Koch, Matthias

Conference Paper

Spatial Filtering and Model Interpretation for Spatial Durbin Models

52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia

Provided in Cooperation with:
European Regional Science Association (ERSA)

Suggested Citation: Koch, Matthias (2012) : Spatial Filtering and Model Interpretation for Spatial Durbin Models, 52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:
https://hdl.handle.net/10419/120747

Spatial Filtering and Model Interpretation for Spatial Durbin Models

Matthias Koch
Institute for Economic Geography and GIScience, Vienna University of Economics and Business, Nordbergstrasse 15/4/A, 1090 Vienna, Austria
Tel: +43-1-31336-4202
Matthias.Koch@wu.ac.at

Abstract

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 desireable 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 tradeoff between the estimator bias and its efficiency, depending on the spectrum of the used spatial weight matrix.

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 approximatly 95% of the available eigenvectors. By using Monte Carlo Simulations we observe that the estimaton procedure has a lower (or equal) bias and smaller (or equal) sample variance as the corresponding Maximum Likelihood estimator based on normality.

1 Introduction

Spatial filtering is a very popular[1] alternative estimation method for spatial (autoregressive) models. These spatial autoregressive models assume

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

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) argues 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) correctly 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 form 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[2] 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

[2]see for example Getis and Griffith (2002)

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 previous 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 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.

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 eigenvaluematrix. A subset of the eigenvectors of $\bar{\mathbf{D}}_n$ is denoted by $\mathbf{D}_n$ and the corresponding eigenvectormatrix by $\mathbf{\Lambda}_q$. The number of

columns of $\mathbf{D}_n$ is denoted by q. The i^{th} diagonal element of the eigenvaluematrix is denoted by λ_i. The Operator $tr()$ applied on 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:

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

In (1) $\mathbf{X}_n$ represents the n by k matrix of (finite) explanatory variables where $rank(\mathbf{X}_n) = k$. The parameters in (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[3]. The SDM incorporates various representations of spatial DGPs like the Spatial Autoregressive Model and the Spatial Error Model[4]. Throughout this paper we maintain additionally the following (central) assumptions:

1. $\mathbf{W}_n$ can be seen as deterministic and is normalized such that the absolute maximum eigenvalue is smaller or equal one
2. $\mathbf{W}_n$ can be approximated by $\mathbf{W}_n \approx \mathbf{D}_n \mathbf{\Lambda}_q \mathbf{D}_n'$
3. $\rho_o \in (-1, 1)$
4. $E[\mathbf{X}_n' \epsilon_n] = 0$

Due to assumptions (1) and (3) we can solve the DGP for $\mathbf{Y}_n$ and end up with Eq. (2):

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

[3] 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.

[4] For more details to the assumptions and properties of the data generating process stated in (1), see Elhorst (2010)

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.. (3) as

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

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 $\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}_n \left(\mathbf{D}_n'\mathbf{D}_n\right)^{-1} \mathbf{D}_n'$. 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 (3) is reduced to a simple linear model where ordinary least squares can be applied

$$\mathbf{M}_D\mathbf{Y}_n \approx \mathbf{M}_D\mathbf{X}_n\beta_0 + \mathbf{M}_D\epsilon_n. \quad (4)$$

However it is not clear how many eigenvectors are necessary for a reasonable model approximation. Given assumption (3) we can write $\mathbf{A}_n$ as a Neumann Series $(\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1} \rho_0 \mathbf{W}_n = \sum_{k=1}^{\infty} \rho_0^k \mathbf{W}_n^k$ and therefore the eigenvectors corresponding to the absolute largest[5] 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). If the approximation given in Eq. (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 (4) and (1). 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.

[5] 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.

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. 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.

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 approximation 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 (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 in (5):

$$\bar{\mathbf{Y}}_n = \rho_o \mathbf{M}_x \mathbf{W}_n \mathbf{Y}_n + \bar{\epsilon}_n \approx \mathbf{M}_x \mathbf{D}_n \theta_0 + \bar{\epsilon}_n \text{ where } \mathbf{M}_x = \mathbf{I}_n - \mathbf{Z}_n \left(\mathbf{Z}_n' \mathbf{Z}_n\right)^{-1} \mathbf{Z}_n',$$
$$\bar{\epsilon}_n = \mathbf{M}_x \epsilon_n,\ \bar{\mathbf{Y}}_n = \mathbf{M}_x \mathbf{Y}_n,\ \theta_0 = \rho_o \mathbf{\Lambda}_q \mathbf{D}_n' \mathbf{Y}_n \text{ and } \mathbf{Z}_n = [\mathbf{X}_n, \mathbf{W}_n \mathbf{X}_n] \quad (5)$$

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.. (5) yields: $\rho_o \widehat{\mathbf{M}_x \mathbf{W}_n} \mathbf{Y}_n = \mathbf{M}_x \mathbf{D}_n \widehat{\theta}$ where $\widehat{\theta} = \left(\mathbf{D}_n' \mathbf{M}_x \mathbf{D}_n\right)^{-1} \mathbf{D}_n' \bar{\mathbf{Y}}_n$. 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 1 in the Appendix derives the asymptotically expected value of $\widehat{\rho}_1$ given that $\mathbf{M}_x \mathbf{W}_n = \mathbf{D}_n \mathbf{\Lambda}_q \mathbf{D}_n'$: $\lim_{n\to\infty} E\left[\hat{\rho}_1\right] = \rho_0 + \frac{\sigma_0^2 tr\left(\mathbf{S}_n' \mathbf{W}_n' \mathbf{M}_x \overline{\mathbf{DD}}_n\right)}{\left(\sigma_0^2 + \sigma_\mu^2\right) tr\left(\mathbf{S}_n' \mathbf{W}_n' \mathbf{M}_x \mathbf{W}_n \mathbf{S}_n\right)}$ 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 $\mathbf{M}_x\mathbf{W}_n = \mathbf{D}_n\mathbf{\Lambda}_n\mathbf{D}_n^{-1}$:

$$\hat{\rho} = \arg \min_{\rho\in(-1,1)} \left| \widehat{\rho}_1 - \rho - \frac{\sigma_0^2 tr\left(\left(\mathbf{I}_n - \rho\mathbf{W}_n'\right)^{-1}\mathbf{W}_n'\mathbf{M}_x\overline{\mathbf{DD}}_n\right)}{\mathbf{Y}_n'\mathbf{W}_n'\mathbf{M}_x\mathbf{W}_n\mathbf{Y}_n} \right| \tag{6}$$

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 in (6) is a computational simple nonlinear minimization problem. In order to optimize (6) 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.. (7):

$$\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\end{pmatrix} = \left(\mathbf{Z}_n'\mathbf{Z}_n\right)^{-1}\mathbf{Z}_n'\left(\mathbf{I}_n - \hat{\rho}\mathbf{W}_n\right)\mathbf{Y}_n \tag{7}$$

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 an abort 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. (8) to (10):

$$direct\ effect_k = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial y_i}{\partial x_{i,k}} \tag{8}$$

$$indirect\ effect_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}\frac{\partial y_i}{\partial x_{j,k}} \tag{9}$$

$$total\ effect_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y_i}{\partial x_{j,k}} \tag{10}$$

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. (10) 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.. (10) 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 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 suggest 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. (8) - Eq. (10). 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 Monte Carlo Design and used performance measures

In our Monte Carlo study, the data generating process is given by (11)

$$\begin{aligned}\mathbf{Y}_n = \rho_o\mathbf{W}_n\mathbf{Y}_n + \mathbf{X}_n\beta_0 + \mathbf{W}_n\mathbf{X}_n\gamma_0 + \epsilon_n \\ \text{where } \epsilon_i \sim i.i.N(0,\sigma_0^2) \text{ and } \mathbf{X}_n \sim (\mathbf{I}_n - \lambda_0\mathbf{W}_n)^{-1}\nu_n\end{aligned} \tag{11}$$

where $\mathbf{X}_n$ has one column. We follow Le Sage et al (2011) by intorducing 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 (11) 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 atuocorrelation in the regrossor introduces problems for the estimation ML estimation method and also has to be considered realistic for typically housing data regressors.

Each configuration of Eq. (11) is simulated $Trials_{MC} = 1000$ times. Since the error term in Eq. (11) 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. (12) as proxy for the information to noise ratio, is fixed across different values of ρ_0.

$$R^2 = \frac{Var\left[\left(\mathbf{I}_n - \rho_0\mathbf{W}_n\right)^{-1}\epsilon_n\right]}{Var\left[\mathbf{Y}_n\right]} \tag{12}$$

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 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[6] 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 (1).

Comparing these two histograms in Figure (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

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

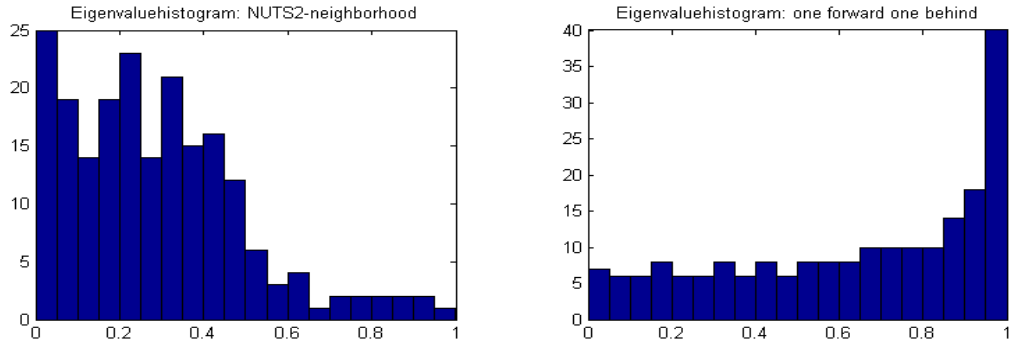

Figure 1: (Figure 1): Eigenvalue-histogram of different spatial weigth matrices

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[7] $(\hat{\sigma}^2_{up-1} - \hat{\sigma}^2_{up})/\hat{\sigma}^2_{up} \leq 0.01$. The Monte Carlo experiment is programed in MATLAB and as comparison we use the sdm-function based on maximum likelihood estimation, where the function is provided by the MATLAB spatial econometrics toolbox[8] 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 sdm 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,

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

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

which is given in Eq. (13)

$$Bias(\hat{\eta}) = \begin{cases} \frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \frac{\hat{\eta}_i - \eta_0}{\eta_0} 100 \text{ if } \eta_0 \neq 0 \\ \frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \hat{\eta}_i \text{ if } \eta_0 = 0 \end{cases} \tag{13}$$

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

$$RMSE(\hat{\eta}) = \begin{cases} \sqrt{\frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \left(\frac{\hat{\eta}_i - \eta_0}{\eta_0}\right)^2} 100 \text{ if } \eta_0 \neq 0 \\ \sqrt{\frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \hat{\eta}_i^2} \text{ if } \eta_0 = 0 \end{cases} \tag{14}$$

Eq. (14) 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 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. (14) 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 analog. This performance measure given in Eq. (15) is similar to the measure given in Eq. (13) since both measure the bias of a parameter estimate where Eq. (13) concerns the estimators first moment and Eq. (15) the estimator's second moment estimate.

$$Bias^2(\hat{\eta}) = \frac{1}{Trials_{MC}} \sum_{i=1}^{Trials_{MC}} \frac{\widehat{std[\hat{\eta}_i]} - std_0[\hat{\eta}_i]}{std_0[\hat{\eta}_i]} 100 \tag{15}$$

In Eq. (15) $\widehat{std[\hat{\eta}_i]}$ 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 analog. Hence Eq. (15) can be interpreted as the percentage bias of the estimated standard deviation of the corresponding parameter estimator.

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 divided into 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 $\lambda_0 = 0.00$, $\lambda_0 = 0.99$. The tables (1)-(4) correspond to the total effects where $\gamma_0 = \beta_0$ and λ_0 is either 0.00 or 0.99 the tables (4)-(12) are given in the Appendix. Since the interpretation of Spatial Durban 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 1: Monte Carlo Results Results Maximum Likelihood for the total effect of x where γ_0=1 and λ_0=0.00

	R^2		.1			.4			.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
	-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%
100	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%
	-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%
200	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%
	-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%
400	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%

Table 2: Monte Carlo Results Results new estimator for the total effect of x where γ_0=1 and λ_0=0.00

	R^2		.1			.4			.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-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%
	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%
200	-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%
	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%
400	-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%
	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%

Comparing tables (1) and (2) we find that both estimators perform equally well regarding bias, RMSEs and Bias^2. Addittionally the performance measures are similar for the MC-case $\gamma_0 = -\rho_0\beta_0$ and $\lambda_0 = 0.00$ given in the Appendix in table (). 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.

Table 3: Monte Carlo Results Results new estimator for the total effect of x where γ_0=1 and λ_0=0.99

	R^2		.1			.4			.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-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%
	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%
200	-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%
	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%
400	-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%
	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%

If the explanatory variable exhibit a high spatial autocorrelation, then the estimation- performance for both estimators regarding the first

Table 4: Monte Carlo Results Results Maximum Likelihood for the total effect of x where γ_0=1 and λ_0=0.99

	R^2	.1			.4			.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%

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) 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 (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 (1)-(12) indicate a similar performance of the proposed estimator compared to the maximum likelihood estimator.

Like Pace et. al. we suggest to use the ARPACK (Lehoucq, Sorensen, and Yang, (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 4 step estimation method unlike 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[9] $round(0.5(n-2k))$. 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[10] 10000 for weight matrices with similar spectrum-properties like the ones reported in Pace et. al. (2011) or Fischer et. al. (2007). Nevertheless we have to conclude that if we want to have a proper model interpretation based on the proposed spatial filtering method, spatial filtering looses the Pace et. al. (2011) reported feasibility for very large data sets.

6 Conclusion and Summary

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 the, by the estimated parameters implied direct/indirect and total effects, 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,

[9] 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.

[10] $n = 1600$ requires with simple programming approximately 15 seconds

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. This contrasts findings for spatial filtering suggested by 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 $n = 1000000$ as feasible for their spatial filtering approach, while our proposed four step estimator can only handle medium sized sample sizes, like $n = 10000$. 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.

A Appendix

A.1 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 $\check{\mathbf{A}}_n$

A.1.1 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. $E\left[\left(\epsilon_n'\mathbf{A}_n\epsilon_n\right)^2\right] = \left(\mu_{\epsilon,4} - 3\sigma^4\right)\sum_i a_{ii}^2 + \sigma^4\left(tr\left(\mathbf{A}_n\right)^2 + tr\left(\mathbf{A}_n^2\right) + tr\left(\mathbf{A}_n\mathbf{A}_n'\right)\right)$

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

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

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

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

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

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

A.1.2 Proof for Lemmas

Proof for Lemma 1,2 and 7. see Lee (1999) ■

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

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

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

Proof for Lemma 9. $Cov\left[\left(y_n'\mathbf{A}_ny_n\right),y_n'\mathbf{B}_n\epsilon_n\right] = E[\left(y_n'\check{\mathbf{A}}_ny_n\ -E\left[y_n'\check{\mathbf{A}}_ny_n\right]\right)$

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

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

A.2 Theorems

Theorem 1 *The asymptotic expected value for $\hat{\rho}_1$ as it is given in Step 2 is given by:* $\lim_{n\to\infty}E\left[\hat{\rho}_1\right] = \rho_0+\ \frac{\sigma_0^2tr\left(\mathbf{S}_n'\mathbf{W}_n'\mathbf{M}_x\overline{\mathbf{DD}}_n\right)}{\left(\sigma_0^2+\sigma_\mu^2\right)tr(\mathbf{S}_n'\mathbf{W}_n'\mathbf{M}_x\mathbf{W}_n\mathbf{S}_n)}$

Theorem 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:* $\mu_\zeta = \sigma^2 tr(\bar{\mathbf{A}}_n)$, $\mu_\vartheta = \sigma^2 tr(\check{\mathbf{B}}_n) + \mu'_n\check{\mathbf{B}}_n\mu_n$,

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

$Var\left(\vartheta\right) = \left(\mu_4 - 3\sigma^4\right)\sum_i \check{b}_{ii}^2 + 2\sigma^4 tr\left(\check{\mathbf{B}}_n^2\right) + 4\sigma^2\mu'_n\check{\mathbf{B}}_n\check{\mathbf{B}}_n\mu_n + 2\mu_3\sum_{ij}\check{b}_{ii}\check{b}_{ij}\mu_{jn}$
$-4\sigma^2 tr(\check{\mathbf{B}}_n)\mu'_n\check{\mathbf{B}}_n\mu_n$ *and* $Cov(\zeta,\vartheta) = \mu'_n\check{\mathbf{A}}_n\mu_n\sigma^4 tr(\bar{\mathbf{B}}_n) + 2\mu_3\sum_{ij}\bar{b}_{ii}\check{a}_{ij}\mu_{jn} +$

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

A.2.1 Proof for Theorems

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

Proof for Theorem 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 $Cov(\zeta,\vartheta) =$
$\mu'_n\check{\mathbf{A}}_n\mu_n\sigma^4 tr(\bar{\mathbf{B}}_n) + 2\mu_3\sum_{ij}\bar{b}_{ii}\check{a}_{ij}\mu_{jn} + 2\sigma^2\sum_l\sum_j\left(\check{a}_{lj}\mu_{jn}\right)\sum_i\left(\mu_{in}\bar{b}_{il}\right) +$

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

A.3 Additional Monte Carlo tables

Table 5: Monte Carlo Results Results Maximum Likelihood for ρ_0 where γ_0=1 and λ_0=0.99

	R^2		.1			.4			.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 6: Monte Carlo Results Results new estimator for ρ_0 where γ_0=1 and λ_0=0.99

	R^2		.1			.4			.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 7: Monte Carlo Results Results Maximum Likelihood for ρ_0 where γ_0=1 and λ_0=0.00

	R^2	.1			.4			.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 8: Monte Carlo Results Results new estimator for ρ_0 where γ_0=1 and λ_0=0.00

	R^2	.1			.4			.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 9: Monte Carlo Results Results Maximum Likelihood for ρ_0 where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

	R^2	.1			.4			.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 10: Monte Carlo Results Results Maximum Likelihood for the total effect of x where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

	R^2	.1			.4			.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-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%
	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%
200	-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%
	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%
400	-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%
	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 11: Monte Carlo Results Results new estimator for ρ_0 where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

	R^2		.1			.4			.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-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%
	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%
200	-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%
	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%
400	-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%
	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 12: Monte Carlo Results Results new estimator for the total effect of x where γ_0=-$\rho_0\beta_0$ and λ_0=0.00

	R^2		.1			.4			.8	
	ρ_0	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2	Bias	RMSE	Bias^2
100	-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%
	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%
200	-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%
	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%
400	-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%
	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%

References

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

[George Casella, 2001] Roger L. Berger George Casella. *Statistical Inference*. Duxbury Press, second edition, 2001.

[Getis, 2002] Daniel Griffith Getis, Arthur. Comparative spatial filtering in regression analysis. *Geographical Analysis*, 34 (2):130–140, 2002.

[Greene, 1997] W. Greene. *Econometric Analysis*. Upper Saddle River: New Jersey: PrenticeHall, third edition, 1997.

[Griffith, 2000] Daniel Griffith. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, 2:141–156, 2000.

[Griffith, 2002] Daniel Griffith. A spatial filtering specification for the auto-poisson model. *Statistics and Probability Letters*, 58:245–251, 2002.

[Griffith, 2003] Daniel Griffith. *Spatial Autocorrelation and Spatial Filtering*. Berlin: Springer Verlag, 2003.

[J. LeSage, 2009] R.K. Pace J. LeSage. *An introduction to spatial econometrics*. Tylor & Francis Group, Boca Raton/London/New York, 2009.

[James P. LeSage, 2011] Shuang Zhu James P. LeSage, R. Kelley Pace. Spatial dependence in regressors and its effect on estimator performance. *working paper*, 2011.

[Jesús Crespo Cuaresma, 2009] Martin Feldkircher Jesús Crespo Cuaresma, Gernot Doppelhofer. The determinants of economic growth in european regions. *working paper*, 2009.

[Jesús Crespo Cuaresma, 2010] Martin Feldkircher Jesús Crespo Cuaresma. Spatial filtering, model uncertainty and the speed of income convergence in europe. *working paper*, 2010.

[Lee, 1999] L. F. Lee. Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models. *Manuscript*, 1999.

[Manfred M. Fischer, 2008a] Daniel A. Griffith Manfred M. Fischer. 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, 2008.

[Manfred M. Fischer, 2008b] Martin Reismann Manfred M. Fischer, Thomas Scherngell. Knowledge spillovers and total factor productivity: Evidence using a spatial panel data model. *Geographical Analysis*, 41:204–220, 2008.

[Michael Tiefelsdorf, 2007] Daniel A Griffith Michael Tiefelsdorf. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, 39:1193–1221, 2007.

[R. Kelley Pace, 2011] Shuang Zhu R. Kelley Pace, James P. LeSage. Interpretation and computation of estimates from regression models using spatial filtering. *working paper*, 2011.

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