there exists an inherent trade-off between the estimator bias and its efficiency, depending on the spectrum of the used spatial weight matrices.

This paper tackles both problems by introducing a new four step estimation procedure based on the eigenvectors of the spatial weight matrix. This new estimation procedure estimates all parameters of interest in a Spatial Durbin model and thus allows for a proper model interpretation. Additionally the estimation procedure's efficiency is only marginally influenced by the number of added eigenvectors, which allows us to use approximately 95% of the available eigenvectors. By using Monte Carlo simulations we observe that the estimation procedure has a lower (or equal) bias and smaller (or equal) sample variance as the corresponding Maximum Likelihood estimator based on normality.

*Keywords***:** spatial filtering, Spatial Durbin Model, alternative estimation technique

JEL-Classification Codes: C13,C18, C31, C35

4.1 Introduction

Spatial filtering is a very popular[32] estimation method for spatial (autoregressive) models. These

31 This paper was submitted to Geographical Analysis in March 2012. Futhermore it was presented at various Dissertationseminars of the Vienna University of Economics and Business and University of Innsbruck.

32 see for example Cuaresma and Feldkircher (2010), Cuaresma et. al. (2009), Tiefelsdorf and Griffith (2007) or Fischer and Griffith (2008).

spatial autoregressive models assume an autocorrelation structure, which is represented by the so called spatial weight matrix $\mathbf{W}_n$. The idea behind spatial filtering is to approximate $\mathbf{W}_n$ via a subset of the corresponding eigenvectors and eigenvalues. If this approximation holds one can use this subset of eigenvectors as explanatory variables in a linear regression framework in order to control for the spatial autoregressive model nature and therefore reduce the potential bias in the ordinary least squares regression. Pace et. al. (2011) for example suggest to apply the Frisch Waugh theorem to the subset of eigenvectors and as a result filter out the spatial dependence in the model. The virtue of this approach is that unlike likelihood based estimation approaches, no distributional assumption about the error term is necessary. Pace et. al. (2011) argue that this is especially useful for models where the dependent variables represent binary, discrete choice outcomes or Poisson distributed counts and where we observe an autoregressive structure in the dependent variable.

Although spatial filtering has desirable properties, this estimation approach faces in general two intrinsic problems: *First*, spatial filtering as the name suggests, filters out the spatial autoregressive term and therefore we are left without an estimate for the spatial autocorrelation parameter. This is a fundamental drawback if the model is spatial autoregressive in the dependent variable, since as LeSage and Pace (2007) point out, the model partial derivatives of the dependent variable with respect to explanatory variables are in general a function of the spatial autocorrelation parameter. Thus, in general the coefficients resulting from spatial filtering lack a proper interpretation. *Second*, spatial filtering is a form of model approximation and therefore the estimation results strongly depend on the approximation quality. The approximation quality however depends on the number of eigenvectors used as explanatory variables. If we would use all available eigenvectors (perfect model approximation) we would have n additional explanatory variables and therefore are left with a model that has to estimate more parameters than actual observations. On the other hand if too few eigenvectors are used the resulting estimation suffers from considerable bias. Therefore the spatial filtering method has always to make a trade-off between estimation bias and estimation efficiency. Hence it's not surprising that different[33] approaches exist for constructing an "optimal" subset of eigenvectors. Finally the performance of spatial filtering strongly depends on the spatial weight matrix's spectrum.

This paper tackles both intrinsic problems associated with spatial filtering for the so-called spatial Durbin model (SDM). We provide a new iterative estimation procedure based on spatial filtering that results in estimates for each parameter associated with the SDM. Therefore we are able to calculate all partial derivatives associated with the model and therefore can correctly interpret the implied model dynamics. The proposed four-step estimation method can, due to its construction, incorporate almost all eigenvectors associated with the spatial weight matrix without affecting the efficiency of the parameter estimation. Therefore the previously described trade-off between estimation bias and estimation efficiency is no longer present.

The paper is organized as follows. The next section provides some background to the SDM and gives a short introduction to spatial filtering as it is suggested in Pace et. al. (2011). The second section describes our proposed iterative four-step estimation method based on spatial filtering and provides a short discussion regarding model interpretation and the calculation of the SDM effect-measures' standard deviation. The following section provides the set up for our Monte Carlo simulation where we compare the four-step estimation procedure with the maximum likelihood estimator based normality. Additionally this section provides three different summary measures for comparing the performance of both estimators. In the section "Monte Carlo results" we find that the four-step estimation procedure performs as well as the maximum likelihood estimator. We also find that for our experimental design, bootstrapping seems to be more preferable as estimation technique for the estimator's variance compared to the Monte Carlo simulation suggested by

33 see for example Getis and Griffith (2002)

LeSage and Pace (2007), which is based on the asymptotic variance of the maximum likelihood estimator. Additionally we discuss some numerical implications of the four step estimation technique and we find that whereas traditional spatial filtering has a trade-off between the estimation bias and estimation efficiency, the four step estimator has a trade-off between estimation bias and computational time. The last section concludes and summarizes this paper.

4.2 The spatial Durbin Model and "classical" spatial filtering

This section first provides the Spatial Durbin model and the associated assumptions for spatial filtering and then provides the intuition behind spatial filtering as suggested by Pace et. al (2011).

Notation: If the matrix $\mathbf{W}_n$ is symmetric then $\mathbf{W}_n$ can be written as $\bar{\mathbf{D}}_n \bar{\mathbf{\Lambda}}_n \bar{\mathbf{D}}_n^{-1}$ where $\bar{\mathbf{D}}_n$ is the matrix containing the eigenvectors and $\bar{\mathbf{\Lambda}}_n$ the eigenvalue matrix. A subset of the eigenvectors of $\bar{\mathbf{D}}_n$ is denoted by $\mathbf{D}_n$ and the corresponding eigenvector matrix by $\mathbf{\Lambda}_q$. The number of columns of $\mathbf{D}_n$ is denoted by q. The i^{th} diagonal element of the eigenvalue matrix is denoted by λ_i. The Operator $tr()$ applied to a matrix is the matrix's trace. The operators $E[]$ and $Var[]$ denote the expected value and the variance respectively. $\mathbf{I}_n$ denotes an identity matrix of dimension n. Let $\mathbf{Y}_n \in \mathbb{R}^{n \times 1}$. One can write $\mathbf{Y}_n$ also as $\mathbf{Y} = (y_1, y_2, ..., y_n)'$. Some proofs and useful Lemmas are given in the (technical) Appendix. If $\mathbf{X}_n$ is a n by k matrix then $rank(\mathbf{X}_n)$ is the number of linear independent columns. If x is a real number $round(x)$ refers to the nearest integer.

The following data generating process is referred to by the literature as Spatial Durbin model and is the focus of our analysis:

$$\mathbf{Y}_n = \rho_o \mathbf{W}_n \mathbf{Y}_n + \mathbf{X}_n \beta_0 + \mathbf{W}_n \mathbf{X}_n \gamma_0 + \epsilon_n \text{ where } \epsilon_i \sim i.i.d(0, \sigma_0^2) \tag{4.1}$$

In Eq. (4.1) $\mathbf{X}_n$ represents the n by k matrix of (finite) explanatory variables where $rank(\mathbf{X}_n) = k$. The parameters in Eq. (4.1) ρ_0, β_0 and γ_0 are the coefficients to be estimated and the ϵ_i are independently and identically distributed with zero mean and finite variance σ_0^2. $\mathbf{W}_n$ represents the symmetric n by n spatial weights matrix of known constants. The diagonal entries of $\mathbf{W}_n$

are assumed to be zero[34]. The SDM incorporates various representations of spatial DGPs like the Spatial Autoregressive Model and the Spatial Error Model[35]. Throughout this paper we maintain additionally the following (central) assumptions:

i.) $\mathbf{W}_n$ can be seen as deterministic and is normalized such that the absolute maximum eigenvalue is smaller or equal one

ii.) $\mathbf{W}_n$ can be approximated by $\mathbf{W}_n \approx \mathbf{D}_n \mathbf{\Lambda}_q \mathbf{D}_n'$

iii.) $\rho_o \in (-1, 1)$

iv.) $E[\mathbf{X}_n' \epsilon_n] = 0$

Due to assumptions (i) and (iii) we can solve the DGP for $\mathbf{Y}_n$ and end up with Eq. (4.2):

$$\mathbf{Y}_n = (\mathbf{I}_n - \rho_o \mathbf{W}_n)^{-1} (\mathbf{X}_n \beta_0 + \mathbf{W}_n \mathbf{X}_n \gamma_0 + \epsilon_n) \text{ where } \epsilon_i \sim i.i.d(0, \sigma_0^2) \tag{4.2}$$

Due to the following fundamental identity $(\mathbf{I}_n - \rho_o \mathbf{W}_n)^{-1} = \mathbf{I}_n + (\mathbf{I}_n - \rho_o \mathbf{W}_n)^{-1} \rho_o \mathbf{W}_n$ we can write Eq. (4.3) as

$$\mathbf{Y}_n = \mathbf{X}_n \beta_0 + \epsilon_n + \mathbf{A}_n (\mathbf{X}_n \beta_0 + \mathbf{W}_n \mathbf{X}_n \gamma_0 + \epsilon_n)$$
$$\text{where } \mathbf{A}_n = (\mathbf{I}_n - \rho_o \mathbf{W}_n)^{-1} \rho_o \mathbf{W}_n \tag{3}$$

The main idea behind Griffith's approach (see Griffith (2003)) is to use an approximation for the matrix $\mathbf{A}_n$ in order to construct a projector $\mathbf{M}_D$ such that: $\mathbf{M}_D \mathbf{A}_n \approx \mathbf{0}$ where $\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}_n (\mathbf{D}_n' \mathbf{D}_n)^{-1} \mathbf{D}_n'$. Since eigenvectors are orthogonal per construction $\mathbf{M}_D$ can be simplified to: $\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}_n \mathbf{D}_n'$. Given such a projector the estimation problem written in Eq. (4.3) is reduced to a simple linear model where ordinary least squares can be applied

$$\mathbf{M}_D \mathbf{Y}_n \approx \mathbf{M}_D \mathbf{X}_n \beta_0 + \mathbf{M}_D \epsilon_n. \tag{4.4}$$

However it is not clear how many eigenvectors are necessary for a reasonable model approximation. Given assumption (iii) we can write $\mathbf{A}_n$ as a Neumann Series $(\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1} \rho_0 \mathbf{W}_n =$

[34] Although it is possible to derive parameter spaces for $\mathbf{W}_n$ matrices where the diagonal elements are not zero, it is not common in applications.

[35] For more details to the assumptions and properties of the data generating process stated in (4.1), see Elhorst (2010)

$\sum_{k=1}^{\infty} \rho_0^k \mathbf{W}_n^k$ and therefore the eigenvectors corresponding to the absolute largest[36] eigenvalues are a good approximation for $\mathbf{A}_n$. Getis and Griffith (2002) for example compare the Moran's I statistic with the Getis G_i local statistic.

Given a reasonable number of eigenvectors for approximating $\mathbf{M}_D \; \mathbf{A}_n \approx \mathbf{0}$ Pace et. al. (2011) use the simple OLS estimator, $\hat{\beta} = (\mathbf{X}_n'\mathbf{M}_D\mathbf{X}_n)^{-1} \; \mathbf{X}_n'\mathbf{M}_D\mathbf{Y}_n$, as an estimator for β_0 given in the DPG by Eq. (4.4). If the approximation given in Eq. (4.4) holds, this estimator is unbiased since $E[\hat{\beta}] = \beta_0 + E\left[(\mathbf{X}_n'\mathbf{M}_D\mathbf{X}_n)^{-1} \mathbf{X}_n'\mathbf{M}_D\epsilon_n\right] = \beta_0$ where the last equal sign is due to assumptions (iv) and (i). Pace et. al. (2011) provide examples of $\mathbf{W}_n$-matrices where the estimator $\hat{\beta}$ has these desirable properties.

This approach however faces two inherent weaknesses: First the $\hat{\beta}$ coefficient vector lacks a proper interpretation, since ceteris paribus model interpretations are associated with the partial derivatives of $\mathbf{Y}_n$ with respect to the explanatory variables $\mathbf{X}_n$. Pace et. al. (2011) label this "a philosophical issue regarding the spatial filtering method". The second weakness of spatial filtering stems from the inherent estimator trade-off between estimation bias and estimation efficiency. Each added eigenvector adds a degree of freedom to the OLS estimation. If all eigenvectors were used, we would have n equations for estimating $n + 2k$ unknowns. This is reflected in the projector $\mathbf{M}_D$ so that if $q \to n$ then $\mathbf{M}_D = \mathbf{0}$. Therefore not all $\mathbf{W}_n$-matrices are suitable for spatial filtering, depending on the number of absolute eigenvalues near one. These conceptual weaknesses motivate the next section where a new estimation procedure, based on the spatial filtering model approximation is presented, which can tackle both weaknesses of the spatial filtering approach.

4.3 Estimation procedure and model interpretation

This section first provides the estimation details regarding the new estimator for $(\rho_o, \beta_0, \gamma_0)'$. This estimator is based on four different estimation steps. Finally this section provides an approxima-

[36] Note if $|\rho_0| \geq 1$ then one should no longer use the biggest absolute eigenvalues, since the choice of the eigenvectors depends on the true parameter value ρ_0.

tion for the variances associated with model parameters and their implied direct/indirect and total effects.

First Step

In order to estimate the Spatial Durbin model given in Eq. (4.1), we first use a projector for eliminating the explanatory variables $\mathbf{X}_n$ and $\mathbf{W}_n\mathbf{X}_n$ and end up with the data generating process given by Eq. (4.5):

$$\bar{\mathbf{Y}}_n = \rho_o \mathbf{M}_x \mathbf{W}_n \mathbf{Y}_n + \bar{\epsilon}_n \approx \mathbf{M}_x \mathbf{D}_n \theta_0 + \bar{\epsilon}_n \text{ where } \mathbf{M}_x = \mathbf{I}_n - \mathbf{Z}_n \left(\mathbf{Z}_n' \mathbf{Z}_n\right)^{-1} \mathbf{Z}_n',$$

$$\bar{\epsilon}_n = \mathbf{M}_x \epsilon_n,\ \bar{\mathbf{Y}}_n = \mathbf{M}_x \mathbf{Y}_n,\ \theta_0 = \rho_o \mathbf{\Lambda}_q \mathbf{D}_n' \mathbf{Y}_n \text{ and } \mathbf{Z}_n = [\mathbf{X}_n, \mathbf{W}_n \mathbf{X}_n] \quad (4.5)$$

By applying the projector $\mathbf{M}_x$ the resulting model is only influenced by the spatial autoregressive lag and the error term. By using the model approximation via eigenvectors we find the relationship between the spatial lag and the eigenvectors given by $\rho_o \mathbf{M}_x \mathbf{W}_n \mathbf{Y}_n \approx \mathbf{M}_x \mathbf{D}_n \theta_0$. Additionally we are able to find an estimator for σ_0^2 which is given by $\hat{\sigma}^2 = 1/(n - 2k - q)\hat{e}'\hat{e}$ where $\hat{e} = \mathbf{M}_x \mathbf{M}_n \mathbf{Y}$.

Second Step

Applying the OLS- estimator for Eq. (4.5) yields: $\rho_o \widehat{\mathbf{M}_x \mathbf{W}_n} \mathbf{Y}_n = \mathbf{M}_x \mathbf{D}_n \widehat{\theta}$ where $\widehat{\theta} = (\mathbf{D}_n'$ $\mathbf{M}_x$ $\mathbf{D}_n)^{-1}$ $\mathbf{D}_n'$ $\bar{\mathbf{Y}}_n$. Given $\rho_o \mathbf{M}_x \mathbf{W}_n \mathbf{Y}_n \approx \mathbf{M}_x \mathbf{D}_n \theta_0$ the following estimator for ρ_o seems to be "natural": $\widehat{\rho}_1 = (\mathbf{Y}_n' \mathbf{W}_n' \mathbf{M}_x\ \mathbf{W}_n \mathbf{Y}_n)^{-1}\ \mathbf{Y}_n' \mathbf{W}_n'\ \mathbf{M}_x \mathbf{D}_n \widehat{\theta}$ where we regress $\mathbf{M}_x \mathbf{W}_n \mathbf{Y}_n$ on $\mathbf{M}_x \mathbf{D}_n \widehat{\theta}$.

Third Step

However, the estimator $\widehat{\rho}_1$ is only (asymptotically) unbiased if $\sigma_0^2 \to 0$. Theorem C.1 in the Appendix derives the asymptotically expected value of $\widehat{\rho}_1$ given that $\mathbf{M}_x \mathbf{W}_n = \mathbf{D}_n \mathbf{\Lambda}_q \mathbf{D}_n'$: $\lim_{n\to\infty} E\left[\hat{\rho}_1\right] = \rho_0 + \frac{\sigma_0^2 tr\left(\mathbf{S}_n' \mathbf{W}_n' \mathbf{M}_x \overline{\mathbf{DD}}_n\right)}{\left(\sigma_0^2 + \sigma_\mu^2\right) tr(\mathbf{S}_n' \mathbf{W}_n' \mathbf{M}_x \mathbf{W}_n \mathbf{S}_n)}$ where $\overline{\mathbf{DD}}_n = \mathbf{D}_n \left(\mathbf{D}_n' \mathbf{D}_n\right)^{-1} \mathbf{D}_n'$, $\mathbf{S}_n = \left(\mathbf{I}_n - \rho_0 \mathbf{W}_n'\right)^{-1}$ and $\sigma_\mu^2 = Var(\mathbf{X}_n \beta_0 + \mathbf{W}_n \mathbf{X}_n \gamma_0)$. Hence, we can construct the following asymptotically unbiased estimator, given that $\mathbf{M}_x\mathbf{W}_n = \mathbf{D}_n\mathbf{\Lambda}_n\mathbf{D}_n^{-1}$:

$$\hat{\rho} = \arg \min_{\rho\in(-1,1)} \left| \widehat{\rho}_1 - \rho - \frac{\sigma_0^2 tr\left((\mathbf{I}_n - \rho\mathbf{W}_n')^{-1}\mathbf{W}_n'\mathbf{M}_x\overline{\mathbf{DD}}_n\right)}{\mathbf{Y}_n'\mathbf{W}_n'\mathbf{M}_x\mathbf{W}_n\mathbf{Y}_n} \right| \tag{4.5}$$

Since $\mathbf{W}$ is symmetric, the Appendix shows additionally that $tr(\ (\mathbf{I}_n-\ \ \rho\mathbf{W}_n')^{-1}\ \mathbf{W}_n'\mathbf{M}_x\ \overline{\mathbf{DD}}_n) = \sum_{i=\max(2k,q)}^{n} \frac{\lambda_i}{1-\rho\lambda_i}$. Hence the optimization procedure given by Eq. (4.5) is a computational simple nonlinear minimization problem. In order to optimize Eq. (4.5) we need an estimator for σ_0^2, where we use the estimator from Step one.

Fourth Step

Given an estimator for $\hat{\rho}$ we can use the spatially filtered dependent variables in an OLS regression to get estimates for $(\hat{\beta}', \hat{\gamma}')'$, given in the following Eq. (4.6):

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = (\mathbf{Z}_n'\mathbf{Z}_n)^{-1}\mathbf{Z}_n'(\mathbf{I}_n - \hat{\rho}\mathbf{W}_n)\mathbf{Y}_n \tag{4.6}$$

Note that this four step estimation procedure can use a large magnitude of eigenvectors without increasing the overall degrees of freedom, since the second and third step reduce the fitted eigenvectors to a single number, namely the spatial autocorrelation parameter $\hat{\rho}$. We suggest to set the number of chosen eigenvalues equal to: $q = round(0.95(n-2k))$. However, this large number of q might leads to an inefficient estimator for σ_0^2. We therefore suggest to use $\hat{\sigma}^2 = 1/(n-2k-q)\hat{e}'\hat{e}$ (where $\hat{e} = \mathbf{M}_x\mathbf{M}_n\mathbf{Y}_n$) only as an initial estimator and then update it with $\hat{\sigma}_{up}^2 = 1/(n-2k)\hat{e}'\hat{e}$ where $\hat{e} = \mathbf{Y}_n - \hat{\rho}\mathbf{W}_n\mathbf{Y}_n - \mathbf{X}_n\hat{\beta} - \mathbf{W}_n\mathbf{X}_n\hat{\gamma}$. We repeat the estimation of step 3 and step 4. Further, we denote the current estimation of σ_0^2 with $\hat{\sigma}_{up}^2$ and the estimation from the previous step with $\hat{\sigma}_{up-1}^2$. Finally, we use as a stop criterion $(\hat{\sigma}_{up-1}^2 - \hat{\sigma}_{up}^2)/\hat{\sigma}_{up}^2 \leq 0.01$.

LeSage and Pace (2007) suggest to use direct/indirect or total effects for the model interpretation of a SDM model, which are given by the following Eqs. (4.7) to (4.9):

$$direct\ effect_k = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial y_i}{\partial x_{i,k}} \tag{4.7}$$

$$\textit{indirect effect}_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1, i\neq j}^{n}\frac{\partial y_i}{\partial x_{j,k}} \tag{4.8}$$

$$\textit{total effect}_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y_i}{\partial x_{j,k}} \tag{4.9}$$

where the partial derivative of y_i with respect to $x_{j,k}$, denoted by $\frac{\partial y_i}{\partial x_{j,k}}$ can be written as $\frac{\partial y_i}{\partial x_{j,k}} = \mathbf{S}^k_{i,j}$ where the matrix $\mathbf{S}^k$ is equal to $\mathbf{S}^k = (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1}(\mathbf{I}_n\beta_k + \mathbf{W}_n\gamma_k)$. Furthermore Eq. (4.9) can additionally be written in matrix notation as $total\ effect_k = \frac{1}{n}\boldsymbol{\iota}_n'\mathbf{S}^k\boldsymbol{\iota}_n$ and therefore, if $\mathbf{W}_n$ is row normalized, the corresponding total effect simplifies to $total\ effect_k = \frac{1}{1-\rho_0}(\beta_k + \gamma_k)$.

In order to derive the standard deviations for the direct/indirect or total effects LeSage and Pace (2007) suggest Monte Carlo Simulations. Note that if $\mathbf{W}_n$ is not row standardized, Eq.. (4.9) can not be further simplified and therefore each Monte Carlo simulation step needs the calculation of $(\mathbf{I}_n - \hat{\rho}\mathbf{W}_n)^{-1}$. Since almost all symmetric $\mathbf{W}_n$ matrices used in applied cases are not row standardized the calculation $(\mathbf{I}_n - \hat{\rho}\mathbf{W}_n)^{-1}$ is almost always necessary. In that light we regard Bootstrapping as a useful estimation method for calculating the standard deviations of $(\hat{\rho}, \hat{\beta}', \hat{\gamma}')'$ and the implied indirect/direct and total effects. Given we use bootstrapping we estimate that the computational time will approximately double or triple compared to the standard Monte Carlo approach suggested in LeSage and Pace (2007).

Alternatively Theorem 3.2 in the Appendix derives the standard deviation of $\widehat{\rho}_1$ as a first step for calculating the variance covariance matrix of $(\hat{\rho}, \hat{\beta}', \hat{\gamma}')'$. However our Monte Carlo experience suggests that this estimator has a very high variance itself and therefore we regard this first step analytical solution as an imprecise variance approximation.

A third possibility would be to use the variance covariance matrix of the Maximum Likelihood estimator as an approximation for the estimator given in Eq. (4.7) - Eq. (4.9). As the Monte Carlo simulations in the next section suggest that the proposed estimator is in small samples equally efficient as the ML counterpart.

4.4 Monte Carlo design and used performance measures

In our Monte Carlo study, the data generating process is given by (4.10)

$$\mathbf{Y}_n = \rho_o \mathbf{W}_n \mathbf{Y}_n + \mathbf{X}_n \beta_0 + \mathbf{W}_n \mathbf{X}_n \gamma_0 + \epsilon_n$$

$$\text{where } \epsilon_i \sim i.i.N(0, \sigma_0^2) \text{ and } \mathbf{X}_n \sim (\mathbf{I}_n - \lambda_0 \mathbf{W}_n)^{-1} \nu_n \qquad (10)$$

where $\mathbf{X}_n$ has one column. We follow Le Sage et al (2011) by introducing spatial autocorrelation in $\mathbf{X}_n$, where ν_n is drawn from an uniform distribution and is fixed for a given n while n is set to $100, 200$ and 400. We use these rather small sample sizes in order to reduce the computational burden. We set $\beta_0 = 1$, $\lambda_0 = 0$ and use two different settings for γ_0: once γ_0 is set equal β_0 and then equal to $-\rho_0\beta_0$. Note that in the second case the SDM model given in (4.10) simplifies to the so called spatial error model. Additionally we also consider $\beta_0 = 1$ and $\lambda_0 = 0.99$. Le Sage et. al. (2011) argue that spatial autocorrelation in the regressor introduces problems for the estimation ML estimation method and also has to be considered realistic for typically housing data regressors.

Each configuration of Eq. (4.10) is simulated $Trials_{MC} = 1000$ times. Since the error term in Eq. (4.10) is drawn from an independently distributed normal distribution we can compare the estimator performance of our new estimator with the Maximum likelihood estimator. We follow LeSage et. al. (2011) by fixing σ_0^2 such that the R^2, given in Eq. (4.11) as proxy for the information to noise ratio, is fixed across different values of ρ_0.

$$R^2 = \frac{Var\left[(\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1} \epsilon_n\right]}{Var\left[\mathbf{Y}_n\right]} \qquad (4.11)$$

The spatial autocorrelation parameter ρ_0 takes in our Monte Carlo experiment the following values: $-0.8, -0.4, 0, 0.4, 0.8$. We use as a spatial weight matrix a maximum eigenvalue normalized one forward one behind pattern. The major reason for this rather unrealistic neighborhood structure is its associated eigenvalue density. The traditional spatial filtering approach's (see Griffith (2003) or Pace et. al. (2011)) bias is smaller if most eigenvalues of $\mathbf{W}_n$ are near zero and as a result the

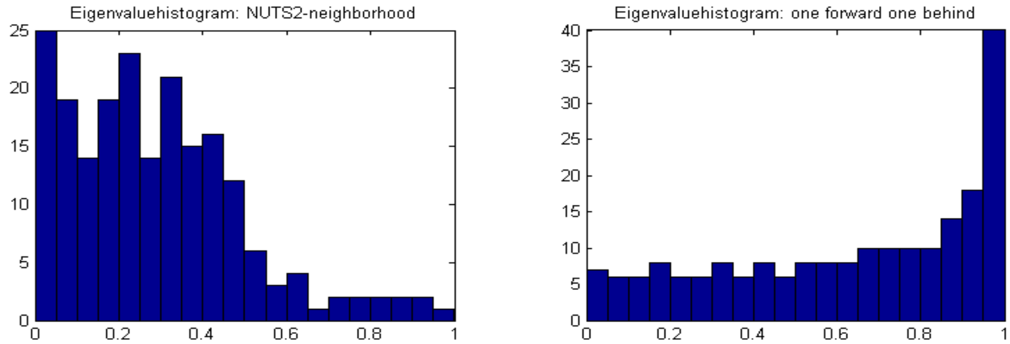

Figure 4.1: Eigenvalue-histogram of different spatial weight matrices

spatial filtering bias depends on the density of the $\mathbf{W}_n$'s spectrum. To compare the one forward one behind pattern we use an originally binary spatial weight matrix[37] reflecting the neighborhood among 203 European NUTS 2 regions, which was used by Fischer et. al. (2008). We compare the unrealistic one forward one behind weight matrix's spectrums density with the weight matrix's spectrum density used in Fischer et. al. in Figure (4.1).

Comparing these two histograms in Figure (4.1) we find out that the one forward one behind pattern is more problematic for spatial filtering methods than for example the pattern reflecting real life neighborhood structure. We observe as a characteristic of the eigenvalue density of the spatial weight matrix used by Fischer et. al. (2008) that indeed a large proportion of the weight matrix's eigenvalues are centered near zero. This characteristic is also observed by Pace et. al. (2011) for other binary neighborhood structures. Hence we are confident that using the one forward one behind pattern imposes an adverse environment for the new estimation procedure, which in return generalizes the Monte Carlo results.

In our Monte Carlo simulation we use approximately 95 per cent of the eigenvectors corresponding to the largest eigenvalues. Additionally we let estimation step 3 and 4 iterate until[38] $(\hat{\sigma}^2_{up-1} - \hat{\sigma}^2_{up})/\hat{\sigma}^2_{up} \leq 0.01$. The Monte Carlo experiment is programed in MATLAB and as com-

37 In Fischer et. al. (2008) "the weights matrix is constructed so that a neighboring region takes the value of 1 and 0 otherwise"

38 We find in our studies that on average 2 to 5 iteration steps are necessary.

parison we use the SDM- function based on maximum likelihood estimation, where the function is provided by the MATLAB spatial econometrics toolbox[39] programmed by James LeSage and Kelly Pace. In each experimental trial we will calculate the models total effects and its estimated standard deviation. The standard deviation is based on 100 bootstrap (Monte Carlo) trials for the new estimation procedure (maximum likelihood based function). Overall the experimental design uses for each sample size $3 \cdot 5 \cdot 3 = 45$ different parameter configurations. We report for each estimator and DGP configuration three different performance measures for the spatial auto correlation parameter and the models total effect.

The first performance measure is the so called relative estimator bias, which is given in Eq. (4.12)

$$Bias(\hat{\eta}) = \begin{cases} \frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \frac{\hat{\eta}_i - \eta_0}{\eta_0} 100 \text{ if } \eta_0 \neq 0 \\ \frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \hat{\eta}_i \text{ if } \eta_0 = 0 \end{cases} \tag{4.12}$$

where $\hat{\eta}$ is either the estimator for the spatial autocorrelation parameter or the estimator for the DGPs total effect. If $\hat{\eta}$ is indexed by i then $\hat{\eta}_i$ corresponds to parameter estimate in the i^{th} Monte Carlo trial. Note that the Bias is adjusted for the relative size of the true parameter value and can be interpreted percentage deviation if the true parameter value is not equal to zero.

The second performance measure is based on the mean squared error and is given by the following equation

$$RMSE(\hat{\eta}) = \begin{cases} \sqrt{\frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \left(\frac{\hat{\eta}_i - \eta_0}{\eta_0}\right)^2 100} \text{ if } \eta_0 \neq 0 \\ \sqrt{\frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \hat{\eta}_i^2} \text{ if } \eta_0 = 0 \end{cases} \tag{4.13}$$

Eq. (4.13) measures the average squared difference between the parameter estimated in the i^{th} Monte Carlo trial and the true parameter value. Given that the true parameter value is not equal to

[39] toolbox can be downloaded at www.spatialeconometrics.com . For details regarding the toolbox see LeSage and Pace (2007)

zero, the average squared difference is normalized by dividing it with the squared true parameter value, taking the square root of the resulting average and then multiplying it by 100 such that it can be interpreted as a percentage value. Note that if the average estimated parameter value over the MC trials is unbiased Eq. (4.13) can be written either as $\sqrt{VAR[\hat{\eta}]}\frac{100}{\eta_0}$ if $\theta_0 \neq 0$ or $\sqrt{VAR[\hat{\eta}]}100$ otherwise.

The third performance measure compares the estimated second moment of the estimator with its corresponding Monte Carlo second moment sample analogue. This performance measure given in Eq. (4.14) is similar to the measure given in Eq. (4.12) since both measure the bias of a parameter estimate where Eq. (4.12) concerns the estimators first moment and Eq. (4.14) the estimator's second moment estimate.

$$Bias^2(\hat{\eta}) = \frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \frac{\widehat{std[\hat{\eta}_i]} - std_0[\hat{\eta}_i]}{std_0[\hat{\eta}_i]} 100 \tag{4.14}$$

In Eq. (4.14) $\widehat{std[\hat{\eta}_i]}$ denotes for each Monte Carlo trial i the corresponding estimated standard deviation of $\hat{\eta}_i$. The true standard deviation of the estimator $\hat{\eta}_i$ is denoted by $std_0[\hat{\eta}_i]$. Note that the SDM function uses an approximation for $std_0[\hat{\eta}_i]$ which converges only asymptotically. Since there are no small sample analytical solutions available for $std_0[\hat{\eta}_i]$ we approximate this number by its Monte Carlo in- sample analogue. Hence Eq. (4.14) can be interpreted as the percentage bias of the estimated standard deviation of the corresponding parameter estimator.

4.5 Monte Carlo results

We compare the performance of our new estimator with the maximum likelihood estimator suggested by LeSage and Pace (2007). First we compare the estimator's performance by the estimation of the spatial autocorrelation parameter ρ_0 and second via the estimation performance for the total effect implied by the DGP. Hence the main Monte Carlo results are presented in form of 12 different tables. Each Table reports different performance measures for each estimator and weather $\gamma_0 = \beta_0$, $\gamma_0 = -\rho_0\beta_0$ or $\lambda_0 = 0.00$, $\lambda_0 = 0.99$. Tables (4.1)-(4.4) correspond to the total effects where $\gamma_0 = \beta_0$ and λ_0 is either 0.00 or 0.99 the tables (C.1)-(C.8) are given in the Appendix. Since the interpretation of Spatial Durbin models primarily depend on their implied effects, we are focusing our discussion of the Monte Carlo results on the implied total effects. In all tables the performance measures improve with increasing sample size and R^2.

Table 4.1: Monte Carlo Results: Maximum Likelihood for the total effect of x where γ_0=1 and λ_0=0.00

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-0.8	0.08%	4.54%	-1.68%	0.00%	1.81%	-2.98%	0.01%	0.54%	1.07%
	-0.4	-0.11%	7.56%	-0.90%	0.04%	3.37%	-0.38%	0.01%	1.43%	-1.96%
	0	0.29%	11.33%	-0.04%	0.07%	3.98%	4.03%	0.06%	1.96%	0.60%
	0.4	0.36%	12.89%	7.91%	-0.09%	4.51%	2.25%	0.10%	2.33%	-2.93%
	0.8	-0.82%	17.07%	6.93%	-0.28%	7.33%	2.21%	0.00%	2.05%	-0.33%
200	-0.8	-0.02%	2.45%	-6.56%	0.01%	0.99%	0.15%	0.01%	0.35%	0.55%
	-0.4	-0.01%	4.54%	-1.32%	-0.01%	1.51%	2.07%	-0.01%	0.76%	0.21%
	0	0.09%	5.15%	0.99%	-0.10%	2.26%	0.67%	-0.02%	0.96%	-0.47%
	0.4	0.11%	7.21%	2.57%	0.06%	2.80%	4.23%	0.00%	1.24%	-0.29%
	0.8	-0.24%	7.55%	0.18%	0.00%	3.27%	4.30%	0.01%	1.22%	-0.62%
400	-0.8	-0.07%	1.78%	0.02%	0.00%	0.71%	1.14%	0.01%	0.28%	-0.39%
	-0.4	0.00%	3.44%	0.22%	-0.10%	1.52%	1.53%	-0.04%	0.58%	2.47%
	0	0.00%	5.18%	1.54%	-0.01%	2.05%	0.40%	-0.03%	0.81%	-1.32%
	0.4	-0.11%	6.42%	2.04%	0.09%	2.74%	-2.54%	0.03%	1.04%	2.27%
	0.8	0.44%	7.29%	1.55%	-0.02%	3.51%	-0.61%	0.03%	1.32%	-1.28%

Comparing Table (4.1) and Table (4.2) we find that both estimators perform equally well regarding bias, RMSEs and Bias^2. Additionally the performance measures are similar for the MC-case $\gamma_0 = -\rho_0\beta_0$ and $\lambda_0 = 0.00$ given in the Appendix. Therefore we conclude for $\lambda_0 = 0.00$ that both estimators have a similar performance regarding the estimation of the first and second moment of the total effects.

If the explanatory variable exhibit a high spatial autocorrelation, then the estimation- performance for both estimators regarding the first moment of the total effects is similar well. However the asymptotic variance of the maximum likelihood estimator seems to be no longer an appropriate approximation for the small sample estimator variation. Especially in small sample sizes we observer in Table (4.4) deviations as large as 50 per cent. Since the standard deviation of our proposed estimator is based on bootstrapping we do not observe such weaknesses in Table (4.3). Therefore we conclude, given medium sample sizes where the estimation time is not an issue, that bootstrapping should be used for empirical applications. Overall we conclude that the Tables (C.1)-(C.8) indicate a similar performance of the proposed estimator compared to the maximum likelihood estimator.

Table 4.2: Monte Carlo Results: New estimator for the total effect of x where γ_0=1 and λ_0=0.00

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-0.8	0.08%	4.54%	-2.18%	0.00%	1.81%	-3.70%	0.01%	0.54%	0.06%
	-0.4	-0.11%	7.56%	-2.39%	0.04%	3.37%	-1.13%	0.00%	1.43%	-2.59%
100	0	0.29%	11.33%	-2.38%	0.07%	3.98%	1.17%	0.06%	1.96%	-1.68%
	0.4	0.37%	12.90%	3.85%	-0.08%	4.51%	-1.43%	0.11%	2.33%	-6.14%
	0.8	-0.81%	17.06%	-3.34%	-0.26%	7.33%	-6.09%	0.00%	2.05%	-4.97%
	-0.8	-0.02%	2.45%	-6.58%	0.00%	0.99%	-0.25%	0.01%	0.35%	-0.03%
	-0.4	-0.01%	4.54%	-2.47%	-0.01%	1.50%	1.02%	-0.02%	0.76%	-0.66%
200	0	0.09%	5.15%	-0.30%	-0.10%	2.26%	-0.48%	-0.02%	0.96%	-1.28%
	0.4	0.11%	7.21%	0.72%	0.07%	2.80%	2.85%	0.01%	1.24%	-1.87%
	0.8	-0.24%	7.55%	-4.00%	0.01%	3.27%	0.29%	0.01%	1.22%	-3.62%
	-0.8	-0.07%	1.78%	0.44%	0.00%	0.71%	0.61%	0.01%	0.28%	-0.62%
	-0.4	0.00%	3.44%	-0.35%	-0.11%	1.52%	1.36%	-0.04%	0.58%	1.82%
400	0	0.00%	5.18%	0.88%	-0.01%	2.05%	-0.99%	-0.03%	0.81%	-2.03%
	0.4	-0.11%	6.42%	1.21%	0.09%	2.74%	-2.69%	0.03%	1.04%	1.46%
	0.8	0.44%	7.29%	-0.77%	-0.01%	3.52%	-2.98%	0.02%	1.32%	-2.75%

Table 4.3: Monte Carlo Results: New estimator for the total effect of x where γ_0=1 and λ_0=0.99

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-0.8	-0.17%	3.56%	0.96%	-0.06%	1.91%	-3.38%	-0.01%	0.58%	-1.29%
	-0.4	-0.19%	9.88%	-1.13%	0.00%	3.70%	-0.55%	-0.04%	1.61%	1.90%
100	0	0.21%	17.85%	-3.26%	-0.06%	6.63%	3.10%	-0.01%	2.87%	-7.02%
	0.4	1.00%	21.10%	-1.34%	0.01%	8.80%	-2.38%	0.04%	3.72%	-1.14%
	0.8	0.34%	30.83%	0.63%	-0.03%	12.55%	-4.16%	0.13%	7.48%	-6.94%
	-0.8	0.10%	2.38%	0.73%	-0.08%	1.11%	2.24%	0.00%	0.48%	0.34%
	-0.4	-0.14%	6.27%	-1.52%	0.01%	2.61%	-2.91%	0.06%	1.19%	-0.59%
200	0	0.27%	10.97%	-3.46%	-0.05%	4.13%	-0.14%	0.05%	1.69%	0.44%
	0.4	0.08%	15.56%	-2.63%	-0.25%	6.55%	0.59%	0.04%	2.65%	0.29%
	0.8	0.28%	23.05%	-3.98%	-0.22%	11.42%	0.14%	-0.16%	3.64%	-2.59%
	-0.8	0.18%	2.04%	1.07%	-0.03%	0.83%	-0.14%	-0.01%	0.32%	0.79%
	-0.4	0.14%	5.63%	-2.03%	0.04%	2.06%	1.25%	-0.01%	0.93%	-3.79%
400	0	0.14%	8.08%	-2.82%	-0.07%	3.31%	-3.57%	0.04%	1.30%	3.28%
	0.4	0.25%	12.70%	-0.26%	0.12%	5.01%	0.35%	0.08%	1.94%	2.97%
	0.8	-0.03%	19.46%	-1.59%	0.14%	8.14%	0.13%	0.04%	2.87%	-2.80%

Like Pace et. al. (2011) we suggest to use the ARPACK (Lehoucq et. al. (1998)) public domain software package for calculating the eigenvalues and eigenvectors of the sparse matrix $\mathbf{W}_n$. In our view the usefulness of these algorithms lies in their ability to handle sparse matrices. If $\mathbf{W}_n$ had to be handled as a full matrix then, given a Windows operational system one could only handle matrices reflecting $n < 3000$ in MATLAB. Nevertheless the four step estimation method unlike

Table 4.4: Monte Carlo Results: Maximum Likelihood for the total effect of x where γ_0=1 and λ_0=0.99

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-0.8	-0.05%	3.86%	-56.77%	0.00%	1.99%	-57.45%	-0.02%	0.68%	-48.95%
	-0.4	-0.33%	9.63%	-48.30%	0.16%	3.82%	-50.44%	0.04%	1.43%	-48.83%
	0	0.17%	10.38%	-39.11%	-0.14%	5.80%	-41.08%	-0.10%	2.11%	-39.82%
	0.4	-0.68%	15.90%	-30.74%	0.04%	5.22%	-29.82%	0.09%	2.68%	-27.08%
	0.8	0.24%	17.25%	-9.12%	0.46%	7.85%	-8.18%	0.09%	3.19%	-9.85%
200	-0.8	-0.02%	1.59%	-34.53%	-0.04%	0.61%	-32.20%	0.00%	0.26%	-34.00%
	-0.4	0.20%	3.47%	-24.42%	-0.08%	1.29%	-25.14%	0.02%	0.61%	-26.00%
	0	-0.10%	5.15%	-21.42%	-0.06%	2.01%	-19.92%	-0.05%	0.84%	-19.13%
	0.4	0.23%	5.72%	-8.15%	-0.16%	2.80%	-9.40%	0.00%	1.09%	-14.07%
	0.8	0.20%	11.03%	-2.75%	0.15%	4.59%	-3.84%	-0.04%	1.77%	-2.72%
400	-0.8	0.01%	0.61%	-10.35%	0.00%	0.28%	-14.94%	0.00%	0.10%	-14.11%
	-0.4	-0.03%	1.47%	-8.33%	-0.02%	0.61%	-12.42%	0.00%	0.26%	-10.09%
	0	0.03%	2.42%	-9.40%	-0.03%	0.98%	-6.20%	0.01%	0.39%	-7.08%
	0.4	0.10%	3.16%	-2.98%	0.04%	1.30%	-3.28%	0.03%	0.59%	-5.01%
	0.8	0.01%	5.66%	-0.69%	-0.07%	2.09%	-1.06%	0.01%	0.88%	-0.22%

the filtering approach suggested by Pace et. al. (2011), raises some computational issues.

Pace et. al. (2011) report that for their sparse weight matrix the computational time for calculating the 100 largest eigenvalues increases with an order of $O(n^{1.1})$. Therefore they regard models with a sample size of one million as feasible. In contrast our results suggest that the new estimation method requires an increasing amount of eigenvalues. We find for example that our MC- results regarding the performance measures of the total effects do not change for the one forward one behind pattern if we set q to[40] $round(0.5(n-2k))$. Therefore we have to conclude that our method has an computational order of $O(n^{2.1})$ for calculating the necessary eigenvalues/eigenvectors. In other words, unlike classical spatial filtering methods where we observe the trade off between estimation bias and estimation efficiency, we find in our proposed method a trade-off between estimation bias and computational time. Still we are confident that the proposed estimator can tackle, if well programmed, sample sizes up to[41] $10,000$ for weight matrices with similar spectrum-properties like the ones reported in Pace et. al. (2011) or Fischer et. al. (2008). Nevertheless we have to conclude that if we want to have a proper model interpretation based on the proposed spatial filtering method, spatial filtering loses the Pace et. al. (2011) reported feasibility for very large data sets.

40 Given the weight matrix reflecting the NUTS-2 regions given Fischer et. al. (2008), we could set q to $round(0.05(n-2k))$ without changing the performance measures associated with the total effects.

41 $n = 1600$ requires with simple programming approximately 15 seconds

4.6 Concluding remarks

This paper outlined a new four step estimation method based on spatial filtering for spatial Durbin models. This estimator overcomes two inherent weaknesses of classical spatial filtering, which Pace et. al. (2011) label "a philosophical issue regarding the spatial filtering method". First it is possible to calculate thedirect/indirect and total effects implied by the estimated parameters and therefore allow for a proper model interpretation. Second, by the four step estimator's construction there exists no longer the inherent trade-off between estimation bias and estimation efficiency, which reduces the feasibility of spatial filtering method. We showed in our Monte Carlo experiments that the estimator can incorporate approximately 95 per cent of all eigenvectors without reducing the estimation efficiency. Therefore the estimation method is much more independent from the weight matrices' spectrum density. Additionally compared to classical spatial filtering, the four step estimator allows for a proper model interpretation and should be feasible for a much broader class of spatial weight matrices.

A Monte Carlo simulation was conducted in order to compare the new estimation method with the performance of the corresponding maximum likelihood estimator. We used a spatial one forward one behind pattern as the data generating process' weight matrix, since it creates an adverse environment for spatial filtering. Even under these conditions, we find that the proposed four step estimation method has similar estimation properties as the ML regarding estimation bias and efficiency. Since ML is, unlike the new four step estimator, based on the correctly specified model likelihood, the four step estimation method can be an especially useful estimation method for spatial Durbin models, where the dependent variable represent binary, discrete choice outcomes or Poisson distributed counts.

The paper discussed different possibilities for calculating the standard deviation of the implied direct/ indirect and total effects. Our Monte Carlo setting indicates that the simulation approach suggested by LeSage and Pace (2007) performs relatively poorly for estimating these standard deviations. Hence we are confident that our suggestion to use bootstrapping, which performs quiet well, is worth the additional computational time.

Our Monte Carlo results also suggest that the proposed four step estimator needs a fixed proportion of eigenvectors relative to the sample size in order to maintain its desirable estimation properties which contrasts the findings for spatial filtering of Pace et. al. (2011). Here the computational effort increases by $O(n^{1.1})$, since the 100 eigenvectors corresponding to the 100 largest absolute eigenvalues are deemed sufficient for (some) realistic weight matrices. Therefore Pace et. al. (2011) deem sample sizes where $n = 1,000,000$ as feasible for their spatial filtering approach, while our proposed four step estimator can only handle medium sized sample sizes, like $n = 10,000$. We have to conclude that there exists a trade-off for the four step estimator between the estimation bias and computational burden for large data sets.

4.7 Adapting the estimation method to the second paper's parameter space

If we change the assumption $\rho_o \in (-1, 1)$ to $\rho_o \in \Theta$, where Θ is the parameter space like it was given in the second paper of this thesis. Still we can write the following fundamental identity $(\mathbf{I}_n - \rho_o \mathbf{W}_n)^{-1} = \mathbf{I}_n + (\mathbf{I}_n - \rho_o \mathbf{W}_n)^{-1} \rho_o \mathbf{W}_n$. The main idea behind spatial filtering is to find a projector $\mathbf{M}_D$ such that $\mathbf{M}_D \mathbf{A}_n \approx \mathbf{0}$ where $\mathbf{A}_n = (\mathbf{I}_n - \rho_o \mathbf{W}_n)^{-1} \rho_o \mathbf{W}_n$. The projector $\mathbf{M}_D$ has the form $\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}_n \mathbf{D}_n'$ where $\mathbf{D}_n$ are the chosen eigenvectors necessary for the approximation. We can write $\mathbf{M}_D \mathbf{A}_n = (\mathbf{I}_n - \mathbf{D}_n \mathbf{D}_n') \bar{\mathbf{D}}_n \left(\mathbf{I}_n - \rho_o \bar{\mathbf{\Lambda}}_n\right)^{-1} \bar{\mathbf{D}}_n^{-1}$ $= (\mathbf{I}_n - \mathbf{D}_n \mathbf{D}_n') \bar{\mathbf{D}}_n \check{\mathbf{\Lambda}}_n(\rho_o) \bar{\mathbf{D}}_n^{-1}$ where the typical element of the diagonal matrix $\check{\mathbf{\Lambda}}_n(\rho_o)$ is given by $\frac{1}{1-\rho_o \lambda_i}$. Hence $\mathbf{M}_D \mathbf{A}_n \approx \mathbf{0}$ is fulfilled if the eigenvectors represented by the eigenvectormatrix $\mathbf{D}_n$ are associated with the corresponding largest absolute values of $\check{\mathbf{\Lambda}}_n(\rho_o)$. Therefore we no longer can just use the q eigenvectors corresponding to the q^{th} largest absolute eigenvalues of $\mathbf{W}_n$. The eigenvectors necessary for a good approximation are a function of the unknown parameter ρ_o.

Nevertheless the proposed estimation method can still be used, if we estimate the four step estimator three times, set q very high, like $q = round(0.95(n - 2k))$ and have relative to the sample size only a small number of explanatory variables: *First*, we estimate a model with q eigenvectors corresponding to the q largest absolute eigenvalues. *Second*, we estimate a model with q eigenvectors corresponding to the q smallest absolute eigenvalues. Since a relatively small number of eigenvectors are necessary to achieve reasonable model approximations, one of the estimation results form first or the second step should be reasonably close to the true parameter values. A simple decision rule for evaluating which of both estimations was closer to reality would be a pseudo likelihood based on normality of the error term. Hence we have one estimate $\hat{\rho}$. In the third step we use the q eigenvectors corresponding to the q largest absolute values of $\frac{1}{1-\hat{\rho}\lambda_i}$

Tables (4.5) and (4.6) show some Monte Carlo results where we have the same DGP given in Eq. (4.1) and the same setup as previously described. However we only report results for $n = 100$ in order to save computation time. Furthermore note one should no longer use average direct, indirect and total effects for the model interpretation as it was suggested in the second paper.

Table 4.5: Monte Carlo Results: New estimator for ρ_0 where γ_0=1 and λ_0=0.0

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-5	2.72%	-30.20%	-9.17%	4.87%	-40.43%	29.76%	7.55%	-38.00%	9.12%
	-3.5	-9.01%	-15.67%	-10.46%	-7.49%	-14.10%	-9.65%	0.34%	-4.39%	-15.90%
100	-2	-0.52%	-8.79%	-5.79%	0.15%	-10.96%	-17.90%	0.12%	-7.82%	-13.25%
	3.5	-7.41%	17.02%	-11.62%	-1.80%	17.27%	-22.06%	-1.56%	15.34%	-21.05%
	5	-2.40%	10.07%	11.55%	-0.54%	6.33%	-3.22%	-0.30%	5.15%	-3.72%

Table 4.6: Monte Carlo Results: New estimator for the total effect of x where γ_0=1 and λ_0=0.0

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-5	7.41%	10.17%	18.24%	7.76%	10.43%	17.93%	6.55%	10.83%	-1.64%
	-3.5	0.35%	5.01%	10.80%	0.08%	3.62%	16.84%	-0.03%	0.61%	-6.33%
100	-2	-0.32%	14.91%	1.55%	-0.23%	2.14%	2.80%	-0.26%	1.62%	6.03%
	3.5	-4.82%	-10.25%	10.27%	-2.49%	-4.82%	8.56%	-2.51%	-6.14%	-6.85%
	5	-0.17%	-8.56%	15.12%	-0.11%	-1.62%	20.48%	-0.06%	-1.02%	20.87%

We find that we have practically no difference in the estimation performance compared to the case where $\rho_o \in (-1, 1)$. However one might should increase the number of bootstrapping trials in order to decrease the Bias^2.

Chapter 5 Conclusions

Spatial econometric models, methods and techniques are of great importance in applied spatial econometrics. Since the prespecified parameter space associated with a spatial econometric model validates the corresponding estimation results, the associated parameter space is of great importance. So far the literature has not really dealt with the parameter space in an appropriate way. This thesis argues that the parameter space(s) used in spatial econometrics are inadequate and provides a new parameter space concept. Furthermore it discusses the possible changes in the model interpretation given the new parameter space concept and finally provides a new filtering based estimation method for SDM.

The thesis' Introduction argues from a practitioner's perspective that two currently popular econometric methods cannot find unbiased parameters outside the traditional parameter space boundaries: *First* LeSage et al (2010) find that the instrumental variable estimator suggested by Kelejian and Prucha (1998) might be biased in many applications, especially if the explanatory variables are generated from a spatially autocorrelated process and the associated model is the SDM. *Second* we show that numerical routines like the ones currently implemented by the LeSage MATLAB toolbox for likelihood based estimation methods, such as maximum likelihood or Bayesian spatial econometrics are not capable of dealing with spatial autocorrelation parameters where $|\rho| > 1$. Since the LeSage toolbox is a role model for the numerical implementation in spatial econometric toolboxes, this finding is very general. Hence we have to conclude that the absence of empirical evidence for $|\rho| > 1$ may be misleading and therefore spatial autocorrelation parameters outside traditional boundaries may exist in real world data.

If a real world spatial autoregressive data generating process where $|\rho| > 1$ would exist, an empirical verification would be of great importance: As the second paper shows spatial autoregressive models with a nontraditional parameter space require a different interpretation of their parameter estimates insomuch as direct, indirect and total effects are no longer useful summary measures. This different interpretation is a result of the different spillover "dynamic" associated with the SDM and nontraditional parameter spaces. As it for example turns out, direct effects may have a different sign across the observational unit although the means across the direct effects given the direct effects sign may be significantly different from zero. To be less abstract, let us assume that our observational unit represents economic growth in regions, we are interested in the effect of the variable "openness" on growth and the weight matrix represents the interregional spillovers due to trade. Furthermore we label those regions for which we observe a negative direct effect group 1 and the others group 2. Ignoring the above described "dynamics" could lead to a situation where due to their experience, policy makers from group 2 are arguing that their counterparts form group 1 should increase their regions's openness as well although such a step will actually decrease the economic growth. Therefore if such an empirical result would be found, this would change the perception of spatial autoregressive models. Motivated by such possible results, the third paper provides a filtering based estimation method for SDM with nontraditional parameter spaces in order to estimate the associated parameters.

The first paper discusses the Kelejian and Prucha's parameter space concept for the one forward one behind spatial weight matrix. It shows that asymptotically no islands of stability exist. It proves mathematically that the inverse eigenvalues of the one forward one behind spatial weight matrix are dense in $\mathbb{R}\backslash(-1,1)$ if the number of observations tends to infinity. The paper contributes to the literature that for such patterns the Kelejian Prucha parameter space concept indeed is adequate.

The second paper investigates the parameter space concepts suggested by Kelejian and Prucha (2010) and Lee and Liu (2010). It finds that the Kelejian and Prucha (2010) parameter space can result in nonstationary DGPs, while the parameter space proposed by Lee and Liu (2010) can be too restrictive in applied cases. Furthermore it discusses that the practice of row standardizing lacks a mathematical foundation. The paper provides a new parameter space concept and supplies several applications.

This paper contributes to the literature in the following ways: *First* it shows that the Kelejian and Prucha (2010) parameter space concept can result in nonstationary data generating processes. It applies the concept to a weight matrix where the eigenvalue spectrum is zero and found that the corresponding parameter space would have been $\mathbb{R}$. However, since the data generating process with this weight matrix was actually a first order autoregressive process from time series it can be concluded that the actual parameter space is $(-1, 1)$. Furthermore the parameter space concept is applied to a specific non triangular weight matrix with full spectrum and the maximum absolute element of $(\mathbf{I}_n - 0.95\mathbf{W}_n)^{-1}$ is calculated which tends to increase exponentially over n and if for example $n = 150$ the maximum absolute entry is $\sim 10^{25}$. This is a further hint that the Kelejian Prucha parameter space concept might result in non stationary data generating processes.

Second the paper shows that the parameter space concept proposed by Lee and Liu (2010) can be too restrictive in applied cases. If the Lee and Liu parameter space is applied to a one forward one behind spatial weight matrix the resulting parameter space is only $(-2/3, 2/3)$ instead of $(-1, 1)$.

The *third* contribution to the literature is the following observation: The practitioner's approach of row standardizing has no longer a mathematical foundation, since both parameter space concepts fail to ensure the boundedness of $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$.

Fourth the paper proposes a new parameter space concept. This concept is theoretically in line with the underlying assumptions of the spatial econometric estimation theory. Furthermore it shows the usefulness of this definition by proving that row standardizing can still result in reasonable parameter spaces, given some additional properties of the weight matrix. Two additional applications for the new parameter space definition concerning models with group interaction and panels with fixed cross sectional size are provided. Both applications result in parameter spaces that are substantially larger than the ones the literature has so far considered to be stable.

Fifth the paper discusses implications regarding the model interpretation for parameter spaces outside of $(-1, 1)$ for the so-called Spatial Durbin Model. One consequence is that increasing the explanatory variable would - independent of the associated regression coefficient's sign - for some observations increase the dependent variable while decreasing it for others, if spillover effects or so-called indirect effects are neglected.

The third paper proposes a new estimation method based on Spatial filtering for the Spatial Durbin model. Pace et al. (2011) show that the spatial filtering methods developed by Griffith (2000) have desirable estimation properties for some parameters associated with spatial autoregressive models. However the results associated with spatial filtering lack a proper interpretation. The new estimation method advocated by the third paper has on the one hand desirable estimation properties and the results have a proper interpretation.

The third paper contributes to the literature in the following ways: *First* it provides a new estimation method which has similar estimation properties as the maximum likelihood estimator and allows for a proper model interpretation via direct indirect and total effects. Pace et. al. (2011) argue maximum likelihood based estimators are not useful for models where the dependent variables represent binary, discrete choice outcomes or Poisson distributed counts and where we observe an autoregressive structure in the dependent variable. Therefore these models provide a "natural" application for this new estimation method.

Second the estimation method can be relatively simple adopted for parameter spaces that follow the definition of the second paper. Since maximum likelihood faces numerical difficulties, Spatial Durbin models with a parameter spaces outside $(-1, 1)$ provide the second "natural" application for this new estimation method.

The dissertation enriches the field of spatial econometrics in two ways: *First* it shows that parameter space concepts are inadequate. It proves that only in some cases the traditional parameter space concepts may result in stable data generating processes. It provides an alternative concept which is for example able to prove that the practice of row standardization will in most cases result in stable data generating processes. Furthermore it gives possible applications where a spatial autocorrelation parameter outside of $(-1, 1)$ might be found. Moreover the implications of such a spatial autocorrelation parameter for the model interpretation is discussed. The thesis finds the possibility that, increasing the explanatory variable would - independent of the associated regression coefficient's sign - for some observations increase the dependent variable while decreasing it for others, if spill over effects or so called indirect effects are neglected. This is a completely new view on the "dynamics" associated with spatial autoregessive models. *Second* the dissertation discussed reasons why so far the literature has not found a spatial autocorrelation parameter outside of $(-1, 1)$ and provides an estimation methodology that can cope with such parameter spaces and furthermore is independent from the error term distribution. Moreover the thesis shows that for traditional parameter spaces the estimation methodology has similar estimation properties as the maximum likelihood estimator regarding bias and efficiency.

References

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

Barry, R. and K. Pace (1999). Monte carlo estimates of the log determinant of large sparse matrices. *Linear Algebra and its Applications 289*(1-3), 41–54.

Casella, G. and R. Berger (2001). *Statistical Inference* (second ed.). Belmont: Duxbury Press.

Cuaresma, J. C., G. Doppelhofer, and M. Feldkircher (Jannuary 2009). The determinants of economic growth in european regions. *CESifo Working Paper Series No. 2519.*.

Cuaresma, J. C. and M. Feldkircher. Spatial filtering, model uncertainty and the speed of income convergence in europe. *Journal of Applied Econometrics, forthcoming*.

Elhorst, P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis 5*(1), 9–28.

Fischer, M. and D. Griffith (2008). Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the european union. *Journal of Regional Science 48*(5), 969–989.

Fischer, M., T. Scherngell, and M. Reismann (2009). Knowledge spillovers and total factor productivity: Evidence using a spatial panel data model. *Geographical Analysis 41*(2), 204–220.

Getis, A. and D. Griffith (2002). Comparative spatial filtering in regression analysis. *Geographical*

Analysis 34(2), 130–140.

Greene, W. (1997). *Econometric Analysis* (third ed.). Upper Saddle River, New Jersey: Prentice Hall.

Griffith, D. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems 2*(2), 141–156.

Griffith, D. (2002). A spatial filtering specification for the auto-poisson model. *Statistics and Probability Letters 58*(3), 245–251.

Griffith, D. (2003). *Spatial Autocorrelation and Spatial Filtering*. Berlin, Heidelberg, New York: Springer.

Hamilton, J. (1994). *Time Series Analysis*. Princton (N. J.): Princton University Press Princton.

Johnsen, C. and R. Horn (1985). *Matrix analysis*. Cambridge (UK): Cambridge University.

Kelejian, H. and I. 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*(1), 99–121.

Kelejian, H. and I. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics 157*(1), 53 – 67.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*(6), 1899 Ű 1925.

Lee, L. F. and X. Liu (2010). Efficient gmm estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory 26*(1), 187–230.

Lee, L. F. and J. Yu (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*(2), 165 – 185.

Lehoucq, R. B., D. C. Sorensen, and C. Yang (1998). *ARPACK UsersŠ Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. Philadelphia: SIAM.

LeSage, J. and K. Pace (2009). *An introduction to spatial econometrics*. Boca Raton, London, New York: Tylor Francis Group.

LeSage, J., K. Pace, and Z. Shuang (2011). Spatial dependence in regressors and its effect on estimator performance. *Working Paper*.

Martin, J., R. (1993). Approximations to the determinant term in gaussian maximum likelihood estimation of some spatial models. *Communications in Statistics-Theory and Methods 22*(1), 189–205.

Neumann, C. (1877). *Untersuchungen ueber das logarithmische und Newton'sche Potential*. Leipzig: B. G. Teubner.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association 70*(349), 120–126.

Pace, K. and J. LeSage (1998). Quick computations of spatially autoreressive estimators. *Geographical Analysis 29*(3), 232–247.

Pace, K. and J. LeSage (2004). Chebyshev approximation of log-determinants of spatial weight matrices. *Computational Statistics Data Analysis 45*(2), 179–196.

Pace, K., J. LeSage, and S. Zhu (2011). Interpretation and computation of estimates from regression models using spatial filtering. *Working Paper*.

Tiefelsdorf, M. and D. Griffith (2007). Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environment and Planning A 39*(5), 1193–1221.

A. Appendix to Paper I

LEMMA (A.1.1) *The z_j of theorem 2.1 obey the following equation:*

$$z_j = \frac{\left(\frac{-1-w}{\rho^2}\right)^j (1-2\rho^2+w) - \left(\frac{-1+w}{\rho^2}\right)^j (1-2\rho^2-w)}{\left(\frac{-1-w}{\rho^2}\right)^j (1+w) - \left(\frac{-1+w}{\rho^2}\right)^j (1-w)} \mid w = \sqrt{1-4\rho^2}$$

Proof. We will verify that formula is indeed the right solution:

$z_{i+1} = 1 - \frac{\rho^2}{z_i}$, $z_1 = 1 - \rho^2$ can be written as

$$z_j = \frac{\left(\frac{-1-w}{\rho^2}\right)^j (1-2\rho^2+w) - \left(\frac{-1+w}{\rho^2}\right)^j (1-2\rho^2-w)}{\left(\frac{-1-w}{\rho^2}\right)^j (1+w) - \left(\frac{-1+w}{\rho^2}\right)^j (1-w)} |w = \sqrt{1-4\rho^2}.$$

If $\theta = -1 - w$ and $\vartheta = -1 + w$, then z_j can be written as $z_j = \frac{\vartheta^{j+1} - \theta^{j+1} - 2\rho^2\left(\theta^j - \vartheta^j\right)}{\vartheta^{j+1} - \theta^{j+1}}$ some useful calculation rules: $\theta\vartheta = 4\rho^2$, $2 + \theta = -\vartheta$, $2 + \vartheta = -\theta$ Both formulas have to give the same result for the starting value: $j = 1 : z_1 = 1 - \rho^2$ (iterative formula)

Per definition it follows that: $z_{i+1} = 1 - \frac{\rho^2}{z_i}$. Now we have to proof if $\widetilde{z}_j \stackrel{?}{=} z_i$ is true: $\widetilde{z}_{j+1} = \frac{\vartheta^{j+2} - \theta^{j+2} - 2\rho^2\left(\theta^{j+1} - \vartheta^{j+1}\right)}{\vartheta^{j+2} - \theta^{j+2}} = 1 - \rho^2 2 \frac{\theta^{j+1} - \vartheta^{j+1}}{\vartheta^{j+2} - \theta^{j+2}} = 1 - \frac{\rho^2}{\frac{\vartheta^{j+2} - \theta^{j+2}}{-2\left(\vartheta^{j+1} - \theta^{j+1}\right)}} \stackrel{!}{=} 1 - \frac{\rho^2}{z_j}$

$\widetilde{z}_j = \frac{\vartheta^{j+1} - \theta^{j+1} - 2\rho^2\left(\theta^j - \vartheta^j\right)}{\vartheta^{j+1} - \theta^{j+1}}$ One can see that $\widetilde{z}_j = z_i$ is only valid if $i = j$ and the following is true: $\widetilde{z}_j = \frac{\vartheta^{j+1} - \theta^{j+1} - 2\rho^2\left(\theta^j - \vartheta^j\right)}{\vartheta^{j+1} - \theta^{j+1}} \stackrel{!}{=} z_i = z_j = \frac{\vartheta^{j+2} - \theta^{j+2}}{-2\left(\vartheta^{j+1} - \theta^{j+1}\right)}$

$$-2\vartheta^{j+1} + 2\theta^{j+1} + 4\rho^2\left(\theta^j - \vartheta^j\right) \stackrel{!}{=} \vartheta^{j+2} - \theta^{j+2}$$

$$-2\vartheta^{j+1} + 2\theta^{j+1} + \theta\vartheta\left(\theta^j - \vartheta^j\right) \stackrel{!}{=} \vartheta^{j+2} - \theta^{j+2}$$

$$-\left(\vartheta^{j+1}\right)(-\vartheta) + \theta^{j+1}(-\theta) \stackrel{!}{=} \vartheta^{j+2} - \theta^{j+2}$$

$$\vartheta^{j+2} - \theta^{j+2} \stackrel{!}{=} \vartheta^{j+2} - \theta^{j+2}$$

Therefore $\widetilde{z}_j = z_i$ if $i = j$. ■

LEMMA (A.1.2) *If* $\varphi_\alpha = \pi + \arcsin\left(\frac{\sqrt{4\rho^2 - 1}}{|2\rho|}\right)$,$\varphi_\beta = \frac{\pi}{2} - \arcsin\left(\frac{1-2\rho^2}{2\rho^2}\right)$, $\varphi_\alpha \in \left(\pi, \frac{\pi}{2}\right)$ *and* $\varphi_\beta \in (0, \pi)$*, then:* $\varphi_\beta = 2\varphi_\alpha - 2\pi$.

Proof. $\varphi_\alpha = \pi + \arcsin\left(\frac{\sqrt{4\rho^2-1}}{|2\rho|}\right) \Rightarrow \sin\left(\varphi_\alpha - \pi\right)^2 = \frac{4\rho^2-1}{4\rho^2} \Rightarrow \rho = \sqrt{\frac{1}{4\left(1-\sin(\varphi_\alpha-\pi)^2\right)}}$

$\varphi_\beta = \frac{\pi}{2} - \arcsin\left(\frac{1-2\rho^2}{2\rho^2}\right) \Leftrightarrow \sin\left(\frac{\pi}{2} - \varphi_\beta\right) = \frac{1-2\rho^2}{2\rho^2} \Rightarrow \rho' = \sqrt{\frac{1}{2\left(1+\sin\left(\frac{\pi}{2}-\varphi_\beta\right)\right)}}$

$\rho \stackrel{!}{=} \rho' : \frac{1}{4\left(1-\sin(\varphi_\alpha-\pi)^2\right)} = \frac{1}{2\left(1+\sin\left(\frac{\pi}{2}-\varphi_\beta\right)\right)} \Leftrightarrow 2\left(1 + \sin\left(\frac{\pi}{2} - \varphi_\beta\right)\right) =$

$4\left(1 - \sin\left(\varphi_\alpha - \pi\right)^2\right) \Leftrightarrow \sin\left(\frac{\pi}{2} - \varphi_\beta\right) = 1 - 2\sin\left(\varphi_\alpha - \pi\right)^2 \Leftrightarrow \cos\left(\varphi_\beta\right) =$

$1 - 2\sin\left(\varphi_\alpha - \pi\right)^2 \Leftrightarrow \cos\left(\varphi_\beta\right) = 1 - 2\frac{1}{2}\left(1 - \cos\left(2\varphi_\alpha - 2\pi\right)\right) \Leftrightarrow \cos\left(\varphi_\beta\right) =$

$\cos\left(2\varphi_\alpha - 2\pi\right)$ since $\varphi_\alpha \in \left(\pi, \frac{\pi}{2}\right)$ and $\varphi_\beta \in (0, \pi)$

$\varphi_\beta = 2\varphi_\alpha - 2\pi$ ■

LEMMA (A.1.3) *Due to theorems 2.1 and the assumptions in theorem 2.2 it follows that* $y_1 = \prod_{j=1}^{n-1} \frac{-s_n}{\cos(\varphi_\alpha)} + \frac{-\tan((j+1)\varphi_\alpha)s_n}{\sin(\varphi_\alpha)} + \sum_{i=1}^{n-1}\prod_{j=n-i}^{n-1} \frac{-s_i}{\cos(\varphi_\alpha)} + \frac{-\tan((j+1)\varphi_\alpha)s_i}{\sin(\varphi_\alpha)}$*. We now take a closer look on the argument in the tangent function for* $\lim_{n\to\infty} y_1$*. One can see that the index j takes every value of the natural numbers at least once. The function* $\tan(x)$ *is not defined for values* $x = \left(m + \frac{1}{2}\right)\pi$ *with* $m \in \mathbb{Z}$*. Therefore, the product sum* $\prod_{j=1}^{n-1} \frac{1}{\cos(\varphi_\alpha)} + \frac{\tan((j+1)\varphi_\alpha)}{\sin(\varphi_\alpha)}$ *in* y_1 *is not defined if* $(j+1)\,\varphi_\alpha = \left(m + \frac{1}{2}\right)\pi$ *with* $m \in \mathbb{Z}$*,* $j \in \mathbb{N}$.

Proof. trivial ■

B. Appendix to Paper II

Lemma B.1: *If* $\varrho \in \{1, \infty\}$ *and* $\|\mathbf{\Psi}_{n,p}\|_\varrho < 1$ *where* $\mathbf{\Psi}_{n,p} = \sum_{j=1}^{p} \rho_j \mathbf{W}_{jn}$ *then* $\left\|(\mathbf{I}_n - \mathbf{\Psi}_{n,p})^{-1}\right\|_\varrho \leq \frac{1}{1-\|\mathbf{\Psi}_{n,p}\|_\varrho} < \infty$

Proof. Since $\|\mathbf{\Psi}_{n,p}\|_\varrho < 1$ the Neumann series can be applied in order to write $(\mathbf{I}_n - \mathbf{\Psi}_{n,p})^{-1} = \sum_{k=0}^{\infty} \mathbf{\Psi}_{n,p}^k$. Therefore, $\left\|(\mathbf{I}_n - \mathbf{\Psi}_{n,p})^{-1}\right\|_\varrho = \left\|\sum_{k=0}^{\infty} \mathbf{\Psi}_{n,p}^k\right\|_\varrho \leq \sum_{k=0}^{\infty} \left\|\mathbf{\Psi}_{n,p}^k\right\|_\varrho \leq \sum_{k=0}^{\infty} \|\mathbf{\Psi}_{n,p}\|_\varrho^k = \frac{1}{1-\|\mathbf{\Psi}_{n,p}\|_\varrho} < \infty$. The inequality follows due to the triangle inequality (see Horn and Johnson (1985)) and the second due to the sub- multiplicativity of these matrix norms. ■

Neumann series[42]**:** *If* $\|\mathbf{\Psi}_n\| < 1$ *for any matrix norm it follows*[43]*:* $(\mathbf{I}_n - \mathbf{\Psi}_{n,p})^{-1} = \sum_{k=0}^{\infty} \mathbf{\Psi}_n^k$

Proof. First: Let $\|\mathbf{\Psi}_n\| < 1$ then: $\lim_{k\to\infty} \mathbf{\Psi}_n^k \leq \boldsymbol{\iota}_n \boldsymbol{\iota}_n' \lim_{k\to\infty} \left\|\mathbf{\Psi}_n^k\right\| \leq \boldsymbol{\iota}_n \boldsymbol{\iota}_n' \lim_{k\to\infty} \|\mathbf{\Psi}_n\|^k = \mathbf{0}_{n,n}$ (for the first inequality the sub- multiplicativity of these matrix norms is used). Therefore, $\lim_{k\to\infty} \mathbf{\Psi}_n^k = \mathbf{0}_{n,n}$. Second, it has to be shown $(\mathbf{I}_n - \mathbf{\Psi}_{n,p})^{-1} = \sum_{k=0}^{\infty} \mathbf{\Psi}_n^k$. This is equivalent to: $(\mathbf{I}_n - \mathbf{\Psi}_{n,p}) \sum_{k=0}^{\infty} \mathbf{\Psi}_n^k = \mathbf{I}_n$. Since $\lim_{K\to\infty} (\mathbf{I}_n - \mathbf{\Psi}_{n,p}) \sum_{k=0}^{K} \mathbf{\Psi}_n^k = \lim_{K\to\infty} \mathbf{I}_n - \mathbf{\Psi}_{n,p}^{K+1} = \mathbf{I}_n$ ■

Proof. for $Sp(\overline{\mathbf{W}}_n) = \{0\}$ **if the typical element** $\overline{w}_{ij,n}$ **of** $\overline{\mathbf{W}}_n$ **is defined by (3.2):** The proof rewrites $\overline{\mathbf{W}}_n$: $\overline{\mathbf{W}}_n = F_n \mathbf{\Gamma}_n F_n^{-1}$. If the typical element $f_{i,j}$ of F_n and $\gamma_{i,j}$ of $\mathbf{\Gamma}_n$ are defined by (B.1) and (B.2) it follows due to the Jordan normal form that the diagonal elements of $\mathbf{\Gamma}_n$ are the eigenvalues of $\overline{\mathbf{W}}_n$.

$$f_{i,j} = \begin{cases} 1 \text{ if } i = n+1-j \text{ and } j \in \{1, 2, ..., n\} \\ 0 \text{ otherwise} \end{cases} \tag{B.1}$$

$$\gamma_{i,j} = \begin{cases} 1 \text{ if } j = i+1 \text{ and } i \in \{1, 2, ..., n-1\} \\ 0 \text{ otherwise} \end{cases} \tag{B.2}$$

---

42 The Neumann series was partly developed by Carl Neumann (1832-1925) who used it in the context of potential theory (1877). It is a useful "tool" in functional analysis. It has to pointed out, since some authors use the for the "Neumann series" the incorrect term "Taylor series".

43 The series is even more general: Let Ξ be Banach algebra and $x \in \Xi$ where $\|x\| < 1$. Then the series $\sum_{k=0}^{\infty} x^k$ is absolutely convergent, and $\sum_{k=0}^{\infty} x^k = (e-x)^{-1}$ where e denotes the one-element in Ξ. For more details see for example Heuser (1995)

Note that $\digamma_n = \digamma_n^{-1}$ and $\mathbf{\Gamma}_n$ is a typical Jordan form. It can easily be seen that $\overline{\mathbf{W}}_n = \digamma_n \mathbf{\Gamma}_n \digamma_n^{-1} = \digamma_n \mathbf{\Gamma}_n \digamma_n$ holds and hence $Sp(\overline{\mathbf{W}}_n) = \{0\}$. ■

Proof. for the group model -Part A: For a model like (3.5) the following parameter space fulfills all the 3 parameter space properties: $|\rho_{11}| < 1$, $|\rho_{22}| < 1$ and $\frac{|\rho_{12}|}{|1-\rho_{22}|}\frac{|\rho_{21}|}{|1-\rho_{11}|} < 1$. Let $\mathbf{\Psi}_{n,4} = \rho_{11}\hat{\mathbf{W}}_{11} + \rho_{12}\hat{\mathbf{W}}_{12} + \rho_{21}\hat{\mathbf{W}}_{21} + \rho_{22}\hat{\mathbf{W}}_{22}$. Due to $|\rho_{11}| < 1$ $|\rho_{22}| < 1$ and $\max\{||\hat{\mathbf{W}}_{i,j}||_\varrho$ where $\varrho \in \{1,\infty\}$ and $i,j \in \{1,2\}\} \leq 1$ the inverse of $\mathbf{I}_{n1} - \rho_{11}\mathbf{W}_{n1,n1}$ and $\mathbf{I}_{n2} - \rho_{22}\mathbf{W}_{n2,n2}$ exist. An equation system is used so $\mathbf{y}_1 = \rho_{11}\mathbf{W}_{n1,n1}\mathbf{y}_1 + \rho_{12}\mathbf{W}_{n1,n2}\mathbf{y}_2 + \mathbf{s}_1$ and $\mathbf{y}_2 = \rho_{21}\mathbf{W}_{n2,n1}\mathbf{y}_1 + \rho_{22}\mathbf{W}_{n2,n2}\mathbf{y}_2 + \mathbf{s}_2$. These equations can be solved:

$$\mathbf{y}_1 = \left(\mathbf{I}_{n1} - \rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{W}_{n2,n1}\right)^{-1}\mathbf{A}_1\left(\rho_{12}\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{s}_2 + \mathbf{s}_1\right)$$

$$\mathbf{y}_2 = \left(\mathbf{I}_{n2} - \rho_{12}\rho_{21}\mathbf{A}_2\mathbf{W}_{n2,n1}\mathbf{A}_1\mathbf{W}_{n1,n2}\right)^{-1}\mathbf{A}_2\left(\rho_{21}\mathbf{W}_{n2,n1}\mathbf{A}_1\mathbf{s}_1 + \mathbf{s}_2\right)$$

where $\mathbf{A}_1 = \left(\mathbf{I}_{n1} - \rho_{11}\mathbf{W}_{n1,n1}\right)^{-1}$ and $\mathbf{A}_2 = \left(\mathbf{I}_{n2} - \rho_{22}\mathbf{W}_{n2,n2}\right)^{-1}$.

Therefore, $\mathbf{\Psi}_{n,4}^{-1} = \begin{pmatrix} \widetilde{\Psi}_{11} & \widetilde{\Psi}_{12} \\ \widetilde{\Psi}_{21} & \widetilde{\Psi}_{22} \end{pmatrix}$ where

$\widetilde{\Psi}_{11} = \left(\mathbf{I}_{n1} - \rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{W}_{n2,n1}\right)^{-1}\mathbf{A}_1$,

$\widetilde{\Psi}_{12} = \left(\mathbf{I}_{n1} - \rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{W}_{n2,n1}\right)^{-1}\mathbf{A}_1\rho_{12}\mathbf{W}_{n1,n2}\mathbf{A}_2$,

$\widetilde{\Psi}_{21} = \left(\mathbf{I}_{n2} - \rho_{12}\rho_{21}\mathbf{A}_2\mathbf{W}_{n2,n1}\mathbf{A}_1\mathbf{W}_{n1,n2}\right)^{-1}\mathbf{A}_2\rho_{21}\mathbf{W}_{n2,n1}\mathbf{A}_1$ and

$\widetilde{\Psi}_{22} = \left(\mathbf{I}_{n2} - \rho_{12}\rho_{21}\mathbf{A}_2\mathbf{W}_{n2,n1}\mathbf{A}_1\mathbf{W}_{n1,n2}\right)^{-1}\mathbf{A}_2$. If a Neumann series is applied it can be shown that $\mathbf{\Psi}_{n,4}^{-1}$ exists if $\|\rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n2,n1}\mathbf{A}_2\mathbf{W}_{n2,n1}\|_1 \vee ||\rho_{12}\rho_{21}\ \mathbf{A}_1\mathbf{W}_{n2,n1}\ \mathbf{A}_2\mathbf{W}_{n2,n1}\ ||_\infty < 1$. $\|\rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n2,n1}\mathbf{A}_2\mathbf{W}_{n2,n1}\|_1 \vee \|\rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{W}_{n2,n1}\|_\infty < 1$ is true if $\frac{|\rho_{12}|}{|1-\rho_{22}|}\ \frac{|\rho_{21}|}{|1-\rho_{11}|} < 1$, since Lemma B.1 shows $\|\mathbf{A}_1\|_{1,\infty} < \frac{1}{|1-\rho_{11}|}$and $\|\mathbf{A}_2\|_{1,\infty} < \frac{1}{|1-\rho_{22}|}$. Therefore, the parameter space property 1 is fulfilled if $|\rho_{11}| < 1$, $|\rho_{22}| < 1$ and $\frac{|\rho_{12}|}{|1-\rho_{22}|}\ \frac{|\rho_{21}|}{|1-\rho_{11}|} < 1$.

Since $\|\rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{W}_{n2,n1}\|_1 \wedge \|\rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{W}_{n2,n1}\|_\infty < 1$ hold under the proposed parameter space it follows due to Lemma 1, the boundedness of $\mathbf{A}_1$ and $\mathbf{A}_2$ and $\max\{$ $||\mathbf{W}_{ni,nj}||_\varrho$ where $\varrho \in \{1,\infty\}$ and $i,j \in \{1,2\}\} \leq 1$ that parameter space property 2 is also fulfilled.

The parameter space $|\rho_{11}| < 1$, $|\rho_{22}| < 1$ and $\frac{|\rho_{12}|}{|1-\rho_{22}|}\frac{|\rho_{21}|}{|1-\rho_{11}|} < 1$ obviously fulfills the parameter space property 3.

Part B: The following assumptions are being made: $\mathbf{W}_{n1,n1}$, $\mathbf{W}_{n2,n2}$, $\mathbf{W}_{n2,n1}$ and $\mathbf{W}_{n1,n2}$ are row standardized, $\mathbf{W}_{n1,n1}$ and $\mathbf{W}_{n2,n2}$ can be written as $\mathbf{W}_{n1,n1} = \Lambda_{n1,n1}\ \overline{\mathbf{W}}_{n1,n1}$,$\mathbf{W}_{n1,n2} = \Lambda_{n1,n2}\ \overline{\mathbf{W}}_{n1,n2}$,$\mathbf{W}_{n2,n1} = \Lambda_{n2,n1}\ \overline{\mathbf{W}}_{n2,n1}$ and $\mathbf{W}_{n2,n2} = \Lambda_{n2,n2}\ \overline{\mathbf{W}}_{n2,n2}$ where Λ represents the row- standardizing and both $\overline{\mathbf{W}}_{n1,n1}$ and $\overline{\mathbf{W}}_{n2,n2}$ are symmetric, $\|\overline{\mathbf{W}}\|_{\infty}$, $\|\mathbf{W}_{n1,n2}\|_{\infty}$ $\mathbf{W}'_{n2,n1}$ $= \mathbf{W}_{n1,n2}$ and $\|\mathbf{W}_{n2,n1}\|_{\infty} < \infty$

It has to be shown that $\|\mathbf{I}_{n1} - \rho_{12}\rho_{21}\mathbf{A}_1\mathbf{W}_{n1,n2}\mathbf{A}_2\mathbf{W}_{n2,n1}\|_{\infty} < \infty$: Under the assumptions, it follows (see Theorem 1): $\|\mathbf{A}_1\|_{\infty} = \frac{1}{1-|\rho_{11}|}$ and $\|\mathbf{A}_2\|_{\infty} = \frac{1}{1-|\rho_{22}|}$. Since $||\rho_{12}\ \rho_{21}\ \mathbf{A}_1\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}||_{\infty} \leq \frac{|\rho_{12}\rho_{21}|}{|1-\rho_{22}||1-\rho_{11}|}$ and due to Lemma 1 it follows directly :

$\|\mathbf{I}_{n1} - \rho_{12}\rho_{21}\ \mathbf{A}_1\mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}\|_{\infty} < \infty$.

Additionally the same has to be shown for $\|(\mathbf{I}_{n1} - \rho_{12}\rho_{21}\mathbf{A}_1\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1})^{-1}\|_1$: Since $\|(\mathbf{I}_{n1} - \rho_{12}\rho_{21}\mathbf{A}_1\ \mathbf{W}_{n1,n2}\mathbf{A}_2\ \mathbf{W}_{n2,n1})^{-1}\|_1 \leq \sum_{k=0}^{\infty} |\rho_{12}\ \rho_{21}|^k\ \|(\ \mathbf{A}_1\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}\ )^k\|_1 = \|\mathbf{I}_{n1} + \rho_{12}\rho_{21}\ \mathbf{A}_1\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1} + \sum_{k=2}^{\infty} |\rho_{12}\rho_{21}|^k\ \ \mathbf{A}_1\ (\prod_{j=1}^{k-1}\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}\ \mathbf{A}_1)\ \mathbf{W}_{n1,n2}\mathbf{A}_2\ \mathbf{W}_{n2,n1}\|_1 \leq 1 + \frac{|\rho_{12}\rho_{21}|\overline{\kappa}_{11}\overline{\kappa}_{22}\overline{\kappa}_{12}\overline{\kappa}_{21}}{|1-\rho_{22}||1-\rho_{11}|} + \sum_{k=2}^{\infty} |\rho_{12}\rho_{21}|\ \frac{\overline{\kappa}_{11}\overline{\kappa}_{22}\overline{\kappa}_{12}\overline{\kappa}_{21}}{|1-\rho_{22}||1-\rho_{11}|}\ (|\rho_{12}\rho_{21}|)^{k-1}\ ||\prod_{j=1}^{k-1}\mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}\ \mathbf{A}_1\ ||_1$ and $||\prod_{j=1}^{k-1}\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}\ \mathbf{A}_1||_1 = ||\prod_{j=1}^{k-1}\ \mathbf{A}'_1\ \mathbf{W}_{n2,n1}\ \mathbf{A}'_2\ \mathbf{W}'_{n1,n2}||_{\infty} = ||\prod_{j=1}^{k-1}\mathbf{A}_1\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}||_{\infty} \leq \left(\frac{1}{|1-\rho_{22}||1-\rho_{11}|}\right)^{k-1}$ it follows $\|\mathbf{I}_{n1} - \rho_{12}\rho_{21}\mathbf{A}_1\ \mathbf{W}_{n1,n2}\ \mathbf{A}_2\ \mathbf{W}_{n2,n1}\|_1 \leq (1 + \overline{\kappa}_{11}\overline{\kappa}_{22}\overline{\kappa}_{12}\overline{\kappa}_{21}) \sum_{k=0}^{\infty} \left(\frac{|\rho_{12}\rho_{21}|}{|1-\rho_{22}||1-\rho_{11}|}\right)^k < \infty$ ■

C. Appendix to Paper III

Technical Appendix

Useful Lemmas

Notation: We first provide some useful Lemmas. For these Lemmas we drop the index for the true parameter. Let $\epsilon_i \sim i.i.d(0,\sigma^2)$ we then denote $\mathbf{Y} = (\mathbf{I}_n - \rho_0\mathbf{W}_n)^{-1}\,(\mathbf{X}_n\beta_0 + \mathbf{W}_n\mathbf{X}_n\gamma_0) + (\mathbf{I}_n - \rho_0\mathbf{W}_n)^{-1}\,\epsilon_n = \mathbf{S}_n\,(\mu_n + \epsilon_n)$, where $\mathbf{S}_n = (\mathbf{I}_n - \rho_0\mathbf{W}_n)^{-1}$. Additionally we denote, $\mu_{\epsilon,3} = E[\epsilon_i^3]$, $\mu_{\epsilon,4} = E[\epsilon_i^4]$ and the typical element of μ_n is denoted by μ_{jn}. If $\mathbf{A}_n$ is a symmetric matrix, then we denote $\mathbf{S}_n\mathbf{A}_n$ as $\bar{\mathbf{A}}_n$ and $\mathbf{S}_n\mathbf{A}_n\mathbf{S}_n$ as $\breve{\mathbf{A}}_n$

Lemmas

Lemma (1)-(9): *Let $\mathbf{A}_n$ be a symmetric n by n matrix. We denote the typical element of A_n with a_{ij}. Then the following equations hold:*

(1) $E\left[(\epsilon_n'\mathbf{A}_n\epsilon_n)\right] = \sigma^2 tr(\mathbf{A}_n)$

(2) $E\left[(\epsilon_n'\mathbf{A}_n\epsilon_n)^2\right] = \left(\mu_{\epsilon,4} - 3\sigma^4\right)\sum_i a_{ii}^2 + \sigma^4\left(tr\left(\mathbf{A}_n\right)^2 + tr\left(\mathbf{A}_n^2\right) + tr\left(\mathbf{A}_n\mathbf{A}_n'\right)\right)$

(3) $E\left[(y_n'\mathbf{A}_n\epsilon_n)\right] = \sigma^2 tr(\bar{\mathbf{A}}_n)$

(4) $E\left[(y_n'\mathbf{A}_n y_n)\right] = \sigma^2 tr(\breve{\mathbf{A}}_n) + \mu_n'\breve{\mathbf{A}}_n\mu_n$

(5) $E\left[(y_n'\mathbf{A}_n y_n)^2\right] = \left(\mu_{\epsilon,4} - 3\sigma^4\right)\sum_i \breve{a}_{ii}^2 + \sigma^4\left(tr\left(\breve{\mathbf{A}}_n\right)^2 + 2tr\left(\breve{\mathbf{A}}_n^2\right)\right) + \left(\mu_n'\breve{\mathbf{A}}_n\mu_n\right)^2 + 4\,\sigma^2$ $\mu_n'\breve{\mathbf{A}}_n'\,\breve{\mathbf{A}}_n\mu_n + 2\mu_n'\,\breve{\mathbf{A}}_n\,\mu_n\sigma^2 tr(\breve{\mathbf{A}}_n) + 2\mu_3\sum_{ij}\breve{a}_{ii}\breve{a}_{ij}\mu_{jn}$

(6) $E\left[(y_n'\mathbf{A}_n\epsilon_n)^2\right] = \left(\mu_{\epsilon,4} - 3\sigma^4\right)\sum_i \bar{a}_{ii}^2 + \sigma^4\left(tr\left(\bar{\mathbf{A}}_n\right)^2 + 2tr\left(\bar{\mathbf{A}}_n^2\right)\right) + \sigma^2\,\mu_n'\bar{\mathbf{A}}_n'\bar{\mathbf{A}}_n\mu_n + 2\mu_3$ $\sum_{ij}\bar{a}_{ii}\bar{a}_{ij}\mu_{jn}$

(7) $Var\left[y_n'\mathbf{A}_n y_n\right] = \left(\mu_{\epsilon,4} - 3\sigma^4\right)\sum_i \breve{a}_{ii}^2 + 4\,\sigma^4\,tr\left(\breve{\mathbf{A}}_n^2\right) + 4\,\sigma^2\,\mu_n'\breve{\mathbf{A}}_n\breve{\mathbf{A}}_n\mu_n + 2\mu_3\sum_{ij}\breve{a}_{ii}\breve{a}_{ij}$ $\mu_{jn} - 4\,\sigma^2\,tr(\breve{\mathbf{A}}_n)\,\mu_n'\breve{\mathbf{A}}_n\mu_n$

(8) $Var\left[y_n'\mathbf{A}_n\epsilon_n\right] = \left(\mu_4 - 3\sigma^4\right)\sum_i \bar{a}_{ii}^2 + 2\sigma^4\,tr\left(\bar{\mathbf{A}}_n^2\right) + 4\sigma^2\,\mu_n'\bar{\mathbf{A}}_n\bar{\mathbf{A}}_n\mu_n + 2\mu_3\sum_{ij}\bar{a}_{ii}\bar{a}_{ij}\mu_{jn}$

$- 4\sigma^2\ tr(\bar{\mathbf{A}}_n)\ \mu_n' \bar{\mathbf{A}}_n \mu_n$

(9) $Cov\left[\left(y_n' \mathbf{A}_n y_n\right), y_n' \mathbf{B}_n \epsilon_n\right] = \mu_n' \check{\mathbf{A}}_n \mu_n\ \sigma^4\ tr(\bar{\mathbf{B}}_n) + 2\ \mu_3 \sum_{ij} \bar{b}_{ii} \check{a}_{ij} \mu_{jn} + 2\ \sigma^2 \sum_l \sum_j \left(\check{a}_{lj} \mu_{jn}\right)$

$\sum_i \left(\mu_{in} \bar{b}_{il}\right) + \ \mu_3 \sum_{ij} \check{a}_{ii} \bar{b}_{ij} \mu_{jn} + \ \mu_4 \sum_i \check{a}_{i,i} \bar{b}_{i,i} + \sigma^4 \sum_{i,i,l,l} \check{a}_{i,i} \bar{b}_{l,l} + \ 2\sigma^4 \sum_{i,j} \check{a}_{i,j} \bar{b}_{i,j} - \ 2\ \sigma^2$ $tr(\bar{\mathbf{B}}_n)\ \mu_n' \check{\mathbf{A}}_n \mu_n - 3\ \sigma^4\ tr(\check{\mathbf{A}}_n)\ tr(\bar{\mathbf{B}}_n) - \ \mu_n' \check{\mathbf{A}}_n \mu_n \sigma^2 tr(\bar{\mathbf{B}}_n)$

Proof for Lemmas

Proof. [Proof for Lemma 1,2 and 7] see Lee (2004) ■

Proof. [Proof for Lemma 3] $E\left[\left(y_n' \mathbf{A}_n \epsilon_n\right)\right] = E\left[\left(\mu_n' \bar{\mathbf{A}}_n \epsilon_n\right)\right] + E\left[\left(\epsilon_n' \bar{\mathbf{A}}_n \epsilon_n\right)\right] = \sigma^2 tr(\bar{\mathbf{A}}_n)$ ■

Proof. [Proof for Lemma 4] $E\left[\left(y_n' \mathbf{A}_n y_n\right)\right] = E\left[\left(\mu_n' \check{\mathbf{A}}_n \mu_n\right)\right] + 2E[\mu_n' \check{\mathbf{A}}_n \epsilon_n] + E\left[\left(\epsilon_n' \check{\mathbf{A}}_n \epsilon_n\right)\right] = \sigma^2\ tr(\check{\mathbf{A}}_n) + \ \mu_n' \check{\mathbf{A}}_n \mu_n$ ■

Proof. [Proof for Lemma 5] $E\left[\left(y_n' \mathbf{A}_n y_n\right)^2\right] = E[(\mu_n' \check{\mathbf{A}}_n \mu_n + 2\mu_n' \check{\mathbf{A}}_n \epsilon_n + \epsilon_n' \check{\mathbf{A}}_n \epsilon_n\ )^2] = E[(\mu_n' \check{\mathbf{A}}_n \mu_n$ $+ \mu_n' \left(\check{\mathbf{A}}_n + \check{\mathbf{A}}_n'\right) \epsilon_n\ + \epsilon_n' \check{\mathbf{A}}_n \epsilon_n)^2] = E[(\mu_n' \check{\mathbf{A}}_n \mu_n)^2 + (\mu_n'\ (\check{\mathbf{A}}_n + \check{\mathbf{A}}_n')\ \epsilon_n)^2 + \left(\epsilon_n' \check{\mathbf{A}}_n \epsilon_n\right)^2 + 2\ \mu_n'$ $\left(\check{\mathbf{A}}_n + \check{\mathbf{A}}_n'\right) \epsilon_n\ \mu_n'\ \check{\mathbf{A}}_n\ \mu_n + 2\ \mu_n'\ \left(\check{\mathbf{A}}_n + \check{\mathbf{A}}_n'\right) \epsilon_n\ \epsilon_n'\ \check{\mathbf{A}}_n\ \epsilon_n + 2\ \mu_n'\ \check{\mathbf{A}}_n\ \mu_n\ \epsilon_n'\ \check{\mathbf{A}}_n\ \epsilon_n] = (\mu_4 - 3\sigma^4) \sum_i \check{a}_{ii}^2$ $+ \sigma^4 (tr(\check{\mathbf{A}}_n)^2 + tr(\check{\mathbf{A}}_n^2)\ + tr(\check{\mathbf{A}}_n \check{\mathbf{A}}_n'))\ + (\mu_n' \check{\mathbf{A}}_n \mu_n)^2 + \ 2\mu_n' \check{\mathbf{A}}_n \mu_n\ \sigma^2 tr(\check{\mathbf{A}}_n) + \sigma^2\ \mu_n'\ (\check{\mathbf{A}}_n + \check{\mathbf{A}}_n')^2$ $\mu_n + 2\ \mu_3 \sum_{ij} \check{a}_{ii}\ \check{a}_{ij}\ \mu_{jn}$ ■

Proof. [Proof for Lemma 6] $E[(y_n' \mathbf{A}_n \epsilon_n)^2] = E[\left((\mu_n + \epsilon_n)'\ \bar{\mathbf{A}}_n \epsilon_n\right)^2] = E[(\epsilon_n' \bar{\mathbf{A}}_n \epsilon_n + \mu_n' \bar{\mathbf{A}}_n \epsilon_n)^2]$ $= E[\left(\epsilon_n' \bar{\mathbf{A}}_n \epsilon_n\right)^2 + \left(\mu_n' \bar{\mathbf{A}}_n \epsilon_n\right)^2 + 2\left(\epsilon_n' \bar{\mathbf{A}}_n \epsilon_n \mu_n' \bar{\mathbf{A}}_n \epsilon_n\right)] = E\left[\left(\epsilon_n' \bar{\mathbf{A}}_n \epsilon_n\right)^2\right] + \ \sigma^2\ \mu_n'\ \bar{\mathbf{A}}_n\ \bar{\mathbf{A}}_n\ \mu_n + 2$ $\mu_3 \sum_{ij} \bar{a}_{ii} \bar{a}_{ij} \mu_{jn}$ ■

Proof. [Proof for Lemma 8] $Var\left[y_n' \mathbf{A}_n \epsilon_n\right] = E[(y_n' \mathbf{A}_n \epsilon_n - E[y_n' \mathbf{A}_n \epsilon_n])^2] = E[(y_n' \mathbf{A}_n \epsilon_n - \sigma^2 tr(\bar{\mathbf{A}}_n))^2] = E[(y_n' \mathbf{A}_n \epsilon_n)^2 + \sigma^4 tr(\bar{\mathbf{A}}_n)^2 - 2y_n' \mathbf{A}_n \epsilon_n \sigma^2 tr(\bar{\mathbf{A}}_n)\ ] = (\mu_4 - 3\sigma^4) \sum_i \bar{a}_{ii}^2 +$ $\sigma^4 \left(tr\left(\bar{\mathbf{A}}_n\right)^2 + 2tr\left(\bar{\mathbf{A}}_n^2\right)\right) + \sigma^2 \mu_n' \bar{\mathbf{A}}_n \bar{\mathbf{A}}_n \mu_n + 2\mu_3 \sum_{ij} \bar{a}_{ii} \bar{a}_{ij} \mu_{jn} \mu_n' + \sigma^4 tr(\bar{\mathbf{A}}_n)^2 - 2\sigma^4 tr(\bar{\mathbf{A}}_n)^2$ $= (\mu_4 - 3\sigma^4) \sum_i \bar{a}_{ii}^2 + \ \sigma^4 (tr\left(\bar{\mathbf{A}}_n\right)^2 + tr(\bar{\mathbf{A}}_n^2)) + \sigma^2 \mu_n' \bar{\mathbf{A}}_n \bar{\mathbf{A}}_n \mu_n + 2\mu_3 \sum_{ij} \bar{a}_{ii} \bar{a}_{ij} \mu_{jn} \mu_n'$ ■

Proof. [Proof for Lemma 9] $Cov\left[\left(y_n' \mathbf{A}_n y_n\right)\ , y_n' \mathbf{B}_n \epsilon_n\right] = E[(y_n' \check{\mathbf{A}}_n y_n - E[y_n' \check{\mathbf{A}}_n y_n])\ (y_n' \bar{\mathbf{B}}_n \epsilon_n - E\left[y_n' \bar{\mathbf{B}}_n \epsilon_n\right])] = E[(y_n' \check{\mathbf{A}}_n y_n - \sigma^2\ tr(\check{\mathbf{A}}_n) - \mu_n' \check{\mathbf{A}}_n \mu_n)\ \left(\ y_n' \bar{\mathbf{B}}_n \epsilon_n - \sigma^2 tr(\bar{\mathbf{B}}_n)\ \right)] = E[y_n' \check{\mathbf{A}}_n y_n$ $y_n' \bar{\mathbf{B}}_n \epsilon_n - \sigma^2\ tr(\check{\mathbf{A}}_n)\ y_n' \bar{\mathbf{B}}_n \epsilon_n - \ \mu_n' \check{\mathbf{A}}_n \mu_n\ y_n' \bar{\mathbf{B}}_n \epsilon_n - y_n' \check{\mathbf{A}}_n y_n\ \sigma^2 tr(\bar{\mathbf{B}}_n) - \ \sigma^4\ tr(\check{\mathbf{A}}_n)\ tr(\bar{\mathbf{B}}_n) -$

$\mu_n'\check{\mathbf{A}}_n\mu_n\ \sigma^2\ tr(\bar{\mathbf{B}}_n)] = E[y_n'\check{\mathbf{A}}_n y_n y_n'\bar{\mathbf{B}}_n\epsilon_n] - \ \sigma^4\ tr(\check{\mathbf{A}}_n)\ tr(\bar{\mathbf{B}}_n) - \ \sigma^2\ tr(\bar{\mathbf{B}}_n)\ \mu_n'\check{\mathbf{A}}_n\mu_n - (\sigma^2$ $tr(\check{\mathbf{A}}_n) + \mu_n'\check{\mathbf{A}}_n\mu_n)\ \ \sigma^2\ tr(\bar{\mathbf{B}}_n) - \ \sigma^4\ tr(\check{\mathbf{A}}_n)\ tr(\bar{\mathbf{B}}_n) - \ \mu_n'\ \check{\mathbf{A}}_n\ \mu_n\ \sigma^2\ tr(\bar{\mathbf{B}}_n) = \mu_n'\ \check{\mathbf{A}}_n\ \mu_n$ $\sigma^4 tr(\bar{\mathbf{B}}_n) + \ 2\ \mu_3\ \sum_{ij}\bar{b}_{ii}\check{a}_{ij}\mu_{jn}\ +\ 2\ \sigma^2\ \sum_l\sum_j\left(\check{a}_{lj}\mu_{jn}\right)\ \sum_i\left(\mu_{in}\bar{b}_{il}\right) + \ \ \mu_3\sum_{ij}\check{a}_{ii}\bar{b}_{ij}\mu_{jn}\ +\ \mu_4$ $\sum_i\check{a}_{i,i}\bar{b}_{i,i} + \sigma^4\sum_{i,i,l,l}\check{a}_{i,i}\bar{b}_{l,l} + 2\sigma^4\sum_{i,j}\check{a}_{i,j}\bar{b}_{i,j} - \ 2\sigma^2 tr(\bar{\mathbf{B}}_n)\mu_n'\check{\mathbf{A}}_n\mu_n - 3\ \sigma^4\ tr(\check{\mathbf{A}}_n)\ tr(\bar{\mathbf{B}}_n) -$ $\mu_n'\ \check{\mathbf{A}}_n\ \mu_n\ \sigma^2\ tr(\bar{\mathbf{B}}_n)$, since $E[y_n'\check{\mathbf{A}}_n y_n\ y_n'\bar{\mathbf{B}}_n\epsilon_n] = E[(\mu_n+\epsilon_n)'\ \check{\mathbf{A}}_n\ (\mu_n+\epsilon_n)\ (\mu_n+\epsilon_n)'\ \bar{\mathbf{B}}_n\ \epsilon_n]$ $= E[\mu_n'\ \check{\mathbf{A}}_n\ \mu_n\ (\mu_n+\epsilon_n)'\ \bar{\mathbf{B}}_n\ \epsilon_n] + E[\ 2\ \epsilon_n'\ \check{\mathbf{A}}_n\ \mu_n\ (\mu_n+\epsilon_n)'\ \bar{\mathbf{B}}_n\ \epsilon_n] + \ E[\epsilon_n'\ \check{\mathbf{A}}_n\ \epsilon_n\ (\mu_n+\epsilon_n)'$ $\bar{\mathbf{B}}_n\ \epsilon_n\ ] = E[\mu_n'\ \check{\mathbf{A}}_n\ \mu_n\ \mu_n'\bar{\mathbf{B}}_n\epsilon_n + \ \mu_n'\check{\mathbf{A}}_n\mu_n\ \epsilon_n'\bar{\mathbf{B}}_n\epsilon_n + \ 2\epsilon_n'\check{\mathbf{A}}_n\mu_n\ \epsilon_n'\bar{\mathbf{B}}_n\epsilon_n\ + \ 2\epsilon_n'\check{\mathbf{A}}_n\mu_n\ \mu_n'\bar{\mathbf{B}}_n\epsilon_n$ $+\ \epsilon_n'\check{\mathbf{A}}_n\epsilon_n\ \epsilon_n'\bar{\mathbf{B}}_n\epsilon_n\ +\ \epsilon_n'\check{\mathbf{A}}_n\epsilon_n\ \mu_n'\bar{\mathbf{B}}_n\epsilon_n] = \ E[\mu_n'\check{\mathbf{A}}_n\mu_n\ \epsilon_n'\bar{\mathbf{B}}_n\epsilon_n\ +\ 2\epsilon_n'\check{\mathbf{A}}_n\mu_n\ \epsilon_n'\bar{\mathbf{B}}_n\epsilon_n\ +\ 2\epsilon_n'\check{\mathbf{A}}_n\mu_n$ $\mu_n'\bar{\mathbf{B}}_n\epsilon_n + \epsilon_n'\check{\mathbf{A}}_n\epsilon_n\ \epsilon_n'\bar{\mathbf{B}}_n\epsilon_n\ \ + \epsilon_n'\check{\mathbf{A}}_n\epsilon_n\ \mu_n'\bar{\mathbf{B}}_n\epsilon_n] = \mu_n'\check{\mathbf{A}}_n\mu_n\ \sigma^4 tr(\bar{\mathbf{B}}_n)\ + 2\mu_3\ \sum_{ij}\bar{b}_{ii}\check{a}_{ij}\mu_{jn} + 2\sigma^2$ $\sum_l\sum_j\left(\check{a}_{lj}\mu_{jn}\right)\ \sum_i\left(\mu_{in}\bar{b}_{il}\right) + \mu_3\sum_{ij}\check{a}_{ii}\bar{b}_{ij}\mu_{jn} + \mu_4\sum_i\check{a}_{ii}\bar{b}_{ii} + \sigma^4\sum_{i,i,l,l}\check{a}_{ii}\bar{b}_{ll} + 2\sigma^4\sum_{i,j}\check{a}_{ij}\bar{b}_{i,j}$, $E[2\ \epsilon_n'\check{\mathbf{A}}_n\mu_n\ \mu_n'\bar{\mathbf{B}}_n\epsilon_n] = 2\sigma^2\ \sum_l\sum_j\left(\check{a}_{lj}\mu_{jn}\right)\ \sum_i\left(\mu_{in}\bar{b}_{il}\right)$ and $E[\epsilon_n'\check{\mathbf{A}}_n\epsilon_n\ \epsilon_n'\bar{\mathbf{B}}_n\epsilon_n] = E[\sum_{i,j,k,l}$ $\epsilon_i\check{a}_{ij}\epsilon_j\epsilon_k\bar{b}_{kl}\epsilon_l] = \mu_4\sum_i\check{a}_{ii}\bar{b}_{ii} + \sigma^4\sum_{i,i,l,l}\check{a}_{ii}\bar{b}_{ll} + \sigma^4\sum_{i,j}\check{a}_{ij}\bar{b}_{ij} + \sigma^4\sum_{i,j}\check{a}_{ij}\bar{b}_{ji}$ ■

Theorems

Theorem (3.1) The asymptotic expected value for $\hat{\rho}_1$ as it is given in Step 2 is given by: $\lim_{n\to\infty} E\left[\hat{\rho}_1\right] = \rho_0 + \frac{\sigma_0^2 tr\left(\mathbf{S}_n'\mathbf{W}_n'\mathbf{M}_x\overline{\mathbf{DD}}_n\right)}{\left(\sigma_0^2+\sigma_\mu^2\right)tr\left(\mathbf{S}_n'\mathbf{W}_n'\mathbf{M}_x\mathbf{W}_n\mathbf{S}_n\right)}$

Theorem (3.2) The asymptotic variance for $\hat{\rho}_1$ as it is given in Step 2 is given by: $\lim_{n\to\infty} Var\left(\hat{\rho}\right) = \lim_{n\to\infty} Var\left(\hat{\rho}-\rho_0\right) = Var\left(\frac{\zeta}{\vartheta}\right) \approx \frac{\mu_\zeta^2}{\mu_\vartheta^4}Var\left(\vartheta\right) + \frac{1}{\mu_\vartheta^2}Var\left(\zeta\right) - \frac{2\mu_\zeta}{\mu_\vartheta^3}Cov(\zeta,\vartheta)$ where $\zeta = \mathbf{Y}_n'\mathbf{A}_n\epsilon_n$, $\vartheta = \mathbf{Y}_n'\mathbf{B}_n\mathbf{Y}_n$,where $\mathbf{A}_n = \mathbf{W}_n'\mathbf{M}_x\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{M}_x\mathbf{D}_n\right)^{-1}\mathbf{D}_n\mathbf{M}_x$ and $\mathbf{B}_n = \mathbf{W}_n'\mathbf{M}_x\mathbf{W}_n$. The other variables are given by: $\mu_\zeta = \sigma^2 tr(\bar{\mathbf{A}}_n)$, $\mu_\vartheta = \sigma^2 tr(\check{\mathbf{B}}_n) + \mu_n'\check{\mathbf{B}}_n\mu_n$,

Theorem (3.3) $Var\left(\zeta\right) = \left(\mu_{\epsilon,4} - 3\sigma^4\right)\sum_i\bar{a}_{ii}^2 + \sigma^4\left(tr\left(\bar{\mathbf{A}}_n\right)^2 + 2tr\left(\bar{\mathbf{A}}_n^2\right)\right) + \sigma^2\ \mu_n'\bar{\mathbf{A}}_n'\bar{\mathbf{A}}_n\mu_n$ $+2\mu_3\sum_{ij}\bar{a}_{ii}\bar{a}_{ij}\mu_{jn}$, $Var\left(\vartheta\right) = \left(\mu_4 - 3\sigma^4\right)\sum_i\check{b}_{ii}^2 + 2\sigma^4\ tr\left(\check{\mathbf{B}}_n^2\right) + 4\sigma^2\ \mu_n'\check{\mathbf{B}}_n\check{\mathbf{B}}_n\mu_n + 2\mu_3\sum_{ij}\check{b}_{ii}\check{b}_{ij}\mu_{jn}$ $-4\sigma^2\ \ tr(\check{\mathbf{B}}_n)\ \ \mu_n'\check{\mathbf{B}}_n\mu_n$ and $Cov(\zeta,\vartheta) = \ \ \mu_n'\check{\mathbf{A}}_n\mu_n\ \ \sigma^4\ \ tr(\bar{\mathbf{B}}_n)\ \ +\ \ 2\ \ \mu_3\ \sum_{ij}\bar{b}_{ii}\check{a}_{ij}\mu_{jn} +\ 2\ \ \sigma^2$ $\sum_l\sum_j\left(\check{a}_{lj}\mu_{jn}\right)\ \sum_i\left(\mu_{in}\bar{b}_{il}\right)\ +\ \ \mu_3\ \sum_{ij}\check{a}_{ii}\bar{b}_{ij}\mu_{jn}\ +\ \mu_4\ \sum_i\check{a}_{i,i}\bar{b}_{i,i}\ +\ \sigma^4\ \sum_{i,i,l,l}\check{a}_{i,i}\bar{b}_{l,l}\ +\ \ 2\ \sigma^4$ $\sum_{i,j}\check{a}_{i,j}\bar{b}_{i,j} -\ 2\ \sigma^2\ tr(\bar{\mathbf{B}}_n)\ \mu_n'\check{\mathbf{A}}_n\mu_n\ -3\sigma^4\ tr(\check{\mathbf{A}}_n)\ tr(\bar{\mathbf{B}}_n) -\ \mu_n'\check{\mathbf{A}}_n\mu_n\ \sigma^2\ tr(\bar{\mathbf{B}}_n)$

Proof for Theorems

Proof. [for Theorem 3.1] $\Rightarrow \lim_{n\to\infty} E[\hat{\rho}_1] = \rho_0 + \lim_{n\to\infty} E[\ (\mathbf{Y}_n'\mathbf{W}_n'\ \mathbf{M}_x\ \mathbf{W}_n\mathbf{Y}_n)^{-1}$ $\left(\mathbf{Y}_n'\mathbf{W}_n'\ \mathbf{M}_x\ \bar{\mathbf{D}}_n\ \varepsilon\right)] = \rho_0 + \frac{E\left[\mathbf{Y}_n'\mathbf{W}_n'\mathbf{M}_x\bar{\mathbf{D}}_n\boldsymbol{\varepsilon}\right]}{E[\mathbf{Y}_n'\mathbf{W}_n'\mathbf{M}_n\mathbf{W}_n\mathbf{Y}_n]} \approx \rho_0 + \sigma_\epsilon^2\, tr(\mathbf{S}_n'\mathbf{W}_n'\ \mathbf{M}_n\bar{\mathbf{D}}_n)\ (\ \left(\sigma_\epsilon^2 + \sigma_\mu^2\right) tr(\mathbf{S}_n'\mathbf{W}_n'$ $\mathbf{M}_x\ \mathbf{W}_n\mathbf{S}_n))^{-1} \approx \rho_0 + \frac{\sigma_\epsilon^2 tr\left(\mathbf{S}'\mathbf{W}'\mathbf{M}\bar{\mathbf{D}}\right)}{\mathbf{Y}'\mathbf{W}'\mathbf{M}\mathbf{W}\mathbf{Y}}$ and $\sigma_\mu^2 = Var[\mu_n]$. Since $tr\left(\mathbf{S}_n'\mathbf{W}_n'\mathbf{M}_x\bar{\mathbf{D}}_n\right) = tr(\ (\mathbf{I}_n$ $-\ \rho_0\mathbf{W}_n')^{-1}\ \mathbf{W}_n'\mathbf{M}_x\bar{\mathbf{D}}_n) = tr(\left(\mathbf{I}_n - \rho_0\bar{\boldsymbol{\Lambda}}_n\right)^{-1}\ \bar{\boldsymbol{\Lambda}}_n\mathbf{M}_x\bar{\mathbf{D}}_n) = tr(\left(\mathbf{I}_n - \rho_0\bar{\boldsymbol{\Lambda}}_n\right)^{-1}\ \bar{\boldsymbol{\Lambda}}_n\boldsymbol{\Lambda}_{\mathbf{M}_x,n}\ \boldsymbol{\Lambda}_{\bar{\mathbf{D}},n})$ where $\boldsymbol{\Lambda}_{\mathbf{M}_x,n}$ is the eigenvalue matrix of $\mathbf{M}_x$ and $\boldsymbol{\Lambda}_{\bar{\mathbf{D}},n}$ the eigenvalue matrix of $\bar{\mathbf{D}}_n$. Note the first $2k$ diagonal-entries of matrix $\boldsymbol{\Lambda}_{\mathbf{M}_x,n}$ are zero and the rest are ones. Also the first q diagonal-entries are matrix $\boldsymbol{\Lambda}_{\bar{\mathbf{D}},n}$ are zero and the rest are ones. Therefore, $tr(\left(\mathbf{I}_n - \rho_0\bar{\boldsymbol{\Lambda}}_n\right)^{-1}\ \bar{\boldsymbol{\Lambda}}_n\boldsymbol{\Lambda}_{\mathbf{M}_x,n}$ $\boldsymbol{\Lambda}_{\bar{\mathbf{D}},n}) = \sum_{i=\max(2k,q)}^{n} \frac{\lambda_i}{1-\rho\lambda_i}$. Hence we for use in order to correct the bias:

$$\min_\rho \left|\widehat{\rho}_1 - \rho - \frac{\sigma^2 tr\left((\mathbf{I}_n - \rho\mathbf{W}_n')^{-1}\mathbf{W}_n'\mathbf{M}_x\overline{\mathbf{D}}\overline{\mathbf{D}}_n\right)}{\mathbf{Y}_n'\mathbf{W}_n'\mathbf{M}_x\mathbf{W}_n\mathbf{Y}_n}\right| \ \blacksquare$$

Proof. [for Theorem 3.2] We use the delta method approximation (see Green (1997) or Casella and Berger (2001)) for calculating the variance of $\hat{\rho}$**:** $Var\left(\hat{\rho}\right) = Var\left(\frac{\zeta}{\vartheta}\right) \approx \frac{\mu_\zeta^2}{\mu_\vartheta^4} Var\left(\vartheta\right) + \frac{1}{\mu_\vartheta^2} Var\left(\zeta\right) - \frac{2\mu_\zeta}{\mu_\vartheta^3} Cov(\zeta,\vartheta)$ where $\zeta = \mathbf{Y}_n'\mathbf{A}_n\epsilon_n$, $\vartheta = \mathbf{Y}_n'\mathbf{B}_n\mathbf{Y}_n$,where $\mathbf{A}_n = \mathbf{W}_n'\mathbf{M}_x\mathbf{D}_n\left(\mathbf{D}_n'\mathbf{M}_x\mathbf{D}_n\right)^{-1}\mathbf{D}_n\mathbf{M}_x$ and $\mathbf{B}_n = \mathbf{W}_n'\mathbf{M}_x\mathbf{W}_n$. The following equations follow directly by applying Lemmas (3), (4) and (7) to (9): $\mu_\zeta = \sigma^2 tr(\bar{\mathbf{A}}_n)$, $\mu_\vartheta = \sigma^2 tr(\check{\mathbf{B}}_n) + \mu_n'\check{\mathbf{B}}_n\mu_n$, $Var\left(\zeta\right) = \left(\mu_{\epsilon,4} - 3\sigma^4\right)\sum_i \bar{a}_{ii}^2 + \sigma^4 \left(tr\left(\bar{\mathbf{A}}_n\right)^2 + 2tr\left(\bar{\mathbf{A}}_n^2\right)\right) + \sigma^2\mu_n'\bar{\mathbf{A}}_n'\bar{\mathbf{A}}_n\mu_n + 2\mu_3\sum_{ij}\bar{a}_{ii}\bar{a}_{ij}\mu_{jn}$, $Var\left(\vartheta\right) = \left(\mu_4 - 3\sigma^4\right)\sum_i \check{b}_{ii}^2 + 2\sigma^4\ tr\left(\check{\mathbf{B}}_n^2\right) + 4\sigma^2\ \mu_n'\check{\mathbf{B}}_n\check{\mathbf{B}}_n\mu_n + 2\mu_3\sum_{ij}\check{b}_{ii}\check{b}_{ij}\mu_{jn}\ - 4\sigma^2\ tr(\check{\mathbf{B}}_n)\ \mu_n'\check{\mathbf{B}}_n\mu_n$ and $Cov(\zeta,\vartheta) = \mu_n'\check{\mathbf{A}}_n\mu_n\ \sigma^4 tr(\bar{\mathbf{B}}_n) + 2\mu_3\sum_{ij}\bar{b}_{ii}\check{a}_{ij}\mu_{jn} + 2\ \sigma^2\sum_l\sum_j\left(\check{a}_{lj}\mu_{jn}\right)\sum_i\left(\mu_{in}\bar{b}_{il}\right) + \mu_3\sum_{ij}\check{a}_{ii}\bar{b}_{ij}\mu_{jn} +$

$\mu_4\sum_i \check{a}_{i,i}\bar{b}_{i,i} + \sigma^4\sum_{i,i,l,l}\check{a}_{i,i}\bar{b}_{l,l} + 2\ \sigma^4\sum_{i,j}\check{a}_{i,j}\bar{b}_{i,j} - 2\sigma^2\ tr(\bar{\mathbf{B}}_n)\ \mu_n'\check{\mathbf{A}}_n\mu_n - 3\sigma^4\ tr(\check{\mathbf{A}}_n)\ tr(\bar{\mathbf{B}}_n)$ $-\ \mu_n'\check{\mathbf{A}}_n\mu_n\ \sigma^2 tr(\bar{\mathbf{B}}_n)$ $\blacksquare$

Additional Monte Carlo tables

Table C.1: Monte Carlo Results: Maximum Likelihood for ρ_0 where γ_0=1 and λ_0=0.99

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-0.8	-0.80%	-5.96%	-10.00%	-0.93%	-5.52%	-1.88%	-0.90%	-5.72%	-5.58%
	-0.4	-0.77%	-21.96%	-2.51%	-0.60%	-22.31%	-4.13%	-0.04%	-21.42%	-0.34%
	0	-0.76%	10.02%	-0.68%	-0.64%	10.10%	-1.55%	-0.74%	9.69%	2.70%
	0.4	-3.78%	21.96%	-0.09%	-4.68%	22.13%	0.20%	-4.77%	22.60%	-1.91%
	0.8	-1.98%	6.08%	-4.43%	-1.79%	6.11%	-6.32%	-2.29%	6.51%	-9.34%
200	-0.8	-0.52%	-4.06%	-7.24%	-0.40%	-3.68%	1.80%	-0.42%	-3.69%	1.56%
	-0.4	-0.34%	-15.31%	-0.97%	-0.78%	-15.92%	-4.61%	-0.63%	-14.99%	1.26%
	0	-0.27%	7.13%	-1.15%	-0.50%	7.00%	0.94%	-0.63%	6.93%	2.04%
	0.4	-1.85%	15.49%	-0.95%	-1.82%	16.50%	-7.18%	-2.74%	15.84%	-2.19%
	0.8	-0.88%	3.91%	-0.87%	-1.02%	4.13%	-5.43%	-0.94%	4.00%	-3.00%
400	-0.8	-0.32%	-2.65%	-0.18%	-0.27%	-2.81%	-6.39%	-0.18%	-2.50%	4.98%
	-0.4	-0.39%	-10.80%	-0.54%	-0.88%	-11.49%	-6.20%	-0.51%	-10.67%	0.78%
	0	-0.42%	4.86%	3.02%	-0.27%	4.74%	5.46%	-0.20%	4.93%	1.32%
	0.4	-2.06%	11.03%	-0.41%	-1.21%	10.69%	1.35%	-1.01%	11.18%	-3.50%
	0.8	-0.63%	2.66%	2.52%	-0.50%	2.74%	-2.10%	-0.42%	2.61%	1.72%

Table C.2: Monte Carlo Results: New estimator for ρ_0 where γ_0=1 and λ_0=0.99

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-0.8	-0.59%	-5.72%	-4.01%	-0.73%	-5.64%	-1.49%	-0.47%	-5.45%	0.93%
	-0.4	0.14%	-21.97%	-2.89%	1.06%	-22.13%	-3.54%	0.25%	-22.20%	-3.90%
	0	-1.22%	10.09%	-0.99%	-0.69%	9.72%	2.37%	-0.55%	10.43%	-4.44%
	0.4	-3.49%	22.73%	-2.16%	-4.24%	22.70%	-1.01%	-4.93%	22.29%	1.95%
	0.8	-2.01%	6.09%	4.75%	-1.90%	5.68%	11.87%	-2.01%	6.11%	4.63%
200	-0.8	-0.12%	-3.86%	-2.44%	-0.17%	-3.81%	-1.16%	-0.21%	-3.84%	-1.41%
	-0.4	0.56%	-14.83%	2.08%	-0.13%	-15.30%	-1.14%	-0.02%	-15.84%	-4.41%
	0	-0.33%	7.22%	-2.76%	-0.05%	7.14%	-1.79%	-0.54%	7.17%	-1.88%
	0.4	-2.15%	15.57%	-0.05%	-1.65%	15.29%	1.57%	-1.82%	15.43%	0.80%
	0.8	-0.86%	4.06%	-0.36%	-0.91%	4.08%	-0.71%	-0.82%	3.78%	6.68%
400	-0.8	-0.07%	-10.71%	0.22%	-0.14%	-2.60%	1.76%	-0.23%	-2.58%	2.46%
	-0.4	-0.22%	5.12%	-2.95%	0.93%	-10.87%	-1.40%	-0.01%	-10.76%	-0.85%
	0	-1.13%	10.94%	-0.92%	-0.15%	4.93%	1.07%	-0.08%	4.95%	0.77%
	0.4	-0.58%	2.81%	-1.87%	-1.12%	11.17%	-2.91%	-0.99%	10.47%	3.29%
	0.8	-0.14%	-2.60%	1.76%	-0.37%	2.71%	-0.41%	-0.41%	2.67%	1.52%

Table C.3: Monte Carlo Results: Maximum Likelihood for ρ_0 where γ_0=1 and λ_0=0.00

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-0.8	-1.16%	-5.91%	-7.33%	-0.88%	-5.50%	-2.48%	-1.07%	-5.46%	-8.20%
	-0.4	-2.92%	-22.08%	-1.94%	-1.09%	-21.84%	-4.02%	-1.26%	-19.92%	-4.92%
	0	-0.29%	9.82%	0.55%	-0.57%	9.44%	-0.08%	-0.30%	8.06%	0.61%
	0.4	-0.96%	21.15%	0.06%	-1.93%	20.13%	-4.51%	-0.81%	15.40%	-3.71%
	0.8	-1.73%	5.95%	-6.92%	-1.05%	4.96%	-5.84%	-0.13%	2.17%	-0.75%
200	-0.8	-0.73%	-4.03%	-5.62%	-0.47%	-3.78%	-2.51%	-0.59%	-3.46%	-4.36%
	-0.4	-1.34%	-15.44%	-1.76%	-0.30%	-14.82%	-3.04%	-0.91%	-12.85%	-1.92%
	0	-0.63%	6.91%	1.14%	-0.08%	6.18%	6.65%	0.07%	5.08%	2.79%
	0.4	-2.19%	15.69%	-3.42%	-1.59%	13.90%	-0.03%	-0.51%	10.27%	-1.16%
	0.8	-0.71%	3.66%	-0.85%	-0.35%	3.07%	-2.42%	-0.11%	1.61%	-0.38%
400	-0.8	-0.20%	-2.61%	0.28%	-0.37%	-2.48%	5.17%	-0.13%	-2.43%	-4.52%
	-0.4	-0.73%	-10.90%	-1.73%	-0.79%	-10.91%	-4.06%	-0.41%	-9.38%	-3.36%
	0	-0.22%	4.84%	2.43%	-0.15%	4.65%	2.60%	0.06%	3.79%	3.09%
	0.4	-1.15%	11.40%	-6.20%	-1.15%	10.55%	-4.12%	-0.40%	8.20%	-6.13%
	0.8	-0.41%	2.59%	-0.41%	-0.34%	2.28%	2.32%	-0.08%	1.54%	-5.22%

Table C.4: Monte Carlo Results: New estimator for ρ_0 where γ_0=1 and λ_0=0.00

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-0.8	-1.02%	-5.80%	-1.42%	-0.75%	-5.40%	3.41%	-0.93%	-5.33%	-2.84%
	-0.4	-2.51%	-21.99%	-0.42%	-0.66%	-21.68%	-2.35%	-0.88%	-19.87%	-4.11%
	0	-0.28%	9.88%	0.31%	-0.56%	9.50%	-0.33%	-0.30%	8.12%	-0.06%
	0.4	-0.52%	21.01%	2.06%	-1.64%	19.98%	-2.91%	-0.60%	15.29%	-2.52%
	0.8	-1.66%	5.84%	1.11%	-0.94%	4.84%	-0.42%	-0.14%	2.15%	0.93%
200	-0.8	-0.59%	-3.96%	-2.51%	-0.28%	-3.73%	0.21%	-0.46%	-3.36%	-0.14%
	-0.4	-0.85%	-15.14%	0.39%	0.05%	-14.55%	-1.10%	-0.54%	-12.65%	-0.23%
	0	-0.66%	6.99%	0.11%	-0.10%	6.26%	5.34%	0.07%	5.16%	1.06%
	0.4	-1.87%	15.41%	-1.27%	-1.21%	13.63%	1.79%	-0.33%	10.12%	0.33%
	0.8	-0.66%	3.67%	1.69%	-0.30%	3.09%	-2.40%	-0.12%	1.58%	1.73%
400	-0.8	-0.22%	-10.62%	0.76%	-0.31%	-2.48%	5.98%	-0.07%	-2.42%	-3.73%
	-0.4	-0.25%	4.94%	0.49%	-0.22%	-10.59%	-1.33%	0.01%	-9.12%	-0.60%
	0	-0.77%	11.11%	-3.80%	-0.18%	4.73%	0.41%	0.04%	3.85%	1.02%
	0.4	-0.42%	2.64%	-0.59%	-0.79%	10.21%	-0.92%	-0.15%	8.02%	-4.62%
	0.8	-0.31%	-2.48%	5.98%	-0.31%	2.32%	1.18%	-0.12%	1.47%	-0.20%

Table C.5: Monte Carlo Results: Maximum Likelihood for ρ_0 where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

		R^2=.1			R^2=.4			R^2=.8		
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-0.8	-0.86%	-5.75%	-6.32%	-0.94%	-5.66%	-4.39%	-1.00%	-5.54%	-1.80%
	-0.4	-0.86%	-22.18%	-3.37%	-0.65%	-21.37%	0.27%	-1.21%	-21.80%	-1.43%
	0	0.18%	9.90%	0.43%	-0.11%	9.98%	-0.40%	-0.50%	10.79%	-7.91%
	0.4	-4.08%	22.12%	-0.33%	-2.74%	21.83%	-0.36%	-3.11%	21.47%	1.73%
	0.8	-1.76%	6.10%	-6.27%	-1.65%	5.99%	-5.21%	-1.60%	6.24%	-9.70%
200	-0.8	-0.68%	-3.95%	-3.57%	-0.50%	-3.77%	-0.05%	-0.58%	-3.96%	-4.52%
	-0.4	-0.15%	-15.32%	-1.07%	-1.00%	-15.84%	-3.90%	-1.13%	-15.77%	-3.38%
	0	-0.32%	6.85%	3.05%	0.32%	6.91%	2.22%	-0.18%	7.19%	-1.99%
	0.4	-1.29%	15.88%	-3.86%	-1.72%	15.62%	-1.84%	-1.46%	15.61%	-2.04%
	0.8	-0.68%	3.92%	-2.84%	-0.84%	4.22%	-8.89%	-0.75%	3.88%	-1.07%
400	-0.8	-0.38%	-2.66%	-0.01%	-0.25%	-2.49%	5.65%	-0.43%	-2.70%	-1.07%
	-0.4	-1.16%	-11.14%	-2.82%	-0.72%	-11.65%	-7.64%	-0.68%	-11.01%	-2.25%
	0	0.13%	4.81%	3.84%	0.10%	5.10%	-2.11%	0.10%	4.99%	0.03%
	0.4	-1.27%	10.89%	-0.43%	-1.03%	11.23%	-3.76%	-0.44%	11.23%	-4.32%
	0.8	-0.47%	2.63%	1.78%	-0.35%	2.67%	-0.70%	-0.31%	2.66%	-0.82%

Table C.6: Monte Carlo Results: Maximum Likelihood for the total effect of x where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

			R^2=.1			R^2=.4			R^2=.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-0.8	0.13%	4.37%	1.01%	-0.02%	1.76%	3.19%	-0.01%	0.68%	3.60%
	-0.4	-0.24%	7.78%	0.06%	-0.06%	3.95%	-1.08%	0.01%	1.49%	-2.03%
100	0	-0.91%	15.01%	3.98%	0.28%	5.65%	2.53%	0.07%	2.60%	0.87%
	0.4	0.54%	22.34%	0.90%	0.40%	7.94%	6.76%	-0.17%	4.03%	0.53%
	0.8	1.33%	39.30%	3.37%	0.63%	16.34%	4.68%	0.07%	5.80%	8.54%
	-0.8	-0.05%	2.57%	2.33%	-0.02%	1.04%	0.25%	-0.01%	0.48%	-0.54%
	-0.4	-0.05%	7.46%	-3.26%	0.13%	2.76%	0.20%	0.01%	1.12%	2.67%
200	0	0.13%	11.15%	-2.90%	-0.16%	4.42%	0.66%	-0.05%	1.70%	2.14%
	0.4	0.75%	17.01%	0.40%	-0.29%	7.14%	2.83%	0.06%	2.49%	0.80%
	0.8	-0.68%	25.69%	0.57%	-0.42%	8.93%	6.87%	-0.05%	4.89%	0.66%
	-0.8	-0.01%	2.58%	0.72%	0.00%	0.84%	-0.32%	0.01%	0.36%	-1.13%
	-0.4	-0.01%	5.19%	1.39%	0.06%	2.24%	-2.58%	-0.05%	0.94%	-1.24%
400	0	0.14%	8.43%	1.61%	0.11%	3.50%	-0.06%	0.00%	1.46%	2.44%
	0.4	-0.02%	12.19%	-1.02%	0.09%	5.01%	5.41%	0.06%	2.07%	-1.15%
	0.8	0.95%	19.48%	1.95%	-0.25%	8.05%	3.77%	-0.13%	3.30%	2.60%

Table C.7: Monte Carlo Results: New estimator for ρ_0 where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

			R^2=.1			R^2=.4			R^2=.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-0.8	-0.67%	-5.68%	-2.43%	-0.56%	-5.17%	6.76%	-0.64%	-5.54%	-0.18%
	-0.4	-0.44%	-21.40%	0.40%	0.82%	-21.38%	-0.11%	-0.55%	-21.90%	-2.36%
100	0	-0.60%	9.96%	0.09%	-0.32%	9.76%	2.19%	-0.74%	10.20%	-2.56%
	0.4	-4.37%	22.76%	-1.17%	-4.60%	22.77%	-1.16%	-4.17%	21.27%	5.20%
	0.8	-1.38%	5.78%	5.24%	-2.39%	6.20%	5.10%	-1.50%	5.73%	4.68%
	-0.8	-0.27%	-3.77%	0.37%	-0.32%	-3.71%	1.86%	-0.46%	-3.63%	5.10%
	-0.4	0.39%	-14.48%	4.55%	-0.79%	-15.50%	-2.03%	-0.70%	-15.18%	0.22%
200	0	-0.60%	7.05%	0.18%	-0.64%	7.22%	-2.15%	-0.41%	7.01%	-0.12%
	0.4	-1.27%	15.20%	1.14%	-1.08%	15.56%	-1.57%	-2.34%	15.65%	-1.14%
	0.8	-0.95%	3.89%	4.51%	-0.74%	3.82%	3.69%	-0.77%	3.82%	-0.33%
	-0.8	0.24%	-10.79%	-0.52%	-0.18%	-2.58%	2.47%	-0.12%	-2.57%	2.85%
	-0.4	-0.29%	4.97%	0.40%	-0.31%	-11.00%	-2.77%	-0.41%	-10.89%	-1.48%
400	0	-0.51%	10.81%	0.04%	-0.33%	4.91%	1.96%	-0.47%	5.19%	-4.12%
	0.4	-0.32%	2.59%	4.07%	-1.08%	10.45%	3.65%	-0.88%	11.14%	-3.69%
	0.8	-0.18%	-2.58%	2.47%	-0.44%	2.76%	-2.00%	-0.51%	2.71%	-2.20%

Table C.8: Monte Carlo Results: New estimator for the total effect of x where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

			R^2=.1			R^2=.4			R^2=.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-0.8	0.03%	1.75%	0.86%	0.03%	0.82%	2.30%	0.01%	0.41%	-1.37%
	-0.4	0.06%	5.19%	-0.24%	0.10%	1.96%	-3.37%	-0.03%	0.78%	-1.35%
100	0	0.31%	6.83%	-0.41%	0.01%	3.04%	-0.32%	0.03%	1.25%	0.47%
	0.4	0.54%	10.65%	-6.05%	-0.09%	3.68%	0.54%	0.02%	1.72%	-3.80%
	0.8	-0.42%	15.14%	-3.26%	0.49%	6.63%	-3.37%	0.01%	3.65%	-0.61%
	-0.8	-0.04%	0.99%	-3.55%	0.01%	0.36%	1.06%	0.00%	0.18%	-2.97%
	-0.4	0.05%	2.35%	-0.52%	-0.04%	1.07%	0.62%	0.00%	0.44%	1.27%
200	0	0.03%	3.88%	-4.00%	-0.02%	1.69%	-3.64%	-0.02%	0.67%	-1.99%
	0.4	0.08%	6.04%	-2.92%	0.23%	2.36%	-3.29%	0.01%	0.98%	-3.18%
	0.8	0.74%	9.68%	-0.55%	0.06%	3.31%	0.83%	-0.08%	1.45%	-1.87%
	-0.8	-0.01%	0.62%	-0.83%	0.00%	0.27%	-0.61%	0.00%	0.13%	-4.55%
	-0.4	0.06%	1.47%	0.17%	0.02%	0.63%	-0.59%	0.01%	0.24%	0.90%
400	0	-0.02%	2.45%	-0.62%	-0.01%	1.01%	-0.85%	0.00%	0.41%	-0.44%
	0.4	-0.03%	3.26%	3.12%	-0.01%	1.53%	-5.19%	0.00%	0.55%	1.00%
	0.8	0.03%	5.09%	1.48%	-0.06%	2.34%	-1.24%	0.01%	0.93%	-0.70%