Spatial Econometrics and the Lasso Estimator: Theory and Applications

Achim Ahrens

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Social Sciences

April 6, 2017

Supervised by

Prof. Mark Schaffer and Prof. Arnab Bhattacharjee

Abstract

This thesis links two topics of empirical economics: spatial econometrics and the Lasso estimator. Spatial econometrics is concerned with methods and models accounting for interaction effects between units. The Lasso estimator is a regularisation technique that allows for simultaneous variable selection and estimation in a high-dimensional setting where the number of parameters may exceed the sample size. Three applied and theoretical articles are presented that demonstrate how spatial econometric research can benefit from high-dimensional methods and, specifically, the Lasso. The introduction in Chapter 1 presents a literature review of both fields and discusses the connections between the two topics. Chapter 2 examines the effect of economic growth on civil conflicts in Africa. The Lasso estimator is employed to generate instrumental variables, which account for non-linearity and spatial heterogeneity. The theoretical contribution in Chapter 3 proposes a two-step Lasso estimator that can consistently estimate the spatial weights matrix in a spatial autoregressive panel model. Chapter 4 is an application to the US housing market. A Lasso-based estimation method is considered that controls for spatial effects in a spatial error-correction model. Chapter 5 provides concluding remarks.

To my supportive parents,

and my two lovely sisters.

Acknowledgements

First of all, I would like to thank my supervisors, Mark Schaffer and Arnab Bhattacharjee, for their help during my PhD studies. They have opened up many opportunities and have been supportive throughout. Most importantly, I have had the pleasure of many *SMWS* visits without which the PhD would have been less enjoyable.

I am indebted to Prof. Dr. Robert Jung of the University of Hohenheim who introduced me to the field of econometrics. I thank many researchers, discussants, seminar and conference participants who I had the pleasure of talking to. These include Christopher Adam, Raul Caruso, Kristian Skrede Gleditsch, Stephan Heblich and Sean Holly, as well as several anonymous referees. I am thankful to Alexandre Belloni of Duke University whose email concerning Lemma 7 in Belloni et al. (2012) was of great help. On the financial side, I acknowledge support from an ESRC postgraduate scholarship which made this thesis possible.

I would like to thank my fellow PhD students and friends that I have made in Edinburgh. Particularly, Vana Anastasiadou (for many *Black Cat* visits), Manuel Andersch (for countless memorable study group meetings), Antonio Carvalho (for introducing me to Stochastic Frontier Analysis and Chinese coal), Jan Ditzen (for being helpful at any time), Erkal Ersoy (for being my boss), Marco Lorusso (for the weekly stress-relieving Squash match), Kyle McNabb (for his famous, critically acclaimed Heriot-Watt campus tour), Mark Wang (for inspiring chats about econometrics), and Fran Watson (for her motivational quotes). I also thank all members of the Heriot-Watt 12 o'clock lunch group for the welcome daily distraction.

I would also like to mention the members of the group *P.S.* (Felix Herrmann, Tim Hoffmann, Nils Knofius, Rico Kraft, Simon Kunrath, Simon Reimers), the former tenants of *Teichstraße 73, Erfurt* (Dirk Lampe, Lars Zeigermann) as well as Felix Bender and Magnus Wurm for being great friends.

Lastly, I would like to thank Gerdis Marquardt for many years of support.

HERIOT WATT UNIVERSITY

ACADEMIC REGISTRY
Research Thesis Submission

Name:	Achim Ahrens		
School:	School of Social Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	PhD in Economics

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	

Submission

Submitted By *(name in capitals)*:	ACHIM AHRENS
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals)*:			
Method of Submission *(Handed in to SSC; posted through internal/external mail)*:			
E-thesis Submitted (***mandatory for final theses***)			
Signature:		Date:	

Please note this form should be bound into the submitted thesis.
Academic Registry/Version (1) August 2016

Contents

List of Figures

List of Tables

List of Abbreviations

i.i.d. identically and independently distributed.

LASSO Least Absolute Shrinkage and Selection Operator.

2SLS Two-stage Least Squares.

ADF Augmented Dickey-Fuller.

AIC Akaike Information Criterion.

ATI Average Total Impact.

BIC Bayesian Information Criterion.

CADF Cross-sectionally Augmented ADF.

CCE Common Correlated Effects.

CCEMG Common Correlated Effects Mean Group.

CCEP Common Correlated Effects Pooled.

CIPS Cross-sectionally Augmented IPS.

CV Cross-validation.

DSGE Dynamic Stochastic General Equilibrium.

ECM Error-correction Model.

FE Fixed Effects.

FGMM Focused Generalised Methods of Moments.

FHFA Federal Housing Finance Agency.

FMHPI FreddieMac House Price Index.

GDP Gross Domestic Product.

GMM Generalised Methods of Moments.

GS2SLS Generalised Spatial Two-stage Least Squares.

GWR Geographically Weighted Regression.

HPY Holly, Pesaran, and Yamagata (2010).

IMF International Monetary Fund.

IV Instrumental Variables.

MG Mean Group.

ML Maximum Likelihood.

MSA Metropolitan Statistical Areas.

MSPE Mean-squared Prediction Error.

OLS Ordinary Least Squares.

PDSE Post-double-selection Estimator.

QML Quasi-Maximum Likelihood.

RE Random Effects.

RSS Residual Sum of Squares.

SAR Spatial Autoregressive Model, *also* Spatial Lag Model.

SARAR Spatial Autoregressive Model with Spatial Autoregressive Disturbances.

SDM Spatial Durbin Model.

SEM Spatial Error Model.

SLX Spatial model with spatially lagged exogenous regressors.

SPDP Spatial Dynamic Panel Data.

STIV Self-Tuning Instrumental Variable.

TSL Two-step Lasso.

TSPL Two-step Post-Lasso.

UK United Kingdom.

US United States.

VAR Vector Autoregressive.

Chapter 1

Introduction

...estimating is what you do when you do not know...

– Sherman Kent[1]

This thesis links two topics of empirical sciences: spatial econometrics and the Lasso estimator. Related to these two topics are two major trends that have shaped the empirical sciences in the past years. The first ongoing trend is the increasing demand for statistical methods for analysing complex data sets, which have become more and more prevalent in the digital age. The fields of statistical learning, or machine learning, provides a comprehensive set of methods that often go beyond the regression-based toolbox of econometrics. These methods are not only of scientific interest, but have already a profound impact on modern life. Autopilot systems, spam filters, handwriting recognition and product recommendations, as for example used by Amazon and Netflix, are only a few examples that demonstrate the versatility and capability of statistical learning techniques. In a recent survey article, Google's Chief Economist Hal Varian (2014) points out that machine learning focuses almost exclusively on prediction, while econometrics puts a stronger emphasis on causal inference. Despite, or because of, the different focus, Varian (2014) calls for closer collaboration between machine learning and econometrics.

One of the most popular methods in statistical learning is the Lasso estimator.

[1] Cited in Tetlock and Gardner, 2015.

The Lasso is a regularisation technique which allows for variable selection and estimation of high-dimensional data. High-dimensionality refers to situations in which the number of variables is large relative to the number of observations. The Lasso and statistical learning have only recently started to attract more attention in economics and econometrics. The most prominent example is the seminal work of the research group surrounding Alexandre Belloni (Duke University), Victor Chernozhukov (MIT) and Christian Hansen (University of Chicago), who demonstrate that high-dimensionality is a common phenomenon in economics when taking uncertainty about the correct model specification into account. Indeed, many influential empirical studies have been criticized for a selective choice of control or instrumental variables (Angrist and Krueger, 1991; Donohue and Levitt, 2001; Miguel, Satyanath, and Sergenti, 2004, to name a few). Variable selection techniques, such as the Lasso, offer a promising approach that facilitates the identification of robust model specifications. Most notably, Belloni et al. (2012) and Belloni, Chernozhukov, and Hansen (2014b) develop a framework for applying the Lasso to generate optimal instruments and to select control variables in a regression model, respectively.

The second trend central to this thesis is the increasing awareness of the importance of interdependence. While temporal dynamics have been an integral part of mainstream econometric and economic modelling for many decades—e.g. in financial econometrics and the study of business cycles—spatial, or cross-sectional, dependence has become a central topic in econometrics only more recently. The literature can be divided into two complementary approaches: common factor models and spatial econometric models. The common factor approach assumes that the error term follows unobserved common factors, such as unobserved global economic shocks, which exert heterogeneous effects on units (for a literature overview, see Sarafidis and Wansbeek, 2012). The Common Correlated Effects (CCE) estimator due to Pesaran (2006) is a prominent method, which allows to control for common factors under general assumptions.

The field of spatial econometrics deals with the development and application of econometric methods which account for interdependence and heterogeneity across space in a cross-section or panel data setting (Anselin, 2010). The centrepiece of spatial econometric models is the spatial weights matrix, which determines the strength of interactions between units. In contrast to the common factor approach, which tends to treat cross-sectional dependence as nuisance, spatial econometrics aims at modelling interaction effects explicitly.

In this thesis, I demonstrate how quantitative research in economics can benefit from statistical learning methods. Specifically, I show that the Lasso's capability as a variable selection technique can be of great value in the analysis of spatial panel data. To this end, I present three essays covering applications and theoretical topics.

Chapter 2: Conflict in Africa. The application in Chapter 2 examines the relationship between weather conditions, economic shocks and civil conflicts in Africa. While most studies rely on country-level data sets, I exploit a panel data set of African first-order administrative units covering 1992-2010. Since sub-national Gross Domestic Product (GDP) for Africa is either unavailable or of poor quality, I utilise nighttime light data from satellites to predict economic growth at the sub-national level.

The identification strategy exploits climate variables as instruments in order to estimate the effect of economic growth shocks and spill-over effects on civil conflict. The estimation is made complicated by non-linearity in the economic growth-climate relationship, spatial heterogeneity across regions and weak identification. The Lasso estimator and the methodology of Belloni et al. (2012) successfully address these challenges by selecting suitable instruments from a large set of putative instruments which account for various forms of non-linearity and heterogeneity. Thus, the application in Chapter 2 demonstrates that the Lasso's ability of simultaneous estimation and model selection in a high-dimensional setting can be of use in Instrumental Variables (IV) regressions.

Chapter 3: Estimating the spatial weights matrix. The spatial weights matrix determines which units are linked in a network. The (i, j)th element of the spatial weights matrix, commonly denoted by w_{ij}, represents the effect of unit j on unit i. In theoretical studies, the elements of the spatial weights matrix are treated as known parameters, and the vast majority of empirical studies specify weights based on observable variables such as geographic distance or using trade data. Given the lack of guidance on how to specify the spatial weights matrix, the risk of misspecification is considerable and is likely to lead to misleading inference. Hence, the use of pre-specified weights is a shortcoming of the conflict application in Chapter 2, which in turn motivates the search for methods to estimate the spatial weights matrix.

In the 3$^{\text{rd}}$ Chapter, I propose a Lasso-based method for estimating the spatial weights matrix in a spatial autoregressive panel model. There are two major challenges when the spatial weights matrix is treated as unknown. First, the number of parameters may be larger than the total number of observations, implying that the model is not identified without further assumptions. One advantage of the Lasso estimator is that it can consistently estimate high-dimensional models where the number of parameters exceeds the sample size under the assumption of sparsity. In the context of spatial models, sparsity requires that many w_{ij} parameters are zero. Secondly, since unit i and j may affect each other simultaneously, the spatially lagged dependent variable constitutes an endogenous regressor. To account for endogeneity, a two-step Lasso estimator is developed that is inspired by Two-stage Least Squares (2SLS) and allows for high-dimensionality.

Chapter 4: Spatial analysis of the US housing market. The issue of unknown spatial weights is also the central theme of Chapter 4. The chapter is concerned with a spatial panel model that includes spatially lagged exogenous regressors. Hence, the local outcome is not only allowed to be affected by local exogenous regressors, but does also depend on regressors observed at other locations. To control for these non-local spatial effects, the Post-double-selection Estimator (PDSE) due to Belloni, Chernozhukov, and Hansen (2014b) is considered, which exploits the Lasso to identify confounding factors. The PDSE of the local, low-dimensional parameters is consistent and asymptotically normal in the presence of confounding spatial effects.

The estimation method is applied to the US housing market. Based on the established co-integration relationship between real house price and real per capita income, the PDSE is employed to examine the short-run spatio-temporal dynamics of house prices and income using a spatial Error-correction Model (ECM). The results are compared with the CCE estimator due to Pesaran (2006).

The remainder of this introductory chapter is organised as follows. The two subsequent sections provide a brief overview of the field of spatial econometrics and the Lasso estimator, respectively. The expositions are necessarily incomplete and focus on concepts relevant to this thesis. Section 1.3 discusses examples of how spatial econometrics and the Lasso estimator can be linked in econometric analyses and presents the outline of this thesis, its applications and theoretical contributions.

Notation. I briefly summarise the notation employed in this thesis. Vectors and matrices are written in bold, e.g. $\mathbf{a}$ and $\mathbf{A}$, while scalar terms are in italics (a_{ij}). The ℓ_q-norm of the vector $\mathbf{a} \in \mathbb{R}^M$ is defined as $\|\mathbf{a}\|_q = (\sum_{m=1}^M |a_m|^q)^{1/q}$ for $q = 1, 2$. The number of non-zero elements in $\mathbf{a}$ is denoted by $\|\mathbf{a}\|_0$, and $\|\mathbf{a}\|_\infty$ is the largest element of $\mathbf{a}$. I use $((\cdot))$ to denote the typical element of a matrix, e.g. $\mathbf{A} = ((a_{ij}))$. The Fröbenius norm of $\mathbf{A}$ is $\|\mathbf{A}\|_F = (\sum_{i,j} |a_{ij}|^2)^{1/2}$. The support operator is $\text{supp}(\mathbf{a}) = \{m \in \{1, \ldots, M\} : a_m \neq 0\}$. Let V be a set, then $\bar{V}$ is the complement of V. $\mathbf{a}_V$ is a vector with elements $a_m \mathbb{1}\{m \in V\}$ for $m = 1, \ldots, M$ where $\mathbb{1}\{\cdot\}$ is the indicator function. The typical element of $\mathbf{A}_V$ is $a_{ij}\mathbb{1}\{j \in V\}$. Lastly, I use $x_T \lesssim_{\mathrm{P}} z_T$ to denote $x_T = O_P(z_T)$ and $a \lesssim b$ to denote $a \leq cb$ for some constant $c > 0$.

1.1 Spatial econometrics

In an overview of the history of spatial econometrics, Anselin (2010, p. 3) states that the field "has moved from the margins to the mainstream of applied econometrics and social science methodology". Indeed, empirical applications of spatial econometric methods cover a wide range of social and economic topics, including criminology (Baller, Anselin, and Messner, 2001), petrol markets (Pinkse, Slade, and Brett, 2002), tax competition (Egger, Pfaffermayr, and Winner, 2005), housing markets (Holly, Pesaran, and Yamagata, 2010), agricultural economics (Eberhardt and Teal, 2013) and growth economics (Ertur and Koch, 2007).

Spatial dependence, also referred to as cross-sectional dependence, is a common pattern in socio-economic datasets. Spatial dependence arises from interactions between observable or unobservable variables across units of observation, such as countries, firms or individuals.

Figure 1.1 illustrates different forms of spatial interaction effects for the case of two units, i and j. Case (1) corresponds to the 'classical', non-spatial case in which the outcome of spatial unit i, denoted by y_i, is assumed to exclusively depend on a vector of local observables, $\mathbf{x}_i$, and local unobservables combined in ε_i. In case (2), the outcome at location i is directly affected by the outcome at location j, and vice versa. For instance, a price change at a petrol station may induce a price change at a nearby petrol station (Pinkse, Slade, and Brett, 2002). In case (3), y_i does not only depend on local observables, but also on observable variables at other locations. An example is given by Ertur and Koch (2007), who derive an augmented Solow model that accounts for technological interdependence across countries and suggests that economic growth is related to other countries' macroeconomic variables through knowledge transmission. Case (4) illustrates the case of spatial dependence among unobservable variables.

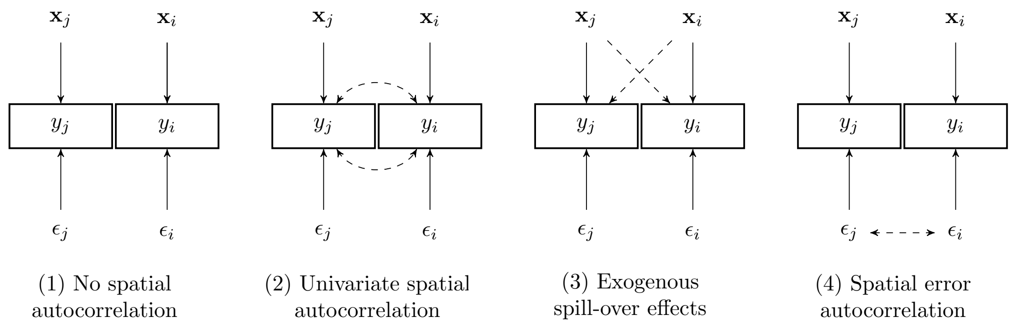


Figure 1.1: *Overview of spatial interaction effects (based on Baller, Anselin, and Messner, 2001). Case (1) corresponds to the 'classical' regression model which ignores interactions between units, while cases (2)-(4) illustrate spatial effects.*

1.1.1 An overview of spatial econometric models

The aim of spatial models is to control for and explicitly estimate spatial interaction effects. The Spatial Autoregressive Model (SAR), or Spatial Lag Model, and the Spatial Error Model (SEM) are arguably the most widely applied spatial models. First introduced by Whittle (1954), they were popularised by Cliff and Ord (1973, 1981) and, hence, are also referred to as Cliff-Ord type models.

Spatial autoregressive model

The SAR is given by

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{1.1}$$

where $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_n)'$ is the $n \times 1$ vector of the dependent variable, $\mathbf{X} = (\mathbf{x}'_1, \ldots, \mathbf{x}'_i, \ldots, \mathbf{x}'_n)'$ is a $n \times p$ full column rank matrix of regressors and $\boldsymbol{\beta}$ is the corresponding $p \times 1$ parameter vector.[2] The error term $\mathbf{e} = (\varepsilon_1, \ldots, \varepsilon_i, \ldots, \varepsilon_n)'$ is assumed to be independently distributed with $\mathrm{E}[\varepsilon_i|\mathbf{x}_i] = 0$ for all i.

The spatial weights matrix, commonly denoted by $\mathbf{W}$, is an n-dimensional square matrix which contains information about the spatial or socio-economic proximity between the spatial units under consideration. Specifically, the typical element w_{ij}

[2]The model in (1.1) is also referred to as the mixed regressive, spatial autoregressive model to emphasise the presence of exogenous regressors (Anselin, 1988; Lee, 2002). Accordingly, the model is termed the *pure* SAR if $\boldsymbol{\beta} = \mathbf{0}$ is assumed.

measures the strength of effects from unit j on unit i for $i \neq j$. The diagonal elements, w_{ii}, are set to zero to reflect that no unit can affect itself. $\mathbf{Wy}$ is referred to as the spatially lagged dependent variable and the spatial autoregressive parameter, λ, determines the strength of spill-over effects. The typical element of the vector $\mathbf{Wy}$ is

$$(\mathbf{Wy})_i = \sum_{j=1}^{n} w_{ij} y_j, \tag{1.2}$$

which illustrates that the spatial lag can be interpreted as a weighted average of outcomes at other locations.

Solving (1.1) for $\mathbf{y}$ yields the reduced form of the SAR

$$\mathbf{y} = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{e}, \tag{1.3}$$

where $\mathbf{I}_n$ is the identity matrix of dimension n and it is assumed that $(\mathbf{I}_n - \lambda\mathbf{W})$ is nonsingular. The reduced form in (1.3) reveals that the outcome y_i of the spatial unit i is not only affected by the explanatory variables in region i, but indirectly also by the explanatory variables in all other regions. In the same manner, y_i depends on unobservable shocks in all regions through the *spatial multiplier* $\mathbf{S} = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}$.

Lee (2002) shows that

$$\mathrm{E}[(\mathbf{Wy})'\mathbf{e}] = \sigma^2 \mathrm{tr}(\mathbf{W}(\mathbf{I}_n - \lambda\mathbf{W})^{-1}), \tag{1.4}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator. If $\lambda \neq 0$, the term $\mathrm{tr}(\mathbf{W}(\mathbf{I}_n - \lambda\mathbf{W})^{-1})$ is in general non-zero, which implies that the spatial lag is an endogenous regressor and the Ordinary Least Squares (OLS) estimator of the SAR model in (1.1) is inconsistent. The intuition for this result is that the model suffers from reverse causality, or simultaneity, since y_i may affect y_j, but at the same time y_j may also affect y_i.

An interesting exception is given if $\mathbf{W}$ is a strictly triangular matrix (Ord, 1975). For example, if $\mathbf{W}$ is a strictly upper triangular matrix, $w_{ij} = 0$ for $i \geq j$. Then,

$\text{tr}(\mathbf{W}(\mathbf{I}_n - \lambda\mathbf{W})^{-1}) = 0$ and OLS is consistent.[3] Intuitively, due to the triangular structure, there are only one-way affects and hence no reverse causality. However, the case of a triangular matrix is rather of theoretical than applied interest.

Spatial error model

The SEM is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{1.5}$$

$$\mathbf{u} = \rho\mathbf{W}\mathbf{u} + \mathbf{e}, \tag{1.6}$$

where ρ is the spatial autoregressive parameter, ε_i is identically and independently distributed (*i.i.d.*) and $(\mathbf{I}_n - \rho\mathbf{W})$ is nonsingular.[4] In this model, spatial effects arise from unobservable shocks operating across units. The parameter vector $\boldsymbol{\beta}$ can be consistently estimated by OLS as long as the regressors are weakly exogenous. However, the error term $\mathbf{u}$ is spatially dependent and heteroskedastic,

$$\text{E}[\mathbf{u}\mathbf{u}'|\mathbf{X}] = (\mathbf{I}_n - \rho\mathbf{W})^{-1}\text{E}[\mathbf{e}\mathbf{e}'|\mathbf{X}](\mathbf{I}_n - \rho\mathbf{W}')^{-1} \tag{1.7}$$

$$= \sigma_\varepsilon^2(\mathbf{I}_n - \rho\mathbf{W})^{-1}(\mathbf{I}_n - \rho\mathbf{W}')^{-1} \neq \sigma_\varepsilon^2\mathbf{I}_n, \tag{1.8}$$

which renders OLS inefficient (Anselin and Bera, 1998, p. 248-249).

The Manski model

A more general spatial cross-section model is discussed in Elhorst (2010a) and labelled as the Manski model (due to Manski, 1993):

$$\mathbf{y} = \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{u} \tag{1.9}$$

$$\mathbf{u} = \rho\mathbf{W}\mathbf{u} + \mathbf{e}, \tag{1.10}$$

[3] If the spatial weights matrix is strictly upper triangular, then $\mathbf{I}_n - \lambda\mathbf{W}$ is a upper triangular matrix with ones on the diagonal and the inverse, $(\mathbf{I}_n - \lambda\mathbf{W})^{-1}$, is also a upper triangular matrix. Furthermore, since $\mathbf{W}$ is a strictly upper triangular matrix and $(\mathbf{I}_n - \lambda\mathbf{W})^{-1}$ is a upper triangular matrix, the product of the two matrices has zeros on the diagonal. Thus, $\text{tr}(\mathbf{W}(\mathbf{I}_n - \lambda\mathbf{W})^{-1}) = 0$.

[4] An alternative, less common approach to modelling dependence in the errors is the spatial moving average process proposed by Haining (1978). Estimation by IV/GMM is discussed in Fingleton (2008).

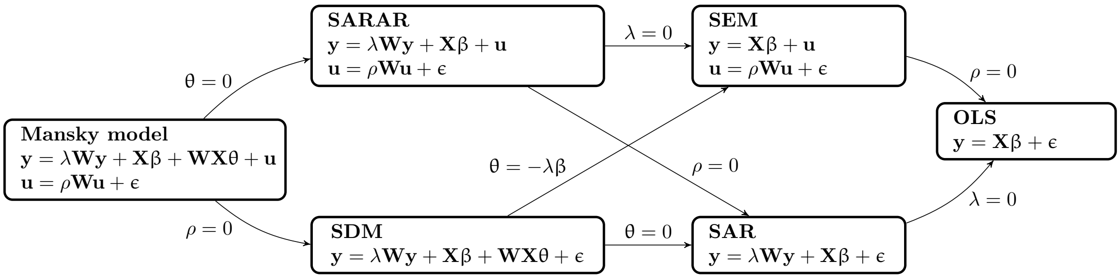


Figure 1.2: *Overview of spatial econometric cross-section models. The Manski model is the most general model. It includes a spatially lagged dependent variable, spatially lagged exogenous regressors and spatially autoregressive disturbances. The illustration is based on Elhorst (2010a).*

where $\boldsymbol{\theta}$ is the $p \times 1$ coefficient vector corresponding to $\mathbf{WX}$. The parameter space of λ and ρ is restricted by the assumption that $(\mathbf{I}_n - \lambda\mathbf{W})$ and $(\mathbf{I}_n - \rho\mathbf{W})$ are nonsingular (Elhorst, 2010a). Note that $\mathbf{y}$ as well as $\mathbf{u}$ follow spatial-autoregressive processes. It is possible to allow for the three weights matrices in (1.9)-(1.10) to differ as in, e.g., Kelejian and Prucha (1998). For simplicity, I assume the same weight matrices for the lag on $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{u}$ throughout, which is in line with most applied work.

The Manski model allows explicitly for the three types of spatial interaction effects illustrated in Figure 1.1. Specifically, y_i depends on *other* spatial units through the outcome of the dependent variable at other locations (i.e., y_j for some $j \neq i$), independent variables, $\mathbf{x}_j$, and unobservable characteristics, ϵ_j. However, Manski (1993) points out that not all parameters in (1.9) and (1.10) are separately identified. A Monte Carlo study by Elhorst (2010a) supports this finding. For the parameters to be identified, further restrictions have to be imposed.

Figure 1.2 provides an overview of spatial cross-section models which are special cases of the Manski model. The Spatial Durbin Model (SDM) was introduced in Anselin (1988) and estimation of the Spatial Autoregressive Model with Spatial Autoregressive Disturbances (SARAR) was considered in the seminal work of Kelejian and Prucha (1998). LeSage and Pace (2009) favour the SDM over the SARAR. They argue that falsely ignoring spatial dependence in the error (i.e. falsely setting $\rho = 0$) leads to inefficiency, but falsely omitting spatially lagged regressors (i.e., setting $\boldsymbol{\theta} = \mathbf{0}$) tends to have more severe effects as it results in a bias and inconsistency.

Relationship to time-series models

Spatial models, such as the SAR in (1.1), are closely related to time-series models. For example, setting the lower sub-diagonal of $\mathbf{W}$ equal to 1,

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & & \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \\ 0 & & & 1 & 0 \end{bmatrix}, \tag{1.11}$$

is equivalent to the autoregressive model of order 1,

$$y_i = \lambda y_{i-1} + \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i \quad \text{for } i = 2, \ldots, n. \tag{1.12}$$

However, in contrast to time-series data, spatial cross-section data has no natural ordering. Furthermore, in time-series analysis, the elements in the upper triangle of $\mathbf{W}$ are naturally zero, since future events cannot affect today's outcome, whereas when analysing spatial data all off-diagonal element are potentially non-zero. Thus, the question arises how the spatial weights matrix should be specified.

1.1.2 The spatial weights matrix

The majority of theoretical and applied work relies on the assumption that the spatial weights matrix is known. In empirical studies, the weights are specified *a-priori* on the basis of theoretical considerations using observable variables, such as geographic distance or trade flows. To understand the difficulty of estimating the spatial weights, it is instructive to write the SAR model as a system of equations:

$$
\begin{aligned}
y_1 &= 0 + w_{12}y_2 + w_{13}y_3 + \dots + w_{1n}y_n + \mathbf{x}_1'\boldsymbol{\beta} + \varepsilon_1 \\
y_2 &= w_{21}y_1 + 0 + w_{23}y_3 + \dots + w_{2n}y_n + \mathbf{x}_2'\boldsymbol{\beta} + \varepsilon_i \\
&\dots \\
y_n &= w_{n1}y_1 + w_{n2}y_2 + \dots + w_{n,n-1}y_{n-1} + 0 + \mathbf{x}_n'\boldsymbol{\beta} + \varepsilon_n
\end{aligned}
\tag{1.13}
$$

Here, I implicitly assume $\lambda = 1$, as the spatial autoregressive parameter and the spatial weights matrix are not separately identified (Bhattacharjee and Jensen-Butler, 2013). If the w_{ij}'s are treated as coefficients to be estimated for all $i \neq j$, there are $(n-1)n + p$ unknown parameters on the right-hand side, whereas the number of observations is n. The SAR model is, therefore, not identified if the spatial weights matrix is treated unknown. The result also applies to other spatial models.

To circumvent the identification problem, several specifications have been proposed in the literature. The most common specifications are:

- The spatial neighbour or contiguity matrix is defined such that

$$
w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are neighbours,} \\ 0 & \text{otherwise.} \end{cases} \tag{1.14}
$$

- The kth-order neighbour matrix is a generalisation of the first-order binary neighbour matrix defined in (1.14). For example, the second-order contiguity matrix is defined such that $w_{ij} = 1$ if area i is a neighbour of j or a neighbour of area j's neighbours, 0 otherwise. The matrix is defined accordingly for $k > 2$.
- The inverse distance matrix assumes that the weights are smaller, the farther two units are apart from each other. Specifically, $w_{ij} = d_{ij}^{-1}$ where d_{ij} typically denotes geographic distance, but could also denote a measure of social or economic distance.
- The q-nearest neighbour matrix is defined such that $w_{ij} = 1$ if unit j is among the q closest neighbours of unit i.

(a) *Map of counties in New Hampshire*

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	1	0	1	0	1	0
2	1	0	0	1	1	0	0	0	1	0
3	0	0	0	0	0	1	0	0	0	1
4	0	1	0	0	1	0	0	0	0	0
5	1	1	0	1	0	0	1	0	0	1
6	0	0	1	0	0	0	1	1	0	1
7	1	0	0	0	1	1	0	1	1	1
8	0	0	0	0	0	1	1	0	1	0
9	1	1	0	0	0	0	1	1	0	0
10	0	0	1	0	1	1	1	0	0	0

	1	2	3	4	5	6	7	8	9	10
1	0	.25	0	0	.25	0	.25	0	.25	0
2	.25	0	0	.25	.25	0	0	0	.25	0
3	0	0	0	0	0	.5	0	0	0	.5
4	0	.5	0	0	.5	0	0	0	0	0
5	.2	.2	0	.2	0	0	.2	0	0	.2
6	0	0	.25	0	0	0	.25	.25	0	.25
7	1/6	0	0	0	1/6	1/6	0	1/6	1/6	1/6
8	0	0	0	0	0	1/3	1/3	0	1/3	0
9	.25	.25	0	0	0	0	.25	.25	0	0
10	0	0	.25	0	.25	.25	.25	0	0	0

(b) *Non-standardised (left) and row-standardised (right) binary contiguity weights matrix for New Hampshire counties*

Figure 1.3: *Example of a binary contiguity matrix for counties in New Hampshire, which are pictured in (a). The row-standardisation guarantees that the entries in each row sum to one.*

- The distance cut-off matrix sets $w_{ij} = 1$ if the (geographic) distance between i and j is below a to-be-specified threshold, zero otherwise.

Prior to estimation, the spatial weights matrix is commonly row-standardised. The row-standardised spatial weights matrix is defined as

$$w_{ij}^{\star} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}} \quad \text{for} \quad i, j = 1, \ldots, n. \tag{1.15}$$

For illustration, Figure 1.3b shows the unstandardised and row-standardised spatial weights matrix corresponding to the map of countries in the US state of New Hampshire in Figure 1.3a. If w_{ij} is non-negative for all i and j, the row-normalisation

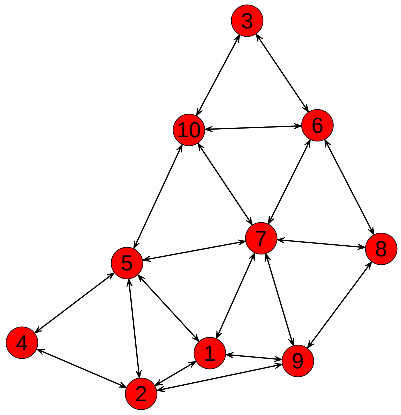


Figure 1.4: *The graph shows the spatial weights matrix in Figure 1.3b represented as network. The figure was created using the `network` package in R (Butts, 2008; Butts, 2015).*

ensures that all standardised weights, $w^{\star}_{ij}$, are between 0 and 1. A row-normalisation also provides that $\mathbf{I}_n - \lambda\mathbf{W}$ is nonsingular for $\lambda \in (-1, 1)$ (Kelejian and Prucha, 2010, Lemma 2). While the row-standardisation facilitates interpretation, it underlies the assumption that the strength of interaction effects is constant across units, which may not be appropriate in some settings. Figure 1.4 represents the spatial weights matrix for New Hampshire as a network system, highlighting the close link between spatial econometrics and network analysis.

A central condition regarding the spatial weights matrix required for asymptotic results was first suggested by Kelejian and Prucha (1998) and Kelejian and Prucha (1999). The condition states that the row and column sums of $\mathbf{W}$ and $(\mathbf{I}_n - \lambda\mathbf{W})^{-1}$ are bounded in absolute value. The row and column sums of a matrix $\mathbf{A}$ with typical element a_{ij} are said to be bounded uniformly in absolute value if there exists a finite constant c such that

$$\sum_{j=1}^{n} |a_{ij}| \leq c \quad \text{and} \quad \sum_{i=1}^{n} |a_{ij}| \leq c \quad \text{for all } n. \tag{1.16}$$

The assumption limits the the strength of interaction effects and the cross-sectional dependence (Kelejian and Prucha, 1998; Kelejian and Prucha, 1999). For example, this condition is satisfied for a binary contiguity matrix or for any row-standardised matrix for which each unit is only affected by a finite number of other units.

Since theory does often not provide sufficient guidance on how to specify the spatial weights matrix, critics have argued that the choice is often arbitrary and may substantially distort estimation results (Harris, Moffat, and Kravtsova, 2011). One approach is to select among a set of alternative specifications, guided by goodness-of-fit measures such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). Kelejian (2008) and Kelejian and Piras (2011) extend the classical Davidson and Mackinnon's J-test (1981) for non-nested models to the spatial setting. The test can be used to select between alternative models, including alternative spatial weights matrix specifications.[5]

An alternative approach is to construct or estimate the spatial weights from the data. Getis and Aldstadt (2004) suggest a data-driven binary contiguity matrix where contiguity is determined by the local clustering statistic $G^{\star}$ introduced in Ord and Getis (1995). Pinkse, Slade, and Brett (2002) consider the model

$$y_i = \sum_{j=1}^{n} w_{ij}(d_{ij})y_j + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \tag{1.17}$$

where the spatial weights are a function of an observable distance measure, d_{ij}. The authors propose a semi-parametric estimator, which they apply to the United States (US) wholesale gasoline market. Bhattacharjee and Jensen-Butler (2013) show that the spatial weights matrix is identified under the assumption of symmetry and propose an estimation procedure for panel data (see also Beenstock and Felsenstein, 2012). Bailey, Holly, and Pesaran (2016) develop a method for estimating the spatial weights matrix in a large T panel under the assumption of sparsity and symmetry. After extracting common factors, the authors use the Holm (1979) multiple testing procedure to identify non-zero correlations.

[5] See also Burridge and Fingleton (2010) and Monte Carlo simulations in Piras and Lozano-Gracia (2012).

1.1.3 Model interpretation and marginal effects

In the linear regression model,

$$y_i = \sum_{k=1}^{p} x_{ik}\beta_k + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1.18}$$

the marginal effects of an increase in x_{ik} on y_i is given by $\partial y_i / \partial x_{ik} = \beta_k$ and the effect of an increase in x_{jk} on the dependent variable in region i is zero, formally $\partial y_i / \partial x_{jk} = 0$ for $i \neq j$. The same holds for an SEM, but marginal effects are more complex if a spatially lagged dependent variable is present.

Consider the reduced form SDM as given in LeSage and Pace (2009, p. 34)

$$\mathbf{y} = \sum_{k=1}^{p} \mathbf{S}_k(\mathbf{W})\mathbf{X}_k + \mathbf{V}(\mathbf{W})\mathbf{e}, \tag{1.19}$$

where $\mathbf{X}_k$ denotes the kth column of $\mathbf{X}$, $\mathbf{V}(\mathbf{W})$ is the spatial multiplier $\mathbf{V}(\mathbf{W}) = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}$ and the $n \times n$ matrices $\mathbf{S}_k(\mathbf{W})$ are equal to $\mathbf{V}(\mathbf{W})(\mathbf{I}_n\beta_k + \mathbf{W}\theta_k)$ for $k = 1, \ldots, p$. Furthermore, let $\mathbf{V}(\mathbf{W})_i$ be the ith row of $\mathbf{V}(\mathbf{W})$ and $S_k(\mathbf{W})_{ij}$ the (i, j)th scalar element of $\mathbf{S}_k(\mathbf{W})$. This notation enables us to write the above model as (LeSage and Pace, 2009, p. 35)

$$y_i = \sum_{k=1}^{p} [S_k(\mathbf{W})_{i1}x_{1k} + S_k(\mathbf{W})_{i2}x_{2k} + \cdots + S_k(\mathbf{W})_{in}x_{nk}] + \mathbf{V}(\mathbf{W})_i\mathbf{e}. \tag{1.20}$$

It becomes evident that

$$\frac{\partial y_i}{\partial x_{ik}} = S_k(\mathbf{W})_{ii} \quad \text{and} \quad \frac{\partial y_i}{\partial x_{jk}} = S_k(\mathbf{W})_{ij}. \tag{1.21}$$

In general $\partial y_i / \partial x_{ik} \neq \beta_k$. This is because local marginal effect include 'feedback loops'. For example, spatial unit i may affect unit j which in turn may affect i, etc. The term $S_k(\mathbf{W})_{ij}$ depends on $\beta_k, \theta_k, \lambda$, but is independent of ρ. Thus, in a pure SEM model, the standard interpretations and marginal effects, as in the linear regression model, apply.

It follows from the discussion that marginal effects vary with i and j, and since $\mathbf{S}_k(\mathbf{W})$ is in general not symmetric, there are up to n^2 distinct marginal effects. To summarize inter-regional effects, LeSage and Pace (2009, p. 36-39) propose two summary statistics. The Average Total Impact (ATI) is given by

$$\text{ATI}_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_k(\mathbf{W})_{ij}. \tag{1.22}$$

To understand the rationale behind the ATI, notice that the sum of the ith row of $\mathbf{S}_k(\mathbf{W})$ represents the total impact on y_i when the kth explanatory variable increases in all other regions by one unit. Averaging across all unit-specific total impacts yields the ATI, which quantifies the average total impact if the kth explanatory variable changes by one in all regions. The Average Direct Impact (ADI) measures the average effect from a change of the kth explanatory variable in region i on y_i, and is given by

$$\text{ADI}_k = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial y_i}{\partial x_{ik}} = \frac{1}{n}\text{tr}(\mathbf{S}_k(\mathbf{W})). \tag{1.23}$$

1.1.4 Model estimation

As discussed above, OLS estimation of spatial models is inconsistent in the presence of a spatial lag and inefficient if the disturbances follow a spatial error structure. The two most popular estimation methods for spatial models are Maximum Likelihood (ML) and Generalised Methods of Moments (GMM).[6]

Maximum Likelihood

The ML estimator for spatial models was first considered by Ord (1975). ML relies on the assumption of *i.i.d.* and normally distributed innovations. If the innovations are non-Gaussian, the ML estimator is the Quasi-Maximum Likelihood (QML) estimator.

Following Lee (2004), I consider the SAR in (1.1), although the exposition indi-

[6]Aside from ML and GMM, Bayesian estimation of spatial models have attracted increasing attention, but are not discussed here. For an overview, see LeSage and Pace (2009).

rectly also applies to the SDM, as $\mathbf{X}$ may be replaced with $[\mathbf{X}, \mathbf{WX}]$. The maximisation procedure can also be easily extended to the case where $\rho \neq 0$ (e.g., Drukker, Prucha, and Raciborski, 2013b). The ML estimator of $\boldsymbol{\beta}$, λ and σ^2 is obtained by maximising the likelihood function $L(\boldsymbol{\beta}, \lambda, \sigma^2 | \mathbf{y}, \mathbf{X})$ which is the likelihood of observing $(\boldsymbol{\beta}', \lambda, \sigma^2)$ conditional on the sample.

To derive the likelihood function, the identity

$$L(\boldsymbol{\beta}, \lambda, \sigma^2 | \mathbf{y}, \mathbf{X}) \equiv f(\mathbf{y} | \boldsymbol{\beta}, \lambda, \sigma^2, \mathbf{X}) \tag{1.24}$$

is of central importance. The latter is the joint conditional distribution of $\mathbf{y}$ and can be expressed as (see, e.g., Johnston and DiNardo, 1997, p. 145)

$$f(\mathbf{y} | \boldsymbol{\beta}, \lambda, \sigma^2, \mathbf{X}) = f(\mathbf{e} | \boldsymbol{\beta}, \lambda, \sigma^2, \mathbf{X}) \left| \frac{\partial \mathbf{e}}{\partial \mathbf{y}} \right|, \tag{1.25}$$

where $| \cdot |$ is the determinant operator. In this case, $\mathbf{e} = (\mathbf{I}_n - \lambda \mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ and hence $\partial \mathbf{e} / \partial \mathbf{y} = (\mathbf{I}_n - \lambda \mathbf{W})$. Assuming that $\mathbf{e}$ is *i.i.d.* and normally distributed and taking the logarithm, the full log-likelihood function can be written as

$$\ln L(\boldsymbol{\beta}, \lambda, \sigma^2 | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |\mathbf{I}_n - \lambda \mathbf{W}| - \frac{\mathbf{e}'\mathbf{e}}{2\sigma^2}. \tag{1.26}$$

In order to simplify the maximisation problem, the log-likelihood function may be maximised ('concentrated') with respect to $\boldsymbol{\beta}$ and σ^2 (LeSage and Pace, 2009, p. 46-48). To illustrate the idea, suppose that the true value of λ is given by $\lambda^\star$. Then, the model can be rewritten as

$$\mathbf{y} - \lambda^\star \mathbf{W}\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \mathbf{e}, \tag{1.27}$$

which suggests to apply OLS to the "spatially filtered dependent variable" (Anselin and Bera, 1998, p. 256)

$$\tilde{\boldsymbol{\beta}}_{\mathrm{ML}}(\lambda^{\star}) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{I}_n - \lambda^{\star}\mathbf{W})\mathbf{y} \tag{1.28}$$

$$\tilde{\sigma}^2_{\mathrm{ML}}(\lambda^{\star}) = \frac{1}{n}(\mathbf{e}(\lambda^{\star})'\mathbf{e}(\lambda^{\star})). \tag{1.29}$$

Substituting $\tilde{\boldsymbol{\beta}}_{\mathrm{ML}}(\lambda^{\star})$ and $\tilde{\sigma}^2_{\mathrm{ML}}(\lambda^{\star})$ into the full log-likelihood function in (1.26) yields the concentrated log-likelihood function,

$$\ln L_c(\lambda|\mathbf{y}, \mathbf{X}) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\tilde{\sigma}^2_{\mathrm{ML}}(\lambda) + \ln|\mathbf{I}_n - \lambda\mathbf{W}|\,. \tag{1.30}$$

The ML estimator of λ, denoted by $\hat{\lambda}_{\mathrm{ML}}$, is defined as the scalar value that maximises $\ln L_c(\lambda|\mathbf{y}, \mathbf{X})$. By substituting the closed form expressions, the maximisation problem is simplified to a univariate problem, which leads to a computational advantage. Estimating $\hat{\lambda}_{\mathrm{ML}}$ allows us to obtain the ML estimators of $\boldsymbol{\beta}$ and σ^2, which are defined as $\hat{\boldsymbol{\beta}}_{\mathrm{ML}} \equiv \tilde{\boldsymbol{\beta}}_{\mathrm{ML}}(\hat{\lambda}_{\mathrm{ML}})$ and $\hat{\sigma}^2_{\mathrm{ML}} \equiv \tilde{\sigma}^2_{\mathrm{ML}}(\hat{\lambda}_{\mathrm{ML}})$, respectively.

Maximisation of the likelihood function requires computing the determinant of the Jacobian term at each step of the iteration, which is computational intensive for large n. For this reason, Ord (1975) suggest to make use of the relation $|\mathbf{I}_n - \lambda\mathbf{W})| = \prod_{i=1}^{n}(1 - \lambda\omega_i)$ where $\{\omega_i\}$ are the eigenvalues of $\mathbf{W}$. However, the computation of eigenvalues may be imprecise if n is large (Anselin, 2001, p. 325). Alternative solutions have been proposed. For further discussion, I refer to Anselin (2001, p. 325) and LeSage and Pace (2009, Ch. 4).

Lee (2004) show that the QML estimator is consistent and asymptotically normal if $\mathbf{e}$ is *i.i.d.*[7] If the assumption of *i.i.d.* errors is violated, the ML estimator is in general inconsistent. Arraiz et al. (2010) illustrate the inconsistency of QML in the presence of heteroskedasticity using Monte Carlo simulations.

[7] See Assumption 1-8 in Lee (2004) for a comprehensive set of assumptions.

Generalised methods of moments

ML estimation can be computationally intensive if n is large (Kelejian and Prucha, 1998; Kelejian and Prucha, 2010). Hence, one reason for the development of the IV/GMM approach were computational considerations (Elhorst, 2010a, p. 15). Kelejian and Prucha (1998) and Kelejian and Prucha (1999) develop an IV/GMM estimator for the SARAR under the assumption of homoskedastic disturbances. The estimation method is termed Generalized Spatial Two-Stage Least Squares (GS2SLS). However, Kelejian and Prucha (2010, p. 54) also point out that the homoskedasticity assumption is often not appropriate given that spatial units are usually heterogeneous with respect to size and other characteristics. For this reason, Kelejian and Prucha (2010) and Arraiz et al. (2010) generalize the estimation method to account for heteroskedasticity. The estimation method also allows to include spatial lags of exogenous variables (Arraiz et al., 2010, p. 594). Fingleton and Le Gallo (2008) and Drukker, Egger, and Prucha (2013) discuss the inclusion of endogenous regressors. Thus, the IV/GMM approach has three advantages over ML. First, it is computationally less demanding, secondly, it allows for heteroskedasticity and, thirdly, it provides a framework that easily allows to include endogenous regressors.

In the following exposition, I consider the SARAR model. For a full set of model assumption, I refer to Kelejian and Prucha (2010) and Arraiz et al. (2010). The SARAR, here reproduced for convenience, is given by

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \tag{1.31}$$

$$\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \mathbf{e}, \tag{1.32}$$

where $\mathbf{Z} = [\mathbf{X}, \mathbf{W}\mathbf{y}]$ and $\boldsymbol{\delta} = (\boldsymbol{\beta}', \lambda)'$. It is also assumed that $\mathrm{E}[\varepsilon_i|\mathbf{x}_i] = 0$, $\mathrm{E}[\varepsilon_i^2|\mathbf{x}_i] = \sigma^2(\mathbf{x}_i)$ for all i, and $\mathrm{E}[\varepsilon_i\varepsilon_j|\mathbf{x}_i] = 0$ for $i \neq j$. Thus, the model allows for non-Gaussian errors and heteroskedasticity, but independence across i is required. Estimation by IV or GMM requires the specification of an instrument matrix, which I denote by $\mathbf{H}$. Valid instruments need to satisfy two crucial conditions.

First, the instruments are required to be orthogonal to the error term (*exclusion restriction*), and, secondly, the instruments must be correlated with the regressors (*instrument relevance*). To identify an appropriate set of instruments, Kelejian and Prucha (1998) use the reduced form of the SARAR in (1.31)-(1.32) to derive the expected value of $\mathbf{y}$,

$$\mathrm{E}[\mathbf{y}|\mathbf{X}] = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} \tag{1.33}$$

$$\mathrm{E}[\mathbf{y}|\mathbf{X}] = (\mathbf{I}_n + \lambda\mathbf{W} + \lambda^2\mathbf{W}^2 + \dots)\mathbf{X}\boldsymbol{\beta}, \tag{1.34}$$

where the second equation assumes $|\lambda| < 1$. This suggests

$$\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \mathbf{W}^3\mathbf{X}, \dots, \mathbf{W}^q\mathbf{X}, \tag{1.35}$$

as valid instruments. It also follows that the GMM/IV approach requires $\boldsymbol{\beta} \neq \mathbf{0}$ unless alternative instruments are available. Note that, for identification, the instrument matrix $\mathbf{H}$ needs to consist of at least $p+1$ columns. On the other hand, to avoid instrument proliferation, q is typically set to 1 or 2 (Arraiz et al., 2010).

Based on Drukker, Prucha, and Raciborski (2013a) and Arraiz et al. (2010), I present the estimation procedure in four steps. Step 1 is to apply 2SLS to (1.31) while ignoring (1.32). The 2SLS estimator of $\boldsymbol{\delta}$ is

$$\tilde{\boldsymbol{\delta}}_{\mathrm{2SLS}} = (\mathbf{Z}'\mathbf{P_H}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P_H}\mathbf{y}, \tag{1.36}$$

where $\mathbf{P_H}$ is the projection matrix $\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$. Although the initial 2SLS estimator $\tilde{\boldsymbol{\delta}}_{\mathrm{2SLS}}$ is consistent, it ignores the error structure in (1.32) and, therefore, is not efficient unless $\rho = 0$.

For this reason, Step 2 applies inefficient GMM to obtain an estimator for ρ, say $\tilde{\rho}$, which is based on the 2SLS residuals, i.e., $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\tilde{\boldsymbol{\delta}}_{\mathrm{2SLS}}$. As proposed by Kelejian and Prucha (2010, p. 56), the GMM estimator exploits the following moment conditions

$$n^{-1}\mathrm{E}[(\mathbf{We})'(\mathbf{We})] = n^{-1}\mathrm{trace}\left\{\mathbf{W}[\mathrm{diag}_{i=1}^{n}(\mathrm{E}[\epsilon_i^2])]\mathbf{W}'\right\} \tag{1.37}$$

$$n^{-1}\mathrm{E}[(\mathbf{We})'\mathbf{e}] = 0. \tag{1.38}$$

Under the assumption of homoskedasticity, the additional condition $n^{-1}\mathrm{E}[\mathbf{e}'\mathbf{e}] = \sigma^2$ is added and the right-hand side of (1.37) simplifies to $\sigma^2 n^{-1}\mathrm{trace}[\mathbf{WW}']$, which yields the moment conditions proposed in Kelejian and Prucha (1999, p. 514). Replacing $\mathbf{e}$ with $\tilde{\mathbf{e}} = \tilde{\mathbf{u}} - \rho\mathbf{W}\tilde{\mathbf{u}}$ yields the sample moment condition of the initial and inefficient GMM estimator of ρ, denoted by $\tilde{\rho}$. Kelejian and Prucha (1998) and Kelejian and Prucha (2010) proof consistency of $\tilde{\rho}$ under homoskedasticity and heteroskedasticity, respectively.

In Step 3, the Generalised Spatial Two-stage Least Squares (GS2SLS) estimator is the 2SLS estimator applied to the Cochrane-Orcutt-type transformation of (1.31) and (1.32). This transformation is obtained by pre-multiplying (1.31) and (1.32) by $(\mathbf{I}_n - \tilde{\rho}\mathbf{W})$. That is, 2SLS is applied to

$$\mathbf{y}^\star(\tilde{\rho}) = \mathbf{Z}^\star(\tilde{\rho})\boldsymbol{\delta} + \mathbf{e}, \tag{1.39}$$

where $\mathbf{y}^\star(\tilde{\rho}) = (\mathbf{I}_n - \tilde{\rho}\mathbf{W})\mathbf{y}$ and $\mathbf{Z}^\star(\tilde{\rho}) = (\mathbf{I}_n - \tilde{\rho}\mathbf{W})\mathbf{Z}$. Hence, the GS2SLS estimator of $\boldsymbol{\delta}$ is given by

$$\hat{\boldsymbol{\delta}}_{\mathrm{GS2SLS}} = [\mathbf{Z}^\star(\tilde{\rho})'\mathbf{P_H}\mathbf{Z}^\star(\tilde{\rho})]^{-1}\,\mathbf{Z}^\star(\tilde{\rho})'\mathbf{P_H}\mathbf{y}^\star(\tilde{\rho}). \tag{1.40}$$

In Step 4, the efficient GMM estimator of ρ is estimated based on $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}_{\mathrm{GS2SLS}}$ where the efficient GMM weighting matrix is estimated using GS2SLS residuals.

1.1.5 Spatial panel models

Spatial econometric models and estimation methods have been extended to the panel data setting. Recent overviews are provided by Anselin, Le Gallo, and Jayet (2008), Lee and Yu (2010c) and Elhorst (2014).

Static spatial panel models

The panel version of the general Manski model is given by

$$\mathbf{y}_t = \lambda \mathbf{W}\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{W}\mathbf{X}_t\boldsymbol{\theta} + \boldsymbol{\alpha} + \mathbf{u}_t \tag{1.41}$$

$$\mathbf{u}_t = \rho \mathbf{W}\mathbf{u}_t + \mathbf{e}_t \qquad \text{for} \quad t = 1, \ldots, T, \tag{1.42}$$

where we add the t subscript to account for the time dimension. The vector $\mathbf{y}_t = (y_{1t}, \ldots, y_{it}, \ldots, y_{nt})'$ contains the observed outcome of the dependent variable at time t for all units. Similarly, the matrix of regressors is $\mathbf{X}_t = (\mathbf{x}'_{1t}, \ldots, \mathbf{x}'_{it}, \ldots, \mathbf{x}'_{nt})'$. The vector of unobserved variables, $\mathbf{u}_t$ and $\mathbf{e}_t$, are defined accordingly. The spatial weights matrix is assumed to be constant over time, although time-varying weights matrices can be accommodated. The model also includes unit-specific unobserved effects, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$.

As in non-spatial panel models with fixed T and $n \to \infty$, the estimation strategy does crucially depend on whether the regressors, $\mathbf{x}_{it}$, are assumed to be uncorrelated to the unobserved time-invariant effects, α_i. This condition is often referred to as the Random Effects (RE) assumption and can be written as $\mathrm{E}[\alpha_i|\mathbf{x}_{it}] = 0$. If $\lambda = 0$, $\rho = 0$ and the RE assumption holds, Pooled OLS and the classical RE estimator are consistent, but only RE is efficient. However, in many situations the RE assumption is not appropriate. Furthermore, it is not possible to estimate the unobserved heterogeneity α_i consistently for fixed T and $n \to \infty$. This is because an increase in n adds another α_i parameter, without providing more information for the identification of each α_i. This issue is commonly referred to as incidental parameter problem. The Fixed Effects (FE) estimator, which applies OLS to the within-transformed model, does allow for $\mathrm{E}[\alpha_i|\mathbf{x}_{it}] \neq 0$. The within-transformation eliminates $\boldsymbol{\alpha}$ from (1.41) by subtracting unit-specific averages from the dependent and explanatory variables.

Kapoor, Kelejian, and Prucha (2007) consider the SEM model ($\lambda = 0$ and $\rho \neq 0$) and extend the moment conditions in Kelejian and Prucha (1998) to the panel setting under the RE assumption. Mutl and Pfaffermayr (2011) consider GMM estimation with and without the RE assumption. The authors also include a spatially lagged dependent variable ($\lambda \neq 0$) and develop a spatial Hausman test that allows for choosing between RE and FE specification. Lee and Yu (2010b) propose a QML estimator that is consistent for fixed or large T.

Spatial dynamic panel models

The model in (1.41)-(1.42) is static in the sense that temporal dynamics are not accounted for. The Spatial Dynamic Panel Data (SPDP) model captures both temporal and spatial effects and is given by

$$\mathbf{y}_t = \gamma \mathbf{y}_{t-1} + \lambda \mathbf{W}\mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{W}\mathbf{X}_t \boldsymbol{\theta} + \boldsymbol{\alpha} + \mathbf{u}_t \tag{1.43}$$

$$\mathbf{u}_t = \rho \mathbf{W}\mathbf{u}_t + \mathbf{e}_t. \tag{1.44}$$

To guarantee stability, $|\gamma| < 1$ is assumed. Let us first consider the estimation problem for a non-spatial panel ($\lambda = 0$, $\rho = 0$) and fixed T. Pooled OLS is inconsistent, even if the random effects assumption holds, since both y_{it} and y_{it-1} depend on α_i by construction, rendering y_{it-1} endogenous. The within transformation employed by the FE estimator eliminates α_i from the equation, but, as famously shown by Nickell (1981), results in a bias of order $O(1/T)$ and is thus inconsistent for fixed T.

Anderson and Hsiao (1981) and Anderson and Hsiao (1982) suggest to apply first-differencing to remove the time-invariant effect α_i and then to instrument the temporal lag, $\Delta y_{it-1} = y_{it-1} - y_{it-2}$ with Δy_{it-2} or y_{it-2}. Starting with this approach, a vast literature on the GMM estimation of dynamic panel data models for fixed T and $n \to \infty$ emerged. Arellano and Bond (1991) propose a two-step GMM estimator that exploits time-specific moment conditions. For example, y_{i1} is a valid instrument for the equation with Δy_{i3} as the dependent variable and both y_{i2} and y_{i1} are valid for the equation corresponding to Δy_{i4}.[8] Further extensions are introduced

[8]It is easy to see that this approach leads to instrument proliferation, even for reasonably small T, as critically discussed in Roodman (2009).

by, among others, Ahn and Schmidt (1995) and Blundell and Bond (1998).[9]

Mutl (2006) combines the non-spatial moment conditions in Arellano and Bond (1991) with the spatial moment conditions proposed in Kapoor, Kelejian, and Prucha (2007) in order to estimate the dynamic panel data model with spatially autoregressive disturbances ($\lambda = 0$, $\rho \neq 0$). The approach is extended by Baltagi, Fingleton, and Pirotte (2014) who allow for a spatially lagged dependent variable ($\lambda \neq 0$).[10]

In addition to GMM estimation, there are numerous studies on ML and QML estimation of spatial dynamic panels. Elhorst (2005) discusses ML estimation of the dynamic panel model with spatial disturbances ($\lambda = 0$, $\rho \neq 0$) under the FE assumption. Focusing on the same model, Su and Yang (2015) derive asymptotic properties of the QML estimator for fixed T under both RE and FE assumption. Yu, Jong, and Lee (2008) and Lee and Yu (2010a) consider the panel-equivalent of the SAR ($\lambda \neq 0$, $\rho = 0$). The authors propose QML estimators for the settings where both T and n tend to infinity. Elhorst (2010b) provides an extensive, comparative simulation study for the case of small T and $\rho = 0$, and concludes that a bias-corrected ML estimator exhibits the best performance for small sample sizes.

Relationship to common factor models

Common factor models, which were first considered by Hotelling (1933) and Stone (1947), provide an alternative approach for capturing cross-sectional dependence which is closely related to spatial models. A multi-factor model can be written as

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it} \qquad \text{for } t = 1, \ldots, T;\ i = 1, \ldots, n. \tag{1.45}$$

The error term is assumed to depend on a finite number of unobservable, time-varying variables which are common to all units and have heterogeneous effects. $\mathbf{f}_t$ denote the m-dimensional vector of factors and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{im})$ is the associated vector of factor loadings. For instance, common factors may represent global eco-

[9]For an extensive literature overview, see Baltagi (2008).

[10]Lee and Yu (2014) study efficient GMM estimation for the case where $\rho = 0$ and T is either finite or large.

nomic shocks affecting all countries, but with varying intensity. It can be easily seen that setting $\mathbf{f}_t = (y_{1t}, \ldots, y_{i-1,t}, y_{i+1,t}, \ldots, y_{nt})$ yields the spatial autoregressive model, while setting $\mathbf{f}_t = (u_{1t}, \ldots, u_{it}, \ldots, u_{nt})$ gives the spatial error model. Therefore, spatial panel models are a special case of common factor models.

A challenge for the estimation of multi-factor models is that, in many applications, common factors are likely to be correlated with observable regressors. The Common Correlated Effects (CCE) estimator introduced by Pesaran (2006) controls for unobserved common factors by augmenting the regression model with cross-sectional averages of the dependent and independent variables.[11] Pesaran and Tosetti (2011) show that the CCE estimator continues to be consistent and asymptotically normal if the error term follows a spatial process and has a multi-factor structure. In an application to the US housing market, Holly, Pesaran, and Yamagata (2010) unify the multi-factor and spatial econometric approach. The authors employ the CCE estimator to eliminate common factors and, subsequently, analyse the residuals for spatial dependence.

Weak and strong cross-section dependence

The need to account for cross-section dependence underlies both the spatial econometric and the common factor literature. It has been increasingly acknowledged that ignoring cross-section dependence may lead to inconsistency and misleading inference. The CCE estimator due to Pesaran (2006) and the Second Generation of Panel Unit Root Tests are examples for how controlling for cross-section dependence has become a central topic in econometrics (Breitung and Pesaran, 2008). However, the assumption of cross-sectional *in*dependence is often unrealistically strong, especially in large n panels. Hence, there is a need to conceptualise different degrees of dependence.

Chudik, Pesaran, and Tosetti (2011) develop the notion of weak and strong cross-section dependence. A double index process $\{x_{it}\}$ is referred to as cross-sectionally

[11] Alternative methods for estimating multi-factor models, which are not discussed here, are based on principal component analysis (Coakley, Fuertes, and Smith, 2002; Bai, 2009) or maximum likelihood (Robertson and Symons, 2007).

weakly dependent if the degree of dependence as measured by the average pairwise correlation coefficient,

$$\bar{\rho} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Corr}(x_{it}, x_{jt}), \tag{1.46}$$

converges to zero as $n \to \infty$ at a sufficiently fast rate (Pesaran, 2015). Pesaran (2004, 2015) introduce a test for cross-sectional dependence, which is given by

$$CD = \left[\frac{Tn(n-1)}{2}\right]^{1/2} \hat{\bar{\rho}}. \tag{1.47}$$

$\hat{\bar{\rho}}$ denotes the sample analogue of $\bar{\rho}$. The exact rate of convergence required for weak dependence, and thus the exact null hypothesis of the CD test, depends on the relative rate at which T and n grow. Pesaran and Tosetti (2011) point out that spatial processes are weakly cross-sectionally dependent if the row and column sums of the spatial weight matrix are bounded in absolute value, which is a standard assumption in spatial econometric models (as discussed in Section 1.1.2). Factor models, on the other hand, may exhibit both weak and strong forms of dependence. For instance, a strong factor could represent unobserved global economic shocks affecting all countries with heterogeneous intensity, independent of how large n is. Because of this distinction, Holly, Pesaran, and Yamagata (2010) and Bailey, Holly, and Pesaran (2016) apply a two-step approach, where in the first step strong dependence is eliminated using the CCE estimator and in the second step spatial dependence is modelled by means of spatial econometric models.

1.2 The Lasso and high-dimensional statistics

The Least Absolute Shrinkage and Selection Operator (Lasso) is a regularisation technique which, since its introduction by Tibshirani (1996), has been widely applied to various fields, in particular genomics (see, e.g., references in Li and Sillanpää, 2012). As of August 2016, Google Scholar lists above 16,000 citations for the article "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), by Tibshirani (1996), which indicates the popularity of the Lasso.

To motivate the Lasso, let us consider the general linear model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.48}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. Furthermore, suppose that only a subset of $s < p$ parameters are non-zero, i.e.,

$$s := \sum_{i=1}^{p} \mathbb{1}\{\beta_i \neq 0\}, \tag{1.49}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. The assumption of sparsity encompasses the typical setting where the number of potential regressors—and, thus, the number of potential models—is large, but only a small number of variables is relevant. The aim of the analysis is to identify the set of relevant regressors.

The standard asymptotic approach in econometrics is to treat the number of regressors, p, as fixed, while the number of observation is assumed to approach infinity. High-dimensional statistics deals with the situation when p is large and may even be larger than the number of observations.[12] If p is large, variable selection is problematic, as hypothesis testing leads to many false positives. If p is large and $p > n$, the ordinary least squares solution is not unique. While OLS requires $p \leq n$, the Lasso can deal with $p \gg n$ under the assumption that $\boldsymbol{\beta}$ is sparse, i.e., $s \leq n$.

Tibshirani (1996) names two further advantages of the Lasso over least squares. First, the Lasso may improve on the prediction accuracy of least squares. Second, if the number of regressors is large, the Lasso can facilitate model interpretation by reducing the number of regressors to a few.

Belloni and Chernozhukov (2011) argue that large p is common in economics, although often not explicitly acknowledged. A typical example is given by cross-country growth regressions, where the number of control variables is large, while the

[12] The discussion in this section allows p to grow with n, although the Lasso is also considered in the fixed p case. For example, Knight and Fu (2000) and Zou (2006).

number of countries is relatively small. Other examples include instrumental variable regression with many instruments or the existence of many competing models.

The objective function of the Lasso is given by

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \,, \tag{1.50}$$

where the first term is the Residual Sum of Squares (RSS) and the second term imposes a penalty on the absolute size of coefficient estimates. The penalty level, λ, controls the strength of penalisation. $\lambda = 0$ yields the OLS solution and $\lambda \to \infty$ yields a null vector. It is common in the Lasso literature to use vector norm notation. Specifically, the penalty function can be compactly written as $\|\boldsymbol{\beta}\|_1$, where $\|\cdot\|_1$ denotes the ℓ_1-norm, and the RSS term is equivalent to $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$.

1.2.1 Lasso and Ridge regression in comparison

A comparison of Lasso and Ridge regression provides insights about the nature of the Lasso penalty. Ridge regression is an alternative regularisation technique, that replaces the ℓ_1-penalty $\sum_j |\beta_j|$ in (1.50) with the ℓ_2-penalty $\sum_j |\beta_j|^2$ (Hoerl and Kennard, 1970).

In this context, it is instructive to write the objective function of Lasso and Ridge regression as

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 \quad \text{subject to} \quad P(\boldsymbol{\beta}) \le \tau, \tag{1.51}$$

where $P(\boldsymbol{\beta}) = \sum_{j=1}^{p} |\beta_j|$ for the Lasso and $P(\boldsymbol{\beta}) = \sum_{j=1}^{p} |\beta_j|^2$ for Ridge regression. Note that, due to Lagrangian duality, there exists a data-dependent correspondence between the constraint parameter τ and the penalty level λ.

Figure 1.5 illustrates, based on (1.51), the geometry underpinning Lasso and Ridge regression for the case of $p = 2$. The red elliptical lines represent RSS contours and the blue lines indicate the Lasso and Ridge constraints. The Lasso constraint $|\beta_1| + |\beta_2| \le \tau$ is diamond-shaped, whereas the Ridge constraint $\beta_1^2 + \beta_2^2 \le \tau$ is circular. Furthermore, $\hat{\beta}_0$ denotes the solution without penalisation, which corre-

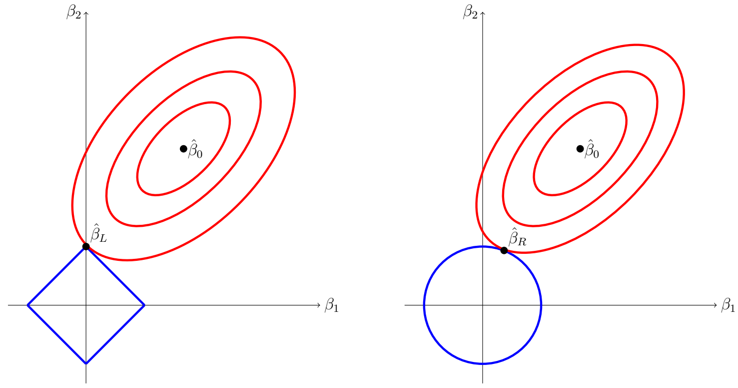


Figure 1.5: *Behaviour of ℓ_1- and ℓ_2-penalty in comparison. Red lines represent RSS contour lines and the blue lines represent the Lasso and Ridge constraint, respectively. $\hat{\beta}_0$ denotes the OLS estimate. $\hat{\beta}_L$ and $\hat{\beta}_R$ are the Lasso and Ridge estimate. The illustration is based on Tibshirani, 1996, Fig. 2.*

sponds to OLS. The Lasso solution at the corner of the diamond implies that, in this example, one of the coefficients is set to zero, whereas Ridge and OLS estimates produce non-zero coefficient estimates.

Figure 1.6 shows a typical solution path for Lasso and Ridge regression for $p = 5$.[13] Each line corresponds to one regressor and the penalty level λ increases from the right to the left-hand side of the horizontal axis. The examples in Figure 1.5 and 1.6 illustrate how the ℓ_1-penalization tends to set some of the coefficient estimates to exactly zero, which makes the Lasso estimator attractive for model selection. The ℓ_1-penalization behaves similar to the ℓ_0-penalty as used in Akaike (1974) and Schwarz (1978), which is known to have good theoretical model selection properties. However, optimisation in the presence of a ℓ_0-penalty is a NP-hard problem and, thus, computationally infeasible even for relatively small p, whereas the Lasso is computationally feasible for large p.

1.2.2 Out-of-sample prediction and Cross-validation

The choice of the penalty level λ determines the complexity of the model. If the focus of the analysis is on prediction, as opposed to the estimation and interpretation of structural parameters, a common approach is to select the value of λ that optimises out-of-sample prediction performance. Consider again the linear model in (1.48) and

[13] The example data set is taken from Hastie, Tibshirani, and Wainwright (2015, Ch. 2) and the estimation was conducted in R using the package `glmnet` (Friedman, Hastie, and Tibshirani, 2010).

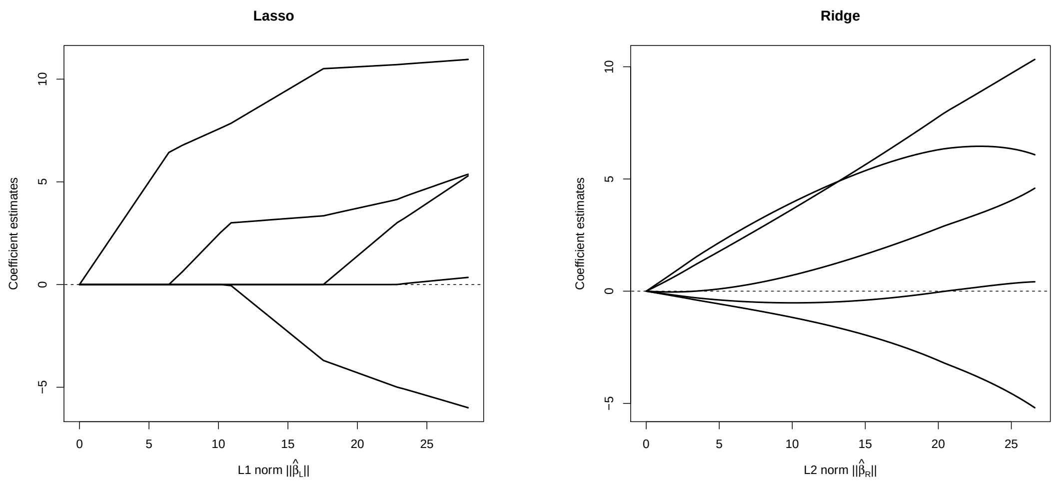


Figure 1.6: *Path of coefficient estimates using ℓ_1- and ℓ_2-penalisation. Initially, the model has five variables (excluding a constant). As the penalty level increases (from right to left), the absolute size of coefficient estimates declines. The Lasso reduces the model complexity by setting some of the coefficient estimates to exactly zero. The illustration is based on Hastie, Tibshirani, and Wainwright, 2015, Fig. 2.1.*

suppose we attempt to predict the dependent variable for a given, observed vector $\mathbf{x}_0$. The expected Mean-squared Prediction Error (MSPE) can be decomposed into

$$\mathrm{E}\Big[(y_0 - \hat{y}_0)^2\Big] = \mathrm{Var}(\hat{y}_0) + \mathrm{Bias}(\hat{y}_0)^2 + \mathrm{Var}(\varepsilon), \tag{1.52}$$

where the squared bias is $\mathrm{Bias}(\hat{y}_0)^2 = \left(\mathrm{E}[\hat{y}_0] - y_0\right)^2$, y_0 is the true value of the dependent variable and $\hat{y}_0$ is an estimator of y_0. It is well known that, under the Gauss-Markov assumptions, OLS is the estimator that exhibits the lowest variance of all unbiased estimators.[14] However, the estimator that minimises the MSPE is not necessarily an unbiased estimator. For illustration, Figure 1.7 compares the prediction performance of four estimation methods. The diagram demonstrates that a prediction method with high bias and low variance (bottom-left square) could outperform a prediction method with low bias and high variance (top-right) in terms of average euclidean distance to the true point.

The Lasso is an example of an estimator that, due to ℓ_1-penalisation, is biased towards zero, but may outperform OLS in terms of MSPE. This is because reducing

[14]The Gauss-Markov assumptions require linearity, exogenous regressors (i.e., $\mathrm{E}[\varepsilon_i|\mathbf{X}] = 0$) and spherical errors (i.e., $\mathrm{E}[\varepsilon_i\varepsilon_j|\mathbf{X}] = 0$ and $\mathrm{E}[\varepsilon_i^2|\mathbf{X}] = \sigma_\varepsilon^2$). The assumption of linearity is implied by equation (1.48).

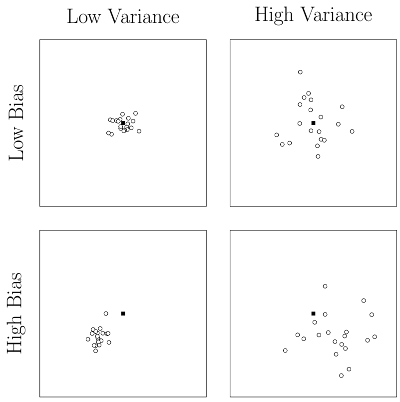


Figure 1.7: *Schematic illustration of bias and variance. The squared points ('▪') indicate the true value and round points ('∘') represent estimates. The diagram shows four estimation methods with either low or high variance and either low or high bias. In some situations, there exists a trade-off between bias and variance. For example, a high bias/low variance estimator may yield predictions that are on average closer to the truth than predictions from a low bias/high variance estimator.*

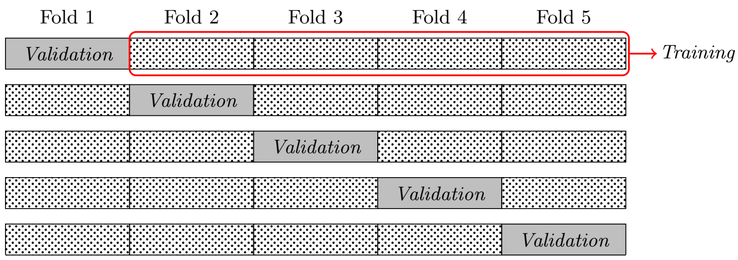


Figure 1.8: *Construction of folds for 5-fold Cross-validation. The data set is divided into 5 sub-sets, referred to as groups or folds. In each iteration, one group is treated as the validation data set (represented by gray boxes), while the remaining groups are the training data set (dotted boxes). Based on James et al. (2014, Fig. 5.5).*

the model flexibility and complexity, as measured by the number of included predictors, tends to decreases variance, while increasing bias (Hastie, Tibshirani, and Friedman, 2009, Ch. 7.3).

Since the prediction error is unobservable, a widely applied approach is to estimate the MSPE using K-fold Cross-validation (Hastie, Tibshirani, and Friedman, 2009). The MSPE estimate serves as an indicator of model performance and can be used to select between alternative model specifications. In particular, as described in the following steps, Cross-validation (CV) can be employed to identify the value of λ that minimises the estimated MSPE:

STEP 1: The dataset of n observations is randomly divided into K groups, also referred to as folds, of approximately equal size. Let $\mathcal{K}_1$ denote the set of all

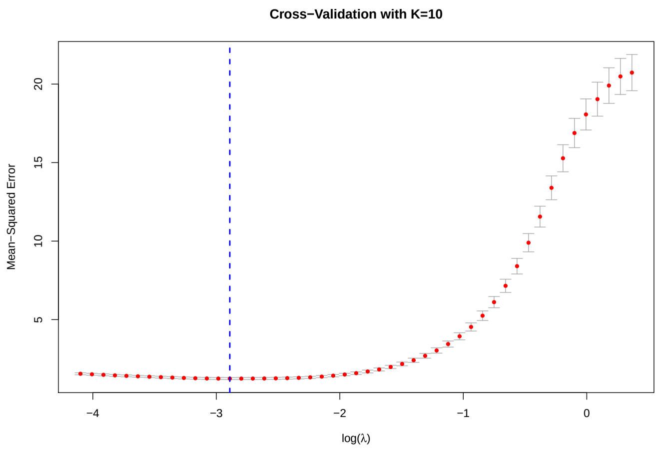


Figure 1.9: *An example of cross-validation for an artificial data set with* $n = 500$*,* $p = 500$ *and* $s = 20$*. The dashed blue line indicates the value of* λ *that minimises the mean-squared error. The number of included predictors decreases from left (low penalty) to right (high penalty).*

observations included in the first group, $\mathcal{K}_2$ denote the set of all observations in the second group, etc. Furthermore, $n_1, \ldots, n_K$ is the group size of group 1 to K.

STEP 2: The first group is treated as the validation data set and, for a given value of λ, the estimation method is applied to the remaining $K-1$ groups, which constitute the training dataset. The MSPE estimate corresponding to group 1 is computed as

$$\widehat{\text{MSPE}}_1(\lambda) = \frac{1}{n_1} \sum_{i \in \mathcal{K}_1} (y_i - \hat{y}_i)^2. \tag{1.53}$$

Then, as illustrated in Figure 1.8, the procedure is repeated treating the second, third, ..., Kth group as the validation data set. Thus, $\widehat{\text{MSPE}}_2(\lambda)$, ..., $\widehat{\text{MSPE}}_K(\lambda)$ are computed.

STEP 3: The K-fold Cross-validation estimate of the MSPE is

$$\widehat{\text{CV}}_K(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \widehat{\text{MSPE}}_k. \tag{1.54}$$

STEP 4: STEP 2 and STEP 3 are re-iterated for a range of λ values in order to identify the penalty level that minimises $\text{CV}_K(\lambda)$.

The special case of $K = n$ is referred to as Leave-one-out Cross-validation. However, due to computational constraints, $K = 5$ or $K = 10$ are commonly chosen

(Hastie, Tibshirani, and Friedman, 2009, p. 242). Figure 1.9 presents results for a 10-fold Cross-validation applied to an artificial data set generated for $n = 500$, $p = 500$ and $s = 20$.[15] The red dots show the estimated MSPE for given values of λ and the attached bars display a two standard deviation range around the point estimate. As indicated by the dashed blue line, the CV reveals that $\hat{\lambda}_{\text{min}} = 0.056$ is the penalty level that minimises the mean-squared prediction error. The penalty level $\hat{\lambda}_{\text{min}}$ yields 93 non-zero coefficient estimates. All true predictors are successfully identified, but 73 variables are falsely included as predictors.

1.2.3 Theoretical results

The theoretical properties of the Lasso are assessed on the basis of its prediction performance, the performance with regard to parameter estimation and the performance in recovering the support of the true parameter vector, $\boldsymbol{\beta}^\star$.[16] Predictive performance is typically measured by the ℓ_2-prediction norm $\|\mathbf{X}(\boldsymbol{\beta}^\star - \hat{\boldsymbol{\beta}})/\sqrt{n}\|_2$, whereas $\|\boldsymbol{\beta}^\star - \hat{\boldsymbol{\beta}}\|_2$ or $\|\boldsymbol{\beta}^\star - \hat{\boldsymbol{\beta}}\|_1$ quantify the loss from estimating the parameter vector.

Non-asymptotic performance bounds

The following exposition develops non-asymptotic bounds for the Lasso and introduces two central conditions.[17] The first condition relates to the choice of the penalty level, λ, and the second to the Gram matrix of regressors, $\mathbf{X}'\mathbf{X}/n$. For simplicity, we assume that $\mathbf{X}$ is non-stochastic.

Theoretical results with regard to the prediction norm and parameter norm rely on the *basic inequality*, which follows from the observation that, since $\hat{\boldsymbol{\beta}}$ minimises

[15] The data generating process is given by eq. (1.48) with $\beta_j = 1$ for $j = 1, \dots, 20$ and $\beta_j = 0$ for $j = 21, \dots, p$. Furthermore, $x_{il} \sim \mathcal{N}(0, 1)$ and $\varepsilon_i \sim \mathcal{N}(0, 1)$ for $l = 1, \dots, p$ and $i = 1, \dots, n$.

[16] In the following, I use a '$\star$'-superscript to emphasis that $\boldsymbol{\beta}^\star$ corresponds to the true parameter vector.

[17] The discussion is based on Hastie, Tibshirani, and Wainwright (2015, Chapter 11) and Bühlmann and Van de Geer (2011, Chapter 6).

the Lasso objective function,

$$\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2/n + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^\star\|_2^2/n + \lambda\|\boldsymbol{\beta}^\star\|_1. \tag{1.55}$$

Rewriting the above property yields the basic inequality

$$\|\mathbf{X}\hat{\boldsymbol{\delta}}\|_2^2/n \leq 2\boldsymbol{\varepsilon}'\mathbf{X}\hat{\boldsymbol{\delta}}/n + \lambda(\|\boldsymbol{\beta}^\star\|_1 - \|\hat{\boldsymbol{\beta}}\|_1), \tag{1.56}$$

where $\hat{\boldsymbol{\delta}} := \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star$. The first term on the right-hand side includes the random term $\boldsymbol{\varepsilon}$ and, by Hölder's inequality, $|\boldsymbol{\varepsilon}'\mathbf{X}\hat{\boldsymbol{\delta}}| \leq \|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty\|\hat{\boldsymbol{\delta}}\|_1$. The idea is to select the penalty level such that it dominates the random part of the equation,

$$2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n \leq \lambda. \tag{1.57}$$

Furthermore, the second term in equation (1.56) is equivalent to $\|\hat{\boldsymbol{\delta}}_\Omega\|_1 - \|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1$. It follows that

$$\|\mathbf{X}\hat{\boldsymbol{\delta}}\|_2^2/n \leq \lambda 2\sqrt{s}\|\hat{\boldsymbol{\delta}}\|_2. \tag{1.58}$$

To establish a bound for the prediction norm $\|\mathbf{X}\hat{\boldsymbol{\delta}}\|_2/\sqrt{n}$ and the parameter norm $\|\hat{\boldsymbol{\delta}}\|_2$, we need to relate the two expressions. This is straightforward in the low-dimensional setting where $p \leq n$ and the Gram matrix $\mathbf{X}'\mathbf{X}/n$ is well-behaved. Then, the Gram matrix is positive-definite, its smallest eigenvalue is positive, i.e.

$$\min_{\boldsymbol{\delta}\in\mathbb{R}^p:\boldsymbol{\delta}\neq 0} \frac{\|\mathbf{X}\boldsymbol{\delta}\|_2}{\|\boldsymbol{\delta}\|_2} > 0, \tag{1.59}$$

and we can substitute for either $\|\mathbf{X}\hat{\boldsymbol{\delta}}\|_2/\sqrt{n}$ or $\|\hat{\boldsymbol{\delta}}\|_2$. However, in the high-dimensional setting where p is larger than n, the Gram matrix $\mathbf{X}'\mathbf{X}/n$ is necessarily rank-deficient and the smallest eigenvalue is zero.[18] Therefore, the OLS assumption that the matrix of regressors is of full column rank needs to be replaced by a weaker assumption. In their seminal contribution, Bickel, Ritov, and Tsybakov

[18] To see why, note that the $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'\mathbf{X}) \leq \min(n, p)$.

(2009) propose the *restricted eigenvalue condition*,

$$\kappa_{\min} := \min_{\boldsymbol{\delta} \in \mathcal{C}(c_0, \Omega)} \frac{\|\mathbf{X}\boldsymbol{\delta}\|_2}{\sqrt{n}\|\boldsymbol{\delta}\|_2} > 0, \quad \mathcal{C}(c_0, \Omega) = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}_{\overline{\Omega}}\|_1 \leq c_0 \|\boldsymbol{\delta}_\Omega\|_1\}, \qquad (1.60)$$

where $\kappa_{\min}$ denotes the restricted minimum eigenvalue and $c_0 \geq 1$ is a parameter. It can be shown that $\hat{\boldsymbol{\delta}}$ lies in the restricted set $\mathcal{C}(c_0, \Omega)$ (see Lemma A.2 in Appendix A.1).

Different variants of the RE condition have been proposed in the literature (e.g., Belloni et al., 2012, or the compatibility condition in Bühlmann and Van de Geer, 2011). Bühlmann and Van de Geer (2011) provide an overview over the RE condition and related conditions. The RE condition is shown to hold under quite general conditions. For example, one sufficient condition is that appropriate sub-matrices of $\mathbf{X}'\mathbf{X}/n$ are invertible (see Bickel, Ritov, and Tsybakov, 2009, p. 1710). Other sufficient conditions are given in, for example, Raskutti, Wainwright, and Yu (2010).

By substituting the restricted minimum eigenvalue $\kappa_{\min}$ into (1.58), we can establish the ℓ_2-prediction norm and ℓ_1-parameter norm as summarised in Theorem 1.1 (see proof in Appendix A.1).

THEOREM 1.1. *Suppose the non-stochastic matrix* $\mathbf{X}$ *satisfies the RE condition in* (1.60) *and* λ *is chosen such that* $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n \leq \lambda$, *then*

$$\frac{1}{\sqrt{n}}\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)\|_2 \leq \frac{2\lambda\sqrt{s}}{\kappa_{\min}}, \qquad (1.61)$$

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\|_2 \leq \frac{2\lambda\sqrt{s}}{\kappa_{\min}^2}. \qquad (1.62)$$

Theorem 1.1 suggests that the penalty level should be chosen such that it dominates the term $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n$ with high probability. This approach thus differs from selecting the penalty level by cross-validation as discussed in Section 1.2.2.

Since $\boldsymbol{\varepsilon}$ is unobservable, the question remains which value of λ ensures that the event $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n < \lambda$ occurs with high probability or asymptotically. At the same time, the penalty level should not be too high in order to avoid unnecessary bias. The classical approach is to assume that $\boldsymbol{\varepsilon}$ is *i.i.d.* and Gaussian (e.g., Bickel, Ritov, and Tsybakov, 2009), which allows to derive the distribution of $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n$ and set the penalty level accordingly (see, e.g., Lemma 6.2 in Bühlmann and Van de Geer, 2011). However, the assumption of homoskedastic and Gaussian errors is strong and, as shown by Belloni et al. (2012) can be weakened. Allowing for non-Gaussian and heteroskedastic errors, Belloni et al. (2012) make use of the moderate deviation theory for self-normalised sums developed in Jing, Shao, and Wang (2003) to derive the smallest penalty that ensures $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n < \lambda$ as $n \to \infty$.

Model selection performance

It is important to emphasise that model selection consistency is qualitatively different from consistent prediction or parameter estimation. Both the prediction norm $\|\mathbf{X}(\boldsymbol{\beta}^\star - \hat{\boldsymbol{\beta}})/\sqrt{n}\|_2$ and the parameter norm $\|\boldsymbol{\beta}^\star - \hat{\boldsymbol{\beta}}\|_2$ may be small, even if the support of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^\star$ differ (Hastie, Tibshirani, and Wainwright, 2015, p. 301). Model selection consistency is formally given if

$$P\left(\text{supp}(\hat{\boldsymbol{\beta}}) = \text{supp}(\boldsymbol{\beta}^\star)\right) \to 1 \text{ as } n \to \infty. \tag{1.63}$$

A slightly stronger, but more tractable property is sign consistency,

$$P\left(\text{sign}(\hat{\boldsymbol{\beta}}) = \text{sign}(\boldsymbol{\beta}^\star)\right) \to 1 \text{ as } n \to \infty. \tag{1.64}$$

A sufficient and (almost) necessary condition for sign consistency is formalised by Zhao and Yu (2006) (see also Meinshausen and Bühlmann, 2006 and Wainwright, 2009) and referred to as *Irrepresentable condition.* Let us partition the Gram matrix $\frac{1}{n}\mathbf{X}'\mathbf{X}$ such that

$$\frac{1}{n}\mathbf{X}'\mathbf{X} = \begin{pmatrix} \frac{1}{n}\mathbf{X}[\Omega]'\mathbf{X}[\Omega] & \frac{1}{n}\mathbf{X}[\bar{\Omega}]'\mathbf{X}[\Omega] \\ \frac{1}{n}\mathbf{X}[\Omega]'\mathbf{X}[\bar{\Omega}] & \frac{1}{n}\mathbf{X}[\bar{\Omega}]'\mathbf{X}[\bar{\Omega}] \end{pmatrix} \tag{1.65}$$

where the matrix $\mathbf{X}[\Omega] \in \mathbb{R}^{n\times s}$ is composed of the columns that correspond to the active set $\Omega = \text{supp}(\boldsymbol{\beta}^\star)$ and s is the cardinality of Ω. $\bar{\Omega}$ is the complementary set and $\mathbf{X}[\bar{\Omega}] \in \mathbb{R}^{n\times(p-s)}$ is defined accordingly. The irrepresentable condition requires the existence of an incoherence parameter $\gamma > 0$ such that

$$\max_{j\in\bar{\Omega}} \left\|(\mathbf{X}[\Omega]'\mathbf{X}[\Omega])^{-1}\mathbf{X}[\Omega]'\mathbf{X}_j\right\|_1 \leq 1-\gamma. \tag{1.66}$$

Note that the vector $(\mathbf{X}[\Omega]'\mathbf{X}[\Omega])^{-1}\mathbf{X}[\Omega]'\mathbf{X}_j$ is the OLS estimate from regressing the jth column of $\mathbf{X}[\bar{\Omega}]$ against $\mathbf{X}[\Omega]$ and thus a measure of correlation between the columns in $\mathbf{X}[\Omega]$ and $\mathbf{X}_j$. In the case of $p < n$, if all columns of $\mathbf{X}[\bar{\Omega}]$ are orthogonal to $\mathbf{X}[\Omega]$, the condition is satisfied with $\gamma = 1$. On the other hand, if $(\mathbf{X}[\Omega]'\mathbf{X}[\Omega])^{-1}\mathbf{X}[\Omega]'\mathbf{X}_j$ is large for some j, there may not exist a positive value of γ such that condition (1.66) holds. Intuitively, the irrepresentable condition states that none of the variables in the inactive set are allowed to be highly correlated with variables in the active set, since the Lasso would not be able to distinguish between irrelevant and relevant regressors.

From (1.66) also follows that $\frac{1}{n}\mathbf{X}[\Omega]'\mathbf{X}[\Omega]$ is required to be positive definite, implying the condition $s < n$. A violation of positive definiteness would mean that the true model is not identified, even if the true support were known. Under the irrepresentable condition (1.66), positive definiteness of $\frac{1}{n}\mathbf{X}[\Omega]'\mathbf{X}[\Omega]$ and assuming that the penalty level is appropriately chosen, Zhao and Yu (2006) and Wainwright (2009) show that the Lasso solution is unique and sign consistent. The irrepresentable condition is much stronger than the restricted eigenvalue condition, which allows for a higher degree of correlation. Bühlmann and Van de Geer (2011, Chapter 7) conclude that the Lasso achieves consistent variable selection in only very specific cases and not in many applications where the regressors in question exhibit strong correlation.

1.2.4 Related methods

The Lasso estimator has inspired the development of alternative penalised regression techniques, which in some situations dominate the Lasso in terms of prediction or model selection performance. Some of the most prominent methods are briefly discussed in this section.

Post-Lasso. The ℓ_1-penalisation imposed by the Lasso shrinks coefficient estimates towards zero. An intuitive approach for reducing the bias resulting from penalisation is the Post-Lasso OLS estimator, which applies OLS to the model selected by the Lasso. Specifically, the Post-Lasso estimator is defined as

$$\tilde{\boldsymbol{\beta}} = \arg\min \frac{1}{n}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \quad \text{subject to} \quad \text{supp}(\boldsymbol{\beta}) = \text{supp}(\hat{\boldsymbol{\beta}}), \tag{1.67}$$

where $\hat{\boldsymbol{\beta}}$ is the Lasso solution. Belloni and Chernozhukov (2013) show that the convergence rates of the Post-Lasso OLS are at least as good as the Lasso rates, while only relying on a slightly stronger condition with regard to the Gram matrix which is referred to as the restricted sparse eigenvalue condition.

Elastic net. The Lasso has two disadvantages relative to Ridge regression (Zou and Hastie, 2005). First, if $p > n$, the Lasso selects at most p regressors. Secondly, in the presence of high correlation among groups of regressors, the Lasso tends to select only one variable from each group, omitting potentially relevant explanatory variables. The elastic net introduced by Zou and Hastie (2005) uses a weighted average of ℓ_1- and ℓ_2-penalty and, thus, nests both Lasso and Ridge regression as special cases. Zou and Hastie (2005) demonstrate that the elastic net may outperform the Lasso in terms of prediction performance and variable selection. One disadvantage of the elastic net is that it requires to specify an additional parameter, which determines the relative weight of ℓ_1- to ℓ_2-penalisation. Mol, Vito, and Rosasco (2009) analyse the theoretical properties of the elastic net under the assumption of sparsity.

Square-root Lasso. The theoretical properties of the Lasso summarised in Theorem 1.1 rely on setting the penalty level such that it dominates the term $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n$. Thus, the optimal penalty depends on the noise level, σ_ε. Belloni et al. (2012) propose an algorithm to estimate the ideal penalty iteratively, which can also accommodate heteroskedasticity. The Square-root Lasso, due to Belloni, Chernozhukov, and Wang (2011) and Belloni, Chernozhukov, and Wang (2014), offers an alternative approach. The objective function is given by

$$\min \frac{1}{n}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2 + \lambda\|\boldsymbol{\beta}\|_1 \tag{1.68}$$

and the authors show that the optimal penalty level does not depend on the noise level, implying a substantial practical advantage.

Adaptive Lasso. Zou (2006) proposes the Adaptive Lasso, which minimises

$$\min \frac{1}{n}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\sum_{j=1}^{p}\frac{|\beta_j|}{|\tilde{\beta}_j|^{\nu}}, \tag{1.69}$$

where $\tilde{\beta}_j$ is some initial parameter estimate and $\nu > 0$. If $p < n$, the $\tilde{\beta}_j$ may be set to the OLS estimate. Zou (2006) shows that, in the context of fixed p, the Adaptive Lasso achieves model selection consistency without relying on the irrepresentable condition in (1.66). Huang, Ma, and Zhang (2008) explore the case where p is allowed to grow with n.

Dantzig selector. Candes and Tao (2007) introduce the Dantzig selector, which minimises

$$\min\|\boldsymbol{\beta}\|_1 \quad \text{subject to} \quad \|\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\|_\infty \leq \lambda. \tag{1.70}$$

Bickel, Ritov, and Tsybakov (2009) show that the Lasso and the Dantzig selector are closely related and exhibit similar theoretical properties. Gautier and Tsybakov (2014) propose an extension to the Dantzig selector, referred to as Self-Tuning Instrumental Variable (STIV) estimator, which allows for endogenous regressors in high dimensions.

Shrinkage GMM. Another promising approach in the high-dimensional setting relies on GMM. Similar to penalised regression, shrinkage GMM adds a penalisation term to the GMM objective function (Caner, 2009). More recently, Fan and Liao (2014) propose a penalised GMM estimator, termed Focused Generalised Methods of Moments (FGMM), that can accommodate endogeneity and high-dimensionality.

1.3 Linking spatial econometrics and the Lasso

It may not be immediately clear how the Lasso estimator—and high-dimensional methods in general—can be exploited in spatial econometric models. This section outlines three situations, each of which is considered in the following chapters of this thesis.

Chapter 2: Lasso-generated instruments. In Chapter 2, temperature and rainfall are used as instruments to identify the causal impact of economic growth shocks on conflict. The form of the function linking economic growth and climate conditions is unknown, but is likely to be characterised by nonlinearity and spatial heterogeneity across climate regions. Similarly, the spatial conflict lag, which captures spill-over effects of conflicts, needs to be instrumented due to reverse causality. The framework developed in Belloni et al. (2012) is applicable in this setting and can be employed to generate approximately optimal instruments for economic growth and the spatial conflict lag.

Belloni, Chen, Chernozhukov, and Hansen (2012) consider the model

$$y_i = \mathbf{x}_i'\boldsymbol{\alpha} + \varepsilon_i, \tag{1.71}$$

$$x_{il} = D_l(\mathbf{z}_i) + v_i, \tag{1.72}$$

where $\mathbf{x}_i$ is a vector of regressors correlated with the error term ε_i.[19] The vector $\mathbf{z}_i$ consists of instrumental variables for which $\text{E}[\varepsilon_i|\mathbf{z}_i] = 0$ and $D_l(\mathbf{z}_i)$ is the conditional expectation function

$$D_{il} := D_l(\mathbf{z}_i) := \text{E}[x_{il}|\mathbf{z}_i]. \tag{1.73}$$

In order to approximate the unknown, potentially non-linear function $D_l(\mathbf{z}_i)$, a large number of transformations of $\mathbf{z}_i$ is considered, denoted by $\mathbf{f}_i = (f_1(\mathbf{z}_i), \ldots, f_p(\mathbf{z}_i))'$, where p may be larger than the number of observations, n. Note that $\mathbf{f}_i$ may be set to $\mathbf{z}_i$ if linearity is assumed.

[19] To simplify the notation, I assume without loss of generality that all regressors in $\mathbf{x}_i$ are endogenous, while Belloni et al. (2012) allow $\mathbf{x}_i$ to include exogenous regressors.

In the setting of the conflict application, y_i is a conflict measure, $\mathbf{x}_i$=$(x_{i1}, \ldots, x_{ik})$ includes economic growth and the spatially lagged dependent variable. The unknown functions $D_l(\cdot)$ link observable climate conditions with the endogenous regressors. To account for spatial heterogeneity and nonlinearity, $\mathbf{f}_i$ is constructed using transformations of climate variables such as dummy variables, polynomials and interactions terms as well as spatially lagged climate variables. The Lasso and Post-Lasso can then be applied to

$$x_{il} = \mathbf{f}_i'\boldsymbol{\beta}_l + \nu_{il} \qquad \text{for} \quad l = 1, \ldots, k, \tag{1.74}$$

and approximately optimal instruments can be obtained. The example illustrates how the Lasso estimator can be exploited in situations where the analyst is faced with a large number of variables and with uncertainty about the correct model specification. In particular, the use of a formal instrument selection method reduces the risk of misspecfication substantially.

Chapter 3: The spatial autoregressive panel model. The vast majority of applied and theoretical spatial econometric research relies on the assumption that the spatial weights matrix is known *a priori*. However, in most applications, the researcher has no other option than specifying the weights matrix on an *ad hoc* basis guided by theory. In Chapter 3, I consider a two-step estimation strategy for estimating the $n(n-1)$ interaction effects in the spatial autoregressive panel model

$$y_{it} = \sum_{i \neq j} w_{ij} y_{jt} + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \qquad i = 1, \ldots, n, \;\; t = 1, \ldots, T, \tag{1.75}$$

where w_{ij} is the (i, j)th element of the spatial weights matrix. The identifying assumption is sparsity of the spatial weights matrix, which requires that only a small number of row elements is non-zero. The proposed estimation methodology exploits the Lasso estimator and mimics two-stage least squares (2SLS) to account for endogeneity of the spatial lag.

The developed two-step estimator is of more general interest. It may be used in applications where the number of endogenous regressors and the number of instrumental variables is larger than the number of observations. I derive convergence rates for the two-step Lasso estimator. Monte Carlo simulation results show that the two-step estimator recovers the spatial network structure successfully if the time dimension is reasonably large.

Chapter 4: A spatial panel model with spatially lagged regressors. Chapter 4 considers an alternative spatial model, which is given by

$$y_{it} = x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt} + \varepsilon_{it}, \qquad i = 1, \ldots, n, \;\; t = 1, \ldots, T. \tag{1.76}$$

In the above model, spatial effects arise due to the exogenous regressor x_{jt}, which exerts influence on the response variable observed at other locations. As in the spatial autoregressive panel model, the model parameters are not identified if the spatial dimension is large relative to the time dimension. The Lasso-based double-selection method distinguishes between low dimensional parameters (here, β), for which consistent and asymptotically normal estimates can be obtained, and the high-dimensional spatial effects, which are treated as confounding factors.

The above examples demonstrate that high-dimensionality is a common phenomenon in economics and, especially, in spatial econometrics. Spatial effects can be interpreted as high-dimensional parameters and the Lasso can be exploited to estimate or control for spatial effects.

Appendix of Chapter 1

A.1 Proof of Theorem 1.1

The proof of Theorem 1.1 uses two Lemmas. Note that Lemma A.2 is similar to Lemma 11.1 in Hastie, Tibshirani, and Wainwright (2015).

Lemma A.1. *Under the assumption* $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n \leq \lambda$*, the Lasso solution satisfies* $\|\mathbf{X}\hat{\boldsymbol{\delta}}\|_2^2/n \leq \lambda 2\sqrt{s}\|\hat{\boldsymbol{\delta}}\|_2$.

Proof of Lemma A.1. As explained in the text,

$$\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2/n + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^\star\|_2^2/n + \lambda\|\boldsymbol{\beta}^\star\|_1. \tag{A.1}$$

It follows that

$$\|\mathbf{X}\hat{\boldsymbol{\delta}}\|_2^2/n \leq 2\boldsymbol{\varepsilon}'\mathbf{X}\hat{\boldsymbol{\delta}}/n + \lambda(\|\boldsymbol{\beta}^\star\|_1 - \|\hat{\boldsymbol{\beta}}\|_1). \tag{A.2}$$

With regard to the first term on the right-hand side, note that by the Hölder's inequality $|\boldsymbol{\varepsilon}'\mathbf{X}\hat{\boldsymbol{\delta}}| \leq \|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty\|\hat{\boldsymbol{\delta}}\|_1$. Thus, on the event $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n \leq \lambda$, we get

$$2\boldsymbol{\varepsilon}'\mathbf{X}\hat{\boldsymbol{\delta}}/n \leq 2/n\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty\|\hat{\boldsymbol{\delta}}\|_1 \leq \lambda\|\hat{\boldsymbol{\delta}}\|_1. \tag{A.3}$$

With regard to the second term,

$$\|\boldsymbol{\beta}^\star\|_1 - \|\hat{\boldsymbol{\beta}}\|_1 = \|\boldsymbol{\beta}^\star\|_1 - (\|\hat{\boldsymbol{\beta}}_\Omega\|_1 + \|\hat{\boldsymbol{\beta}}_{\overline{\Omega}}\|_1) = \|\boldsymbol{\beta}^\star\|_1 - (\|\hat{\boldsymbol{\beta}}_\Omega\|_1 + \|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1)$$

where we have used that $\hat{\boldsymbol{\beta}}_{\overline{\Omega}} = \hat{\boldsymbol{\delta}}_{\overline{\Omega}}$. Furthermore, by reverse triangle inequality,

$$\|\hat{\boldsymbol{\beta}}_\Omega\|_1 \geq \|\boldsymbol{\beta}^\star_\Omega\|_1 - \|\hat{\boldsymbol{\delta}}_\Omega\|_1. \tag{A.4}$$

Since $\boldsymbol{\beta}^\star = \boldsymbol{\beta}^\star_\Omega$, it follows that

$$\|\boldsymbol{\beta}^\star\|_1 - \|\hat{\boldsymbol{\beta}}\|_1 \leq \|\boldsymbol{\beta}^\star\|_1 - (\|\boldsymbol{\beta}^\star_\Omega\|_1 - \|\hat{\boldsymbol{\delta}}_\Omega\|_1 + \|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1) \leq \|\hat{\boldsymbol{\delta}}_\Omega\|_1 - \|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1. \tag{A.5}$$

Substituting (A.3) and (A.5) into (A.2) gives the result in Lemma A.1,

$$\|\mathbf{X}\hat{\boldsymbol{\delta}}\|_2^2/n \leq \lambda\|\hat{\boldsymbol{\delta}}\|_1 + \lambda(\|\hat{\boldsymbol{\delta}}_\Omega\|_1 - \|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1) \leq \lambda 2\|\hat{\boldsymbol{\delta}}_\Omega\|_1 \leq \lambda 2\sqrt{s}\|\hat{\boldsymbol{\delta}}\|_2. \tag{A.6}$$

Lemma A.2. *Under the assumption* $2\|\boldsymbol{\varepsilon}'\mathbf{X}\|_\infty/n \leq \lambda$*, the Lasso solution satisfies* $\hat{\boldsymbol{\delta}} \in \mathcal{C}(c_0, \Omega)$ *for* $c_0 = 2$*.*

Proof of Lemma A.2. Since $\lambda > 0$ and using (A.6),

$$0 \leq \lambda\|\hat{\boldsymbol{\delta}}\|_1 + \lambda(\|\hat{\boldsymbol{\delta}}_\Omega\|_1 - \|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1) \tag{A.7}$$

$$0 \leq \|\hat{\boldsymbol{\delta}}\|_1 + (\|\hat{\boldsymbol{\delta}}_\Omega\|_1 - \|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1) \tag{A.8}$$

$$\|\hat{\boldsymbol{\delta}}_{\overline{\Omega}}\|_1 \leq \|\hat{\boldsymbol{\delta}}\|_1 + \|\hat{\boldsymbol{\delta}}_\Omega\|_1 \leq 2\|\hat{\boldsymbol{\delta}}\|_1. \tag{A.9}$$

Proof of Theorem 1.1. The Theorem follows from Lemma A.1, Lemma A.2 and by substituting $\kappa_{\min}$ into equation (1.58).

Chapter 2

Civil Conflicts in Africa[1]

The theoretical literature on civil conflict suggests three major channels through which economic conditions affect civil conflicts—namely, opportunity costs, the expected payoff of rebellion and the state's capacity to prevent insurgence. The relative importance and empirical significance of these mechanisms is disputed. Despite a vast number of empirical studies, there is still no consensus as to whether economic shocks have a significant causal impact on civil conflict. While some studies find that income or price changes affect conflict risk (Miguel, Satyanath, and Sergenti, 2004; Brückner and Ciccone, 2010; Dube and Vargas, 2013), other studies cast doubt on this view (Djankov and Reynal-Querol, 2010; Bergholt and Lujala, 2012; Koubi et al., 2012; Weezel, 2015). In this article, I attempt to contribute to the debate, firstly, by examining a novel sub-national panel data set of African first-order administrative units[2] and, secondly, by considering data-driven climate instruments that account for non-linearities and heterogeneity across regions.

Constrained by the lack of suitable sub-national data, empirical research usually focuses on countries as units of observations. The need for econometric analysis using disaggregated data has been stressed by authors from many disciplines, including conflict research (e.g. Buhaug et al., 2011; Blattman and Miguel, 2010; Jensen and

[1]A short version of this chapter has been published in *Peace Economics, Peace Science and Public Policy*, 2015, 21(4). See Ahrens (2015).

[2]First-order administrative units correspond to states in the United States. In the following, the terms areas, sub-national areas or first-order administrative units are used interchangeably.

Gleditsch, 2009; Sundberg and Melander, 2013). A study by Henderson, Storeygard, and Weil (2012) proposes a framework for predicting Gross Domestic Product (GDP) using nighttime light data from satellites for countries with missing or low quality national accounts data, as well as for sub-national areas (see also Nordhaus and Chen, 2012). I build upon Henderson, Storeygard, and Weil (2012) in order to predict economic growth for African areas.[3] The economic growth predictions are then used to analyse the causal link between income shocks and civil conflicts on the sub-national level.

Based on the identification strategy proposed by Miguel, Satyanath, and Sergenti (2004), I address reverse causality in the economic growth-conflict relationship using rainfall and temperature variables as instruments for economic growth. The economic rationale for exploiting rainfall is that African economies are highly dependent on rain-fed agricultural production.[4] For instance, Barrios, Bertinelli, and Strobl (2010) show that rainfall is an important determinant of economic growth in Africa and recent empirical studies link annual temperature variations to economic output (e.g. Dell, Jones, and Olken, 2012; Heal and Park, 2013; Lanzafame, 2014; Burke, Hsiang, and Miguel, 2015), suggesting that measures of temperature could provide additional instruments. Both rainfall and temperature are widely used as instruments for economic growth or are directly related to political and socio-economic conditions (e.g. Miguel, 2005; Bohlken and Sergenti, 2010; Hidalgo et al., 2010; Aidt and Leon, 2015; Kim, 2014; Burke and Leigh, 2010; Hsiang, Burke, and Miguel, 2013; Brückner and Ciccone, 2011). However, other authors express doubts over whether rainfall and temperature provide relevant instruments arguing that the correlation is not sufficiently strong (e.g. Koubi et al., 2012).

Due to the complexity of the relationship between economic growth and climate,

[3] In recent years, nighttime lights have been frequently employed as proxies for economic activity. See, for example, Michalopoulos and Papaioannou (2013) and Michalopoulos and Papaioannou (2015), Elliott, Strobl, and Sun (2015), as well as Hodler and Raschky (2014) in the context of civil conflicts.

[4] According to the World Bank's World Development Indicators, the average contribution of the agricultural sector to Sub-Saharan Africa's GDP amounts to 18.6% over 1992-2010 and the agricultural sector accounts for 65% of the labour force, see `http://go.worldbank.org/GUJ8RVMRL0` (data retrieved on July 15, 2015). Furthermore, only 6% of the continent's food production is irrigated (Buhaug et al., 2015).

it is not obvious which functional form captures the link adequately. For instance, while agricultural production is expected to be increasing in precipitation, very high rainfall levels may be associated with extreme weather conditions and could have adverse effects. Estimation results thus depend on how weather variables are defined and how extreme weather events, such as droughts, are accounted for. Various climate variable specifications have been employed in the literature—including inter-annual rainfall growth (Miguel, Satyanath, and Sergenti, 2004; Weezel, 2015), rainfall levels (Ciccone, 2011; Ciccone, 2013), deviations from the long-run level (Koubi et al., 2012) and drought dummy variables (Aidt and Leon, 2015). The variety of climate instrument specifications may be one reason for the observed variations and lack of robustness in estimation results. Furthermore, the estimation methodology should accommodate that the economic growth-climate relationship is likely to vary fundamentally across African regions due to heterogeneous soil and climate characteristics.

This study employs the Lasso estimator due to Tibshirani (1996) to address the challenges arising from heterogeneity, non-linearity and weak identification. Belloni et al. (2012) and Belloni et al. (2016) propose the use of the Lasso estimator to obtain first-stage predictions from a very large set of putative instruments in order to approximate the true, complex functional relationship. In this way, the Lasso serves as a data-driven method of generating optimal instruments, while at the same time accounting for non-linearities and spatial heterogeneity. A major advantage of this approach is that the estimation is less susceptible to the researcher's choice and specification of instruments, thereby facilitating transparency and robustness of empirical results. The Lasso-based approach is also related to Couttenier and Soubeyran (2014), who make use of the PDSI drought index due to Palmer (1965) which captures local climate conditions. In this context, the Lasso instruments can be interpreted as a climate index, which is motivated from econometric theory and the literature on optimal instruments.

Commodity price changes provide another exogenous source of variation in economic conditions (Dube and Vargas, 2013; Bazzi and Blattman, 2014; Brückner and Ciccone, 2010), which I exploit in this study. Since commodity prices may affect civil conflicts through channels other than income (in particular, via state revenues), commodity prices are not used as an instrument for economic growth, but are treated as exogenous regressors (Bazzi and Blattman, 2014). An advantage of exploiting commodity price data is that it allows to identify the effect of prices on civil conflict for different commodity classes. This is especially important since, as pointed out by Dal Bó and Dal Bó (2011), conflict risk may be decreasing in prices for labour-intensive goods, but increasing in prices for capital-intensive commodities—a view which is empirically supported by Dube and Vargas (2013).

Another focus of this study is on the spatial dimension of conflicts. It is a well known empirical phenomenon that civil conflicts are clustered in space (e.g. Buhaug and Gleditsch, 2008). Thus, instead of treating sub-national areas as isolated units, the spatial econometric model employed allows conflict risk to depend on conflicts in spatially close areas. Note that, in this study, 'space' does not only refer to geographic distance. Specifically, spatial weights matrices based on political and ethnic distance are considered, providing new insights into the drivers of conflict diffusion on the sub-national level.

Estimation results suggest that rainfall and temperature have a significant impact on economic growth in African first-order administrative areas. The link is especially strong when considering Lasso instruments which seem to capture the complexity of the relationship much more effectively. There is, however, no evidence that predicted economic growth has a causal effect on civil conflict onset. Furthermore, positive price changes of capital-intensive commodities raise the risk of conflict outbreak, providing support for rapacity as a conflict driver, whereas labour-intensive goods do not seem to matter. The most striking pattern overall are strong conflict diffusion effects, in particular within countries.

This chapter is structured as follows. The next section summarises the theoretical literature on conflict, economic conditions and conflict diffusion. Section 2.2 gives an overview of related empirical studies. Section 2.3 describes the data set and Section 2.4 introduces the estimation methodology. In Section 2.5, economic growth predictions using nighttime light data are obtained. Section 2.6 presents estimation results. Concluding remarks are in Section 2.7.

2.1 Theoretical background

2.1.1 Economic shocks and conflict

Rational choice theories predict two opposing effects of the level of income on civil conflict. According to *opportunity cost theories*, which have their roots in rational choice theories of crime (Becker, 1968), countries with relatively low income levels have a higher propensity to civil conflicts due to low opportunity cost on the individual level, rendering violent activities relatively attractive (Collier and Hoeffler, 1998; Collier and Hoeffler, 2004; Collier, Hoeffler, and Rohner, 2009). In contrast, the *state as a prize*, or *rapacity mechanism*, predicts that the higher national income, the higher are expected returns from rebellion (Grossman, 1995; Fearon, 2008). The two theories are not contradictory and may operate simultaneously. Which mechanism dominates in the short run is likely to depend on the distribution of income and the type of income changes over time. For instance, income changes induced by rainfall and temperature shocks will presumably have the most profound effect on low income subsistence farmers with no access to irrigation techniques. A significant negative coefficient on economic growth instrumented by climate variables would therefore support opportunity cost theories.

The view that not only the direction, but also the type of income changes matters is supported by Dal Bó and Dal Bó (2011), who link the opportunity cost and the rapacity mechanism by developing a formal model of an economy with a labour-intensive, a capital-intensive and an unproductive appropriation sector. The model suggests that a positive income shock may increase conflict risk if the change favours capital-intensive industries, which implies a reduction of relative wages in the labour-intensive sector. Thus, the model of Dal Bó and Dal Bó (2011) predicts a negative coefficient on labour-intensive commodity prices and a positive coefficient on capital-intensive commodity prices.

The temporal dimension is also crucial for understanding the income-conflict relationship. Chassang and Miquel (2009) argue that opportunity costs, rather than the rapacity mechanism, drive the relationship between economic growth and conflict in the short run. Intuitively, since the value of assets and the stock of wealth are less volatile than wages in the short run, the 'prize' remains more or less equally attractive, whereas opportunity costs may drop substantially during an economic crisis. Since this study focuses on short-run economic shocks, the argument by Chassang and Miquel (2009) would suggest that the opportunity cost channel dominates the rapacity mechanism.

Another channel through which income may affect conflict stresses the importance of the *state's capacity* to prevent or repress insurgence, which is argued to be related to national income via state revenues (Fearon and Laitin, 2003). If state revenues depend highly on natural resource rents, a rise in prices for export commodities may improve fiscal conditions, strengthen the state's counter-insurgence capacities and thereby diminish conflict risk. Thus, a negative coefficient on prices for capital-intensive goods would support the state capacity theory, while a positive coefficient would provide direct evidence for the rapacity mechanism.

2.1.2 Spatial dimension of conflict

A common empirical pattern is the clustering of violence in space. The existence of civil conflict clusters as such, however, does not imply a causal relationship between conflict events, since many conflict determinants, such as income, typically also exhibit spatial clustering. Various mechanisms through which conflicts may spread are discussed in the literature (e.g. Lake and Rothchild, 1998; Buhaug and Gleditsch, 2008).

First, conflicts can induce a direct change in the balance of power, the availability of resources or the socio-economic conditions in another area. For instance, Murdoch and Sandler (2004) show that conflicts have a substantial negative impact on economic growth, not only in the conflict country, but also in nearby countries through the negative effects on trade and investment, which in turn may make nearby areas more susceptible to civil conflict. In addition, cross-border population movements may foster conflict diffusion by altering the ethnic composition in the host country, exerting pressure on economic conditions and creating tensions between host and refugee population (Salehyan and Gleditsch, 2006). Furthermore, conflicts may facilitate the trade of arms, illicit drugs and the spread of diseases across borders (Blattman and Miguel, 2010).

The second mechanism, often referred to as *demonstration effect*, is due to the spread of information. While observing conflict events in other areas, groups update their beliefs and expectations towards the feasibility of rebellion or appropriation (Lake and Rothchild, 1998). This effect is expected to be particularly strong if there exist ethnic ties between these areas or if the areas are in geographical proximity, which is likely to facilitate the spread of information (Buhaug and Gleditsch, 2008).

Thirdly, independent of whether the focus is on countries or sub-national areas as in this study, it is important to stress that geographical entities are not isolated from each other in a political sense. The observation of conflict diffusion may reflect that the underlying issues causing the civil conflict affect groups on both sides of the border due to economic, political or ethnic ties (Salehyan and Gleditsch, 2006). Since civil conflicts are in most cases not confined to a particular sub-national area, this mechanism is likely to be particularly strong in the sub-national data set examined in this study.

2.2 Previous empirical studies

It is well known that the economic growth-conflict relationship suffers from reverse causality. While economic shocks may trigger conflicts, violence or even the prospect thereof are likely to adversely affect economic growth. In a seminal study, Miguel, Satyanath, and Sergenti (2004) instrument GDP growth with rainfall shocks, which they define as the percentage change of rainfall from the previous year. The authors find significantly positive coefficients on contemporaneous and lagged rainfall shocks in the first stage with GDP growth as the dependent variable. Results from Instrumental Variables (IV) estimation using a sample of African countries covering 1981-1999 suggest that a 5 percent drop in GDP causes the likelihood of a civil war to increase by 10 percentage points. This approach is critically discussed by Ciccone (2011) and Ciccone (2013) who argues that, because rainfall is strongly mean-reverting, a specification using rainfall in levels is more appropriate. In a response, Miguel and Satyanath (2010) and Miguel and Satyanath (2011) justify the use of rainfall shocks, arguing that economic actors often react to changes in economic conditions, and also show that main results do not change when using rainfall in levels. Brückner and Ciccone (2010) show that the identification strategy in Miguel, Satyanath, and Sergenti (2004) is not robust to the inclusion of time effects and Jensen and Gleditsch (2009) point out that the exclusion of countries involved in civil wars in other states alters the results.

The work of Miguel, Satyanath, and Sergenti (2004) has sparked numerous studies employing similar identification strategies. For example, Bergholt and Lujala (2012) exploit climate-related natural disasters to identify the causal effect of growth on conflict. Their results suggest that natural disasters have a strong effect on economic growth, but economic shocks induced by disasters are unrelated to conflict. Koubi et al. (2012) analyse both the climate-economic growth and the economic growth-conflict link, but find that precipitation and temperature variation do not determine economic growth. According to Weezel (2015), on the other hand, rainfall anomalies have a significant effect on economic growth in the agricultural sector, but the link between economic growth and conflict onset is only weak. Similar to this study, Hodler and Raschky (2014) analyse a sub-national dataset of 5,689 subnational African regions and establish a link between lagged economic shocks and civil conflicts. However, the authors do not account for spatial effects and interpret nighttime lights as a proxy for economic activity instead of obtaining economic growth predictions explicitly.

Against the background of climate projections predicting higher average temperatures and extremer weather conditions (Stern, 2007; IPCC, 2007), there exist a rich literature focusing on the direct impact of climate on conflict (Hendrix and Salehyan, 2012; Fjelde and Uexkull, 2012; O'Loughlin et al., 2012; Hendrix and Glaser, 2007; Raleigh and Kniveton, 2012). For instance, Burke et al. (2009) predict that the increase in conflict due to global warming will results in 390,000 additional battle deaths in sub-Saharan Africa by 2030. These results are contested by Buhaug (2010) who argues that the association between climate and conflict is not robust when controlling for structural variables such as economic development. Harari and La Ferrara (2013) apply spatial econometric models to a panel of sub-national African cells and find that adverse climate shocks taking place during the growing season raise the risk of conflict events.

Another strand of literature exploits commodity price shocks to investigate the link between economic conditions and conflict. The identification strategy relies on the assumption that international commodity prices are not affected by civil conflicts in exporting countries. Brückner and Ciccone (2010) construct an export-weighted commodity price index for sub-Saharan African countries in 1981-2006. Three-year commodity price growth is shown to be significantly related to conflict; both when used as a regressor in a fixed effect estimation and when used as an instrument for economic growth. Dube and Vargas (2013) examine coffee and oil prices shocks in Columbia and show that a fall in coffee prices increases conflict risk, while a negative oil price shock reduces conflict risk, consistent with the view that conflict risk is decreasing in the price of labour-intensive commodities, but increasing in the price for capital-intensive goods (Dal Bó and Dal Bó, 2011). According to Bazzi and Blattman (2014), there is no conclusive evidence that commodity price shocks trigger new conflicts and only limited support for the view that positive shocks promote the likelihood of conflict ending. Berman and Couttenier (2015) relate world demand for agricultural products produced within 0.5×0.5 degree cells to conflict in Africa. They find that conflict intensity and onset are negatively correlated with income shocks, and attribute the link to the opportunity cost mechanism.

The spatial dimension of violence has attracted more attention in recent years. Most prominently, Buhaug and Gleditsch (2008) investigate spill-over effects and find transnational-ethnic ties to be particularly important. Jensen and Gleditsch (2009) re-estimate the model from Miguel, Satyanath, and Sergenti (2004), adding a spatial conflict lag into the estimation equation. The authors discover that the absolute size of the coefficient estimate on economic growth decreases when spatial dependence is accounted for, which highlights the importance of incorporating spill-over effects.

2.3 Data

Data for the dependent variable is taken from the Uppsala Conflict Data Program's Georeferenced Event Dataset (UCDP GED) v.1.9-2015 (Sundberg and Melander, 2010; Sundberg and Melander, 2013). The UCDP GED provides a list of geo-coded violent events in Africa covering 1989-2010. An event is defined as:

> The incidence of the use of armed force by an organised actor against another organized actor, or against civilians, resulting in at least 1 direct death in either the best, low or high estimate categories at a specific location and for a specific temporal duration.
>
> Sundberg and Melander, 2013, p. 524

Each event occurs as part of a larger-scale conflict that, by definition, exceeds a threshold of 25 battle-related deaths in at least one calendar year. The UCDP has collected information on location, timing and the type of conflict for each event, as well as a high, a low and a best casualty estimate. Conflicts are divided into state-based, non-state based and one-sided disputes. If a formally organised group is involved in a violent incident with a state-based actor, the conflict type is coded as state-based (11,137 events). If none of the parties is governmental, but both actors are formally organized, the conflict type is non-state (3,382 events). Accordingly, if only one party is formally organized and violence against non-organized civilians is involved, the conflict is denoted as one-sided (6,838 events). The precision of geo-referencing varies from exact coordinates to events than can only be assigned to a whole country. Events that cannot be related to first-order administrative units are discarded. This affects 1,265 of 21,189 events (6.0%) over the 1992-2010 period. Furthermore, since this study focuses on civil (or intra-state) conflicts, events associated with inter-state disputes are removed (43 events). In the baseline specification, I define a conflict as 'active' if the UCDP GED records at least one battle-related death according to the 'best' estimate in a given year.

The precipitation and temperature data is compiled by Willmott and Matsuura (2013) and downloaded from NOAA/OAR/ESRL PSD in a suitable data format (i.e., NetCDF).[5] The authors have generated a 0.5×0.5 degree global dataset based on 20,782 weather stations which record monthly total precipitation throughout the years 1901-2010. Precipitation is measured in centimetres and temperature in degrees Celsius (°C).

Nighttime light data is made publicly available by the NOAA's National Geophysical Data Center (NOAA-NGDC).[6] The NOAA-NGDC processes raw satellite data from the United States Air Force Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS). The DMSP-OLS's satellites collect light data on a daily basis between 8.30pm and 10pm local time, and the light intensity is recorded on a scale from 0 to 63. There is data from one satellite per year for 1992-1993, 1995-1996 and 2008-2010, and two satellites for the remaining years. The resolution is 30×30 arc seconds, which corresponds to approximately 926×926 metres at the equator. The NOAA-NGDC identifies observation distorted by sunlight, moonlight, clouds, auroral activity and forest fires and the remaining observations are used to obtain annual averages for each 30×30 arc second pixel and each satellite-year.[7] The final product is a raster image in TIF format for each satellite-

[5] See `http://www.esrl.noaa.gov/psd/data/gridded/data.UDel_AirT_Precip.html`. The data was downloaded on July 1, 2015.

[6] Retrieved from `http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html` on July 1, 2015.

[7] Another source of background noise arises from gas flaring which occurs during oil production. The NOAA-NGDC does not exclude observations affected by gas flaring from the dataset. Elvidge et al. (2009a) provides a polygon dataset that can be used to exclude the locations where light emissions are predominantly from gas flaring. The correlation coefficient between average light intensity with and without excluding gas flaring is however close to one which is why, for simplicity, only average light intensity including gas flaring is considered.

	Obs.	Mean	Sd.	Min	Max	Median
Country-level						
Average light emissions	3471	4.29	8.02	0.00	61.71	1.23
Light growth in %	3274	5.95	23.90	-200.15	237.95	2.92
Log. of GDP, constant LCU	3361	26.06	3.38	16.92	36.47	26.37
Growth in GDP in %	3176	3.70	5.97	-69.81	91.62	3.89
Economic growth (predicted) in %	3274	3.75	3.52	-28.86	41.79	3.91
Area-level						
Incidence[1]	16131	0.15	0.36	0.00	1.00	0.00
Incidence[25]	16131	0.07	0.26	0.00	1.00	0.00
Incidence[50]	16131	0.05	0.23	0.00	1.00	0.00
Onset[1]	12977	0.06	0.24	0.00	1.00	0.00
Onset[25]	14120	0.04	0.19	0.00	1.00	0.00
Onset[50]	14442	0.03	0.17	0.00	1.00	0.00
Onset[1], state	13092	0.06	0.24	0.00	1.00	0.00
Onset[1], non-stated	14531	0.03	0.17	0.00	1.00	0.00
Onset[1], one-sided	14985	0.02	0.12	0.00	1.00	0.00
Ending[1]	2305	0.37	0.48	0.00	1.00	0.00
Casualty estimate	16131	30.98	942.30	0.00	108931	0.00
Average light emissions	15808	2.22	7.16	0.00	63.00	0.05
Log. of average light emissions	13213	-2.00	2.68	-10.72	4.14	-2.33
Light growth in %	12110	8.18	41.57	-325.50	408.94	4.42
Economic growth (predicted) in %	12110	4.16	6.45	-49.64	72.68	4.19
Rainfall	16036	7.72	4.96	0.00	31.16	7.68
Temperature	16036	23.25	4.25	0.95	31.51	23.78
Commodity price growth (labor) in %	15282	4.69	16.85	-49.12	75.12	5.36
Commodity price growth (capital) in %	15264	7.08	19.47	-44.99	70.33	8.36

Note: The numbers in brackets indicate the threshold used to code a conflict as active.

Table 2.1: *Summary statistics for the area-level and country-level data set.*

year, covering -180 to 180 degree longitude and -65 to 75 latitude. It is important to note that the light intensity as measured by satellites is not directly comparable across time and satellites due to different, time-varying satellite settings. The framework by Henderson, Storeygard, and Weil (2012) accounts for this by the use of year dummies, as will be discussed in the next section. For further information on nighttime light data, see Henderson, Storeygard, and Weil (2012), Doll (2008) and Elvidge et al. (2009b).

The construction of the export-weighted commodity price index follows Brückner and Ciccone (2010). International commodity prices are from the International Monetary Fund (IMF) and UNCTAD. The trade data used to obtain country-level export shares is downloaded from UNCTAD and averaged across 1995-2012. The effect of commodity price shocks on conflict through economic conditions is likely to substantially differ across commodity groups. While it is expected that some commodities have a strong impact on low-income households (e.g. annual crops), other commodities are likely to disproportionally affect capital owners and state rents (Dal Bó and Dal Bó, 2011; Bazzi and Blattman, 2014; Dube and Vargas, 2013). For this reason, commodities are divided into capital-intensive (e.g. oil, minerals) and labour-intensive commodities (e.g. food crops).[8]

Formally, the commodity price index for labour-intensive commodities is defined as $\sum_j \omega_{cj} P_{jt}$, where ω_{cj} is the time-invariant export share of labour-intensive commodity j and country c. The commodity price index for capital-intensive goods is defined accordingly. To account for the possibility that international commodity prices are influenced by civil conflicts in exporting countries, a threshold is applied such that ω_{cj} is set to zero if the world market share of country c for commodity j is greater than 10%. Since the interest lies in commodity price shocks, the annual percentage change is used in all regressions.

Climate, conflict and nighttime light data is matched with first-order administrative boundaries from the Natural Earth (NE) dataset.[9,10] The NE map reflects the present state of political boundaries on the earth and, thus, the NE dataset does not account for boundary changes over time. While this is clearly a limitation, it is unlikely to have a significant effect on results. There are in total 849 first-order administrative units for mainland Africa and Madagascar in the NE map. Table 2.1 shows summary statistics for the area and country-level data set, respectively. Figure 2.1 shows the development of conflict-related casualties along with the average nighttime light intensity over time.

[8] Labour-intensive commodities: coffee, chocolate, tobacco, cotton, tea, sugar, wheat, fish; capital-intensive commodities: iron, copper, aluminium, nickel, oil, uranium, gold, wood. Note that the list of commodities is based on Brückner and Ciccone (2010), but does not include bananas, livestock, phosphates and ground nuts due to missing data.

[9] Retrieved from `http://www.naturalearthdata.com/` on July 1, 2015.

[10] The data generation process was carried out in R using the package raster (Hijmans and Etten, 2014).

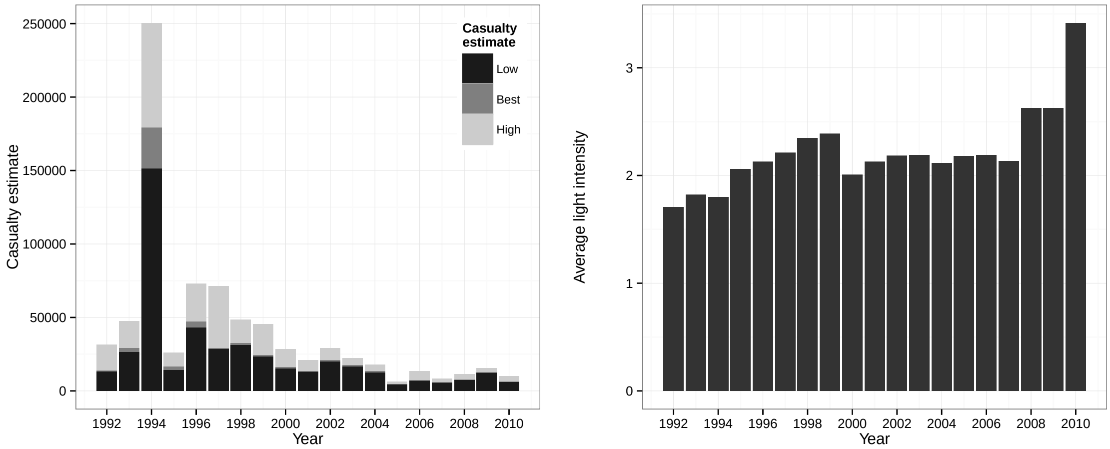


Figure 2.1: *Descriptive graphs. The left figure shows the aggregate low, best and high casualty estimate. The right figure shows light intensity averaged across areas.*

2.4 Estimation methodology

2.4.1 The econometric model

The spatial autoregressive panel model considered in this study is given by

$$o_{ict} = \rho \sum_{j=1}^{N} w_{ij} e_{jct} + \beta \hat{g}_{ict} + p_{L,ct}\gamma_L + p_{K,ct}\gamma_K + \mu_{ic} + \delta_t + \psi_c t + u_{ict}, \tag{2.1}$$

where i, c and t are the area, country and time index, respectively.[11] The variable o_{ict} is a binary conflict *onset* indicator and e_{jct} denotes conflict *incidence*. Formally, the conflict incidence indicator equals one if there is an active conflict in a given year, zero otherwise. The onset indicator, o_{ict}, is set to unity if there is an active conflict in area i and year t, but there was no active conflict in the previous year. If there is no active conflict in neither year t nor year $t-1$, the onset indicator is set to zero for year t. The use of conflict onset as the dependent variable, as opposed to conflict incidence, accounts for the temporal persistence of conflict, which would otherwise induce an estimation bias (Beck and Katz, 2011; Bazzi and Blattman, 2014).

[11] While the least square estimates of the linear probability model are in general inconsistent (Horrace and Oaxaca, 2006), the estimates from least squares tend to be a good approximations of the true marginal effects (Angrist and Pischke, 2009; Wooldridge, 2010). Furthermore, the linear probability model allows to make use of advanced IV methods, including the weak identification testing and the Lasso estimator.

Furthermore, $\hat{g}$ is economic growth predicted using nighttime light data (as discussed in Section 2.5). The IV approach aims at addressing both the measurement error arising from the prediction of economic growth as well as the endogeneity arising from reverse causality and omitted variables. The regressors $p_{L,ct}$ and $p_{K,ct}$ denote the growth rate of the export-weighted commodity price index for labour and capital-intensive commodities (as defined in Section 2.3). In addition, the model allows for area-specific unobserved heterogeneity (μ_{ic}), common time effects (δ_t) and country-specific time trends ($\psi_c t$).

The model assumes that conflict outbreak risk in area i depends on conflict outcomes in other areas through a weighted average, which is referred to as a spatial lag. Note that the spatial lag is applied to the conflict incidence indicator, not to conflict onset, as ongoing conflicts may trigger the outbreak of conflicts in other areas. The spatial autoregressive parameter, ρ, reflects the strength of spill-over effects and provides insights about the drivers of conflict diffusion. Since conflict outcomes are simultaneously determined, the spatial lag of the dependent variable is endogenous. Kelejian and Prucha (1998) suggest spatially lagged exogenous explanatory variables as instruments for the spatial lag. In this application, the estimation of equation (2.1) is more complicated since the main regressor, economic growth, is endogenous. Thus, spatial lags of economic growth are not available as instruments for the spatial conflict lag. However, spatial lags of rainfall and temperature provide suitable instruments and allow for identification of the model parameters.

The spatial weights, w_{ij}, are specified based on geographic, political and ethnic distance. First, the inverse distance matrix is defined as $w_{ij} = 1/d_{ij}$ where d_{ij} is the geographic distance between the centroids of area i and j. Thus, the implicit assumption of the inverse distance matrix is that the interaction between area i and j is decreasing in geographic distance. Second, since civil conflicts are often fought on the national level, strong within-country spill-over effects are expected. The country matrix captures these political spill-overs effects: $w_{ij} = p_j$ if i and j are in the same country, zero otherwise. The inclusion of the population count, denoted by p_j, accounts for that, *ceteris paribus*, it is expected that the larger the population size of area j, the greater the impact of area j on area i.[12] The neighbour weights matrix is a slight modification of the country matrix: $w_{ij} = p_j$ if i and j are in the same country or in contiguous countries, zero otherwise. Hence, the neighbour weights matrix also captures spill-over effects across country borders. A comparison of the country matrix and neighbour matrix allows to assess the size of cross-border diffusion processes relative to within-country spill-overs. Lastly, the ethnic weights matrix is considered: If i and j are populated by at least one common ethnic group, $w_{ij} = p_j$, zero otherwise. The binary ethnic matrix is obtained based on the Geo-Referencing of Ethnic Groups (GREG) dataset by Weidmann, Rød, and Cederman (2010) who use the classical *Atlas Narodov Mira* (1964) to generate maps of ethnic groups.[13] Note also that, as standard in the spatial econometrics literature, all spatial weights matrices are row-standardised prior to generating spatial lags.

2.4.2 The Lasso and instrument selection

The relationship between weather variables and economic growth is complex and it is not obvious which set of instruments is appropriate for the purpose of identifying the effect of economic shocks on conflict risk. Accounting for extreme weather conditions, non-linearities and heterogeneity across regions gives rise to a large number of legitimate instruments—making it difficult, if not impossible, to select the 'correct' set of instruments on theoretical grounds. Following the seminal work of Belloni et al. (2012), I employ the Lasso estimator in a first step to generate optimal instruments from a large set of putative instruments. In the second step, the instruments generated by the Lasso are used in a standard IV regression.

Suppose the true relationship between economic growth predicted by nighttime lights and climate is given by

$$\hat{g}_{ict} = h(w_{ict}) + \xi_{ict}, \tag{2.2}$$

[12] Population count estimates used for the construction of spatial weights matrices are from CIESIN/FAO/CIAT (2005) and refer to the pre-sample year 1990.

[13] For a critical discussion, see Bridgman (2008).

where I omit additive time effects, year effects and country-specific time trends for notational convenience.[14] w_{ict} represents local weather conditions and $h(\cdot)$ is an unknown, potentially non-linear function linking economic growth with the weather. Let $\mathbf{z}_{ict}$ denote a vector consisting of a large number of putative instruments. Specifically, $\mathbf{z}_{ict}$ includes transformations of precipitation and temperature, such as temporal and spatial lags as well as interaction terms accounting for heterogeneity and non-linearities. Under the assumption that there exists a sparse linear approximation of $h(w_{ict})$ and that the approximation error, $r(w_{ict}) = h(w_{ict}) - \mathbf{z}'_{ict}\boldsymbol{\pi}$, is sufficiently small in large samples, the Lasso estimator can obtain approximately optimal instruments.

Formally, the objective function of the Lasso estimator is given by

$$\min_{\boldsymbol{\pi}} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{g}_{ict} - \mathbf{z}'_{ict}\boldsymbol{\pi})^2 + \frac{\lambda}{nT}\sum_{j=1}^{p}\phi_j|\pi_j|. \tag{2.3}$$

The first term is the residual sum of squares and the second term penalizes the absolute size of the coefficient estimates. λ denotes the penalty level and ϕ_j are penalty loadings, which accommodate heteroscedastic errors. Belloni et al. (2012) and Belloni et al. (2016) derive the optimal penalty level, $\lambda^\star$, which is the smallest penalty level that overrules the random part of the data-generating process, and suggest an algorithm for estimating the ideal penalty loadings.

Due to the penalization term, the Lasso sets some of the coefficients in $\boldsymbol{\pi}$ equal to exactly zero, thereby effectively selecting instruments.[15] For example, suppose the Lasso estimator approximates the link function $h(\cdot)$ as

$$\hat{h}_{ict} = \hat{\pi}_0\text{Rainfall}_{ict} + \hat{\pi}_1 D_{1,ict}\text{Rainfall}_{ict}, \tag{2.4}$$

[14] Year effects, fixed effects and country-specific time trends are partialled out prior to Lasso estimation.

[15] The objective function is closely related to the well-known Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978), which impose a penalty on the number of parameters. The advantage of the Lasso estimator is that fast algorithms are available, while AIC and BIC are computationally infeasible in this context, even if the number of putative instruments is reasonably small.

where the dummy variable $D_{1,ict}$ equals one if area i is in country 1, zero otherwise. Thus, this result would suggest that the effect of rainfall is the same for all countries except country 1, whereas other variables—including temperature as well as temporal and spatial lags—are assigned coefficients of zero. The predicted values, $\hat{h}_{ict}$, are then utilised as an instrumental variable. The approach can be easily extended to multiple endogenous regressors, implying that the spatial conflict lag can be instrumented in the same way.

The Post-Lasso is a simple extension of the Lasso estimator. It addresses the shrinkage bias arising from penalization by applying OLS to the regressors selected by the Lasso. Although both Lasso and Post-Lasso are valid for generating instruments, in the following only results based on the Post-Lasso are reported as it performs in most scenarios at least as good as the standard Lasso (Belloni et al., 2012).

In this application, the set of potential instruments in $\mathbf{z}_{ict}$, from which the optimal instruments are generated, includes (i) temporal lags of rainfall and temperature to account for lagged effects of weather conditions on the economy, (ii) spatial lags to allow for indirect effects via spatially connected areas, (iii) threshold dummy variables to capture extreme weather events,[16] (iv) country-specific climate interaction variables to allow for parameter heterogeneity across climate regions and (v) logged, squared and cubed precipitation and temperature levels to capture non-linearities as well as (vi) combinations of the above. The total set of potential instruments includes 1,388 variables.

[16]The rainfall drought dummy variables are set to one if the rainfall level is below the area-specific 10th, 20th and 30th percentile, respectively, zero otherwise. The temperature drought dummy variables are set to one if the temperature level is above the area-specific 70th, 80th and 90th percentile, respectively, zero otherwise.

	(1) ln(GDP)	(2) ln(GDP)	(3) Δln(GDP)	(4) Δln(GDP)	(5) Δln(GDP)	(6) Δln(GDP)	(7) Δln(GDP)
ln(Light)	0.254***	0.384***					-0.000794
	(0.0423)	(0.0967)					(0.00141)
Δln(Light)			0.0471***	0.165***	0.120***	0.166***	0.165***
			(0.0157)	(0.0560)	(0.0395)	(0.0565)	(0.0560)
$\Delta\ln(\text{Light})^-$					0.100		
					(0.0936)		
$\Delta\ln(\text{Light})^2$						-0.106	
						(0.0829)	
Observations	3333	866	3146	820	820	820	820
Countries	183	46	183	46	46	46	46
Sample	All	Africa	All	Africa	Africa	Africa	Africa
R^2	0.754	0.754	0.0872	0.163	0.169	0.177	0.163

Note: All models include year effects. Model 1 and 2 include fixed effects. Standard errors are in parentheses. Standard errors are robust to both arbitrary heteroskedasticity and arbitrary within-region correlation.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.2: *Predicting economic growth using nighttime lights.*

2.5 Predicting GDP with nighttime light data

The method for prediction of GDP using nighttime light data from satellites is based on Henderson, Storeygard, and Weil (2012). The authors consider different forms of

$$\ln(\text{GDP})_{ct} = \psi \ln(\text{Light})_{ct} + \alpha_c + \theta_t + \varepsilon_{ct}, \tag{2.5}$$

where $\ln(\text{GDP})_{ct}$ is the logarithm of GDP in levels as measured by national accounts and $\ln(\text{Light})_{ct}$ is the logarithm of average light intensity.[17] θ_t accounts for variations in satellite settings across time as well as time-specific economic and technological conditions. α_c controls for country-specific unobserved heterogeneity due to cultural and economic characteristics.

Table 2.2, Model 1 corresponds to Table 2, Model 1 in Henderson, Storeygard, and Weil (2012).[18,19] The coefficient on average light intensity is 0.254 which suggests that a 1% increase in light intensity is associated with a 0.254% rise in GDP. Note that the point estimate is similar to the point estimate in Henderson, Storeygard, and Weil (2012) (0.277, with a standard error of 0.031). Model 2 shows that the coefficient on light emission is substantially higher in African countries.

[17] Worldwide GDP data for the prediction of GDP for African first-order administrative units is from the World Bank and in local constant currency. The World Bank data set was downloaded on July 7, 2015 using the R package `WDI` from Vincent Arel-Bundock.

[18] All regression results in this section were obtained using Stata 12 and xtivreg2 (Schaffer, 2012).

[19] Following Henderson, Storeygard, and Weil (2012, fn. 16), Bahrain, Singapore, Equatorial Guinea and Serbia and Montenegro are excluded from the sample. In addition, Norway and Estonia are excluded due to data reliability issues.

The interest of this study lies in the effect of economic growth on conflict. Thus, it seems natural to consider equation (2.5) in first differences,

$$\Delta\ln(\text{GDP})_{ct} = \psi\,\Delta\ln(\text{Light})_{ct} + \theta_t' + \varepsilon_{ct}'. \tag{2.6}$$

A major practical advantage of using (2.6) rather than (2.5) is that the latter does not require estimating the fixed effects for prediction purposes. Hence, it is straightforward to obtain estimates for countries or areas for which no GDP data is available. Model 3-7 in Table 2.2 show estimation results. Model 3 uses the full sample. Model 4-7 are based on Africa only. The coefficient on the log-difference of nighttime light in Model 4 is significantly larger, suggesting that in Africa nighttime lights are less responsive to changes in income than in the rest of the world. For this reason, estimates of economic growth are based on the African sample only.

A concern for the purpose of this study is that the relationship between income and light growth is, due to fixed installation costs, asymmetric in the sense that light is more sensitive to positive growth than to negative economic growth. Model 5 in Table 2.2 shows that the coefficient on negative light growth is not significantly different from the coefficient on positive light growth (see also Table 3, Model 3 in Henderson, Storeygard, and Weil 2012). Model 6 includes squared light growth and Model 7 includes the logarithm of light emission in levels to account for non-linearities. However, both variables are insignificant. Therefore, Model 4 is the preferred model for estimating economic growth using nighttime lights. Formally, predicted economic growth is defined as

$$\hat{g}_{ict} \equiv \hat{\psi}\,\Delta\ln(\text{Light})_{ict} + \hat{\theta}_t'. \tag{2.7}$$

It is well known that conflicts and economic growth measured by national accounts are negatively correlated. If economic growth estimated by nighttime lights is a good proxy for true growth in economic activity, violence should also be reflected in nighttime lights and predicted economic growth. Henderson, Storeygard,

	(1) Growth (pred.)	(2) Growth (pred.)	(3) Growth (pred.)
Incidence	-1.465***	-1.826***	-2.265***
	(-5.93)	(-4.93)	(-5.34)
Casualty threshold	1	25	50
Observations	12099	12099	12099

Note: All models include fixed effects. t-statistics are shown in parentheses. Standard errors are robust to both arbitrary heteroskedasticity and arbitrary within-region correlation.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: *Conflict incidence and economic growth. The results indicate that (predicted) economic growth is lower by 1.5–2.3 percentage points in conflict years.*

and Weil (2012, Fig. 4) show that the Rwandan genocide was associated with a drop in GDP as estimated by nighttime light data (see Figure 2.2).[20] The fixed effects estimation in Table 2.3 is a formal test of whether estimated economic growth is, conditional on fixed effects and year effects, significantly different in conflict years. A significant different mean in conflict years may be interpreted as evidence that violence is reflected in nighttime lights, which in turn supports the use of nighttime lights as a predictor for economic growth. The independent variable is conflict *incidence* with a conflict threshold of 1, 25 and 50 battle deaths, respectively. In all three specifications, the null hypothesis that the conditional mean in conflict years is equal to the conditional mean in peace years is rejected. The results suggest that average income growth in conflict years is lower by between 1.47 and 2.27 percentage points.

2.6 Results

2.6.1 Economic growth, temperature and rainfall

At first, it is insightful to examine the relationship between economic growth predicted using nighttime lights and climate variables. With respect to rainfall, I expect that, all other things equal, higher rainfall levels are associated with higher output due to favourable conditions for agricultural production. However, very high rainfall

[20]For a comparison of nighttime lights before and after the Syrian Civil War, see 'Syria's drained population', published online by *The Economist* on September 30, 2015. `http://www.economist.com/blogs/graphicdetail/2015/09/daily-chart-18` (accessed on October 4, 2015).

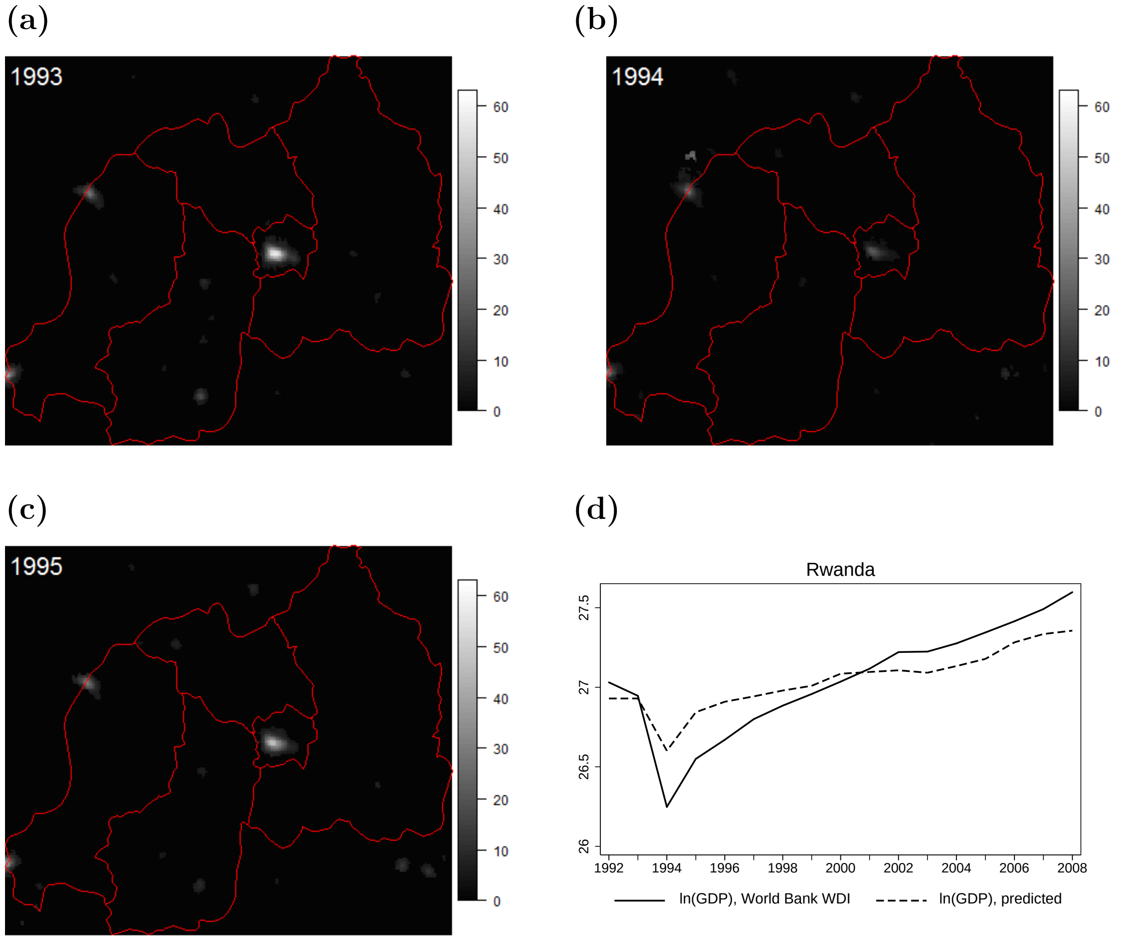


Figure 2.2: *Luminosity and the Rwandan genocide. Figures (a)-(c) show light emissions of Rwanda in 1993-1995, overlayed by first-order administrative borders. Note that the capital Kigali emits a significantly lower level of light during the genocide in 1994. Figure (d) compares official GDP data with GDP predicted using nighttime lights. The illustration follows Henderson, Storeygard, and Weil 2012, Fig. 4.*

	(1) Growth (pred.)	(2) Growth (pred.)	(3) Growth (pred.)	(4) Growth (pred.)
Rainfall	0.00103***	0.000953	-0.00432*	0.00124
	(0.000248)	(0.00136)	(0.00248)	(0.00168)
Rainfall^2	-0.0000683***	-0.000165***	0.0000967	-0.000239***
	(0.0000141)	(0.0000574)	(0.000134)	(0.0000750)
Temperature	-0.000676	0.0127**	0.00550	0.0201**
	(0.00112)	(0.00523)	(0.0145)	(0.00960)
Temperature^2	0.0000320	-0.000219*	-0.0000853	-0.000363
	(0.0000251)	(0.000119)	(0.000335)	(0.000228)
F-test, p-value (Rainfall)	0.000	0.000	0.034	0.000
F-test, p-value (Temp.)	0.000	0.004	0.682	0.007
F-test statistic (all)	23.088	11.541	2.665	12.601
Fixed effects	No	Yes	Yes	Yes
Year effects	No	Yes	Yes	Yes
Country-trends	No	No	Yes	Yes
Spatial Matrix	–	–	–	Country
Observations	9566	12009	2454	12009

Note: Standard errors are in parentheses. Standard errors are robust to both arbitrary heteroskedasticity and arbitrary within-region correlation.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: *Economic growth, rainfall and temperature.*

levels may reflect extreme, adverse weather conditions which suggests a concave relationship between economic growth and rainfall in levels. In Model 1 in Table 2.4, the coefficient on rainfall is significantly positive and the coefficient on squared rainfall is significantly negative, consistent with the notion that very high rainfall levels are associated with adverse weather conditions. Model 2 controls for year effects and fixed effects which renders rainfall in levels insignificant. Model 3 includes country-specific time trends. The F-tests indicate that rainfall and squared rainfall are jointly significant in all three specifications at the 5% level.

The economic rationale for the relationship between temperature and economic growth is also complex. Heal and Park (2013) and Burke, Hsiang, and Miguel (2015) argue for an inverted u-shaped relationship with a single peak around the agricultural and physiological optimum temperature. Physiological studies have shown that human performance significantly deteriorates if temperatures are very high (e.g. Wendt, Loon, and van Marken Lichtenbelt, 2007). Looking at the estimation results, temperature and squared temperature are separately insignificant in Model 1. The F-test, however, shows that temperature and squared temperature are jointly highly significant. Temperature and squared temperature are also jointly significant in Model 2, but not in Model 3, which accounts for country-specific time trends.

An advantage of spatial econometric methods is that additional instruments become available. Economic growth may not only be directly affected by local weather conditions, but also indirectly through the adverse impact on economic output in spatially close areas. Presumably, areas with high population density but negligible agricultural production (i.e., cities and metropolitan areas) are predominantly affected by weather shocks through the impact on spatially close areas with significant agricultural production. Model 4 regresses economic growth on spatially lagged climate variables where the country weights matrix is used. The F-tests indicate that spatial climate lags significantly determine economic growth in African areas.

It is interesting to note that the effect of rainfall on economic growth is predominantly driven by the adverse effect of extreme rainfall levels, and the effect of temperature on economic growth is predominantly driven by the positive effect of temperature. The latter is in contrast to Dell, Jones, and Olken (2008) who consider a linear specification and find that a 1°C increase in the average temperature reduces economic growth by 1.1 percentage points in a sample of low-income countries. However, as pointed out, the relationship is likely to be concave and whether the positive or the negative effect dominates may depend on the data sample. As stated by Hsiang, Burke, and Miguel (2013, p. 8), "the curvature is not apparent in every study, probably because the range of temperatures [...] contained within a sample may be relatively limited." Another explanation for the positive effect of temperature on predicted economic growth is that the relationship between temperature and nighttime light growth may be different to the relationship between temperature and economic growth measured by national accounts. For instance, high temperatures may lead individuals to shift social and economic activities from daytime to nighttime, causing an increase in nighttime light emissions.

There remains strong doubt about whether the quadratic relationship in Table 2.4 provides a reasonable representation of the climate-economic growth relationship. Heterogeneity across climate regions is not accounted for and extreme weather events may not be captured sufficiently by the quadratic relationship. For this reason, I utilise the Lasso estimator in Section 2.6.3 to approximate the true functional relationship from a large number of transformed climate variables. Despite these limitations, there is reasonably strong evidence that economic growth is—in some way—affected by climate variables. Rainfall and squared rainfall are jointly significant in all specification and temperature is only insignificant in Model 3.

2.6.2 Predefined instruments

Estimation in this sub-section is by two-step efficient GMM based on a predefined set of instruments, which comprises of rainfall, squared rainfall, temperature, squared temperature as well as spatial lags thereof.[21] Note that the instrument set is not selected in the belief that the quadratic relationship captures the climate-growth relationship appropriately. The results in this section serve rather as a point of comparison for the Lasso-based estimations presented in the next section.

Commodity price growth is treated as exogenous and not as an instrument for economic growth. The reason is that, as pointed out by Bazzi and Blattman (2014), commodity price growth may affect conflict through channels other than income, in particular through state revenues. The dependent variable is conflict onset which, in the main specifications, is defined using a casualty threshold of 1 (as indicated in brackets). There are three statistical tests reported at the bottom of Table 2.5. First, the Kleibergen-Paap rank test (Kleibergen and Paap, 2006), which evaluates the overall strengths of instruments. As demonstrated by Bound, Jaeger, and Baker (1995), the IV/GMM estimator may be severely biased in finite samples towards OLS if the correlation between endogenous regressors and instruments is only weak, highlighting the importance of weak identification testing. Secondly, the Sanderson-Windmeijer conditional F-test is reported for economic growth and the spatial conflict lag, respectively, which allows to assess the strength of instruments for each endogenous regressors individually (Angrist and Pischke, 2009; Sanderson and Windmeijer, 2016). For both tests, the Stock and Yogo (2005) critical values

[21] IV/GMM regressions in this section were conducted in Stata using the command `xtivreg2` (Schaffer, 2012). Lasso and Post-Lasso estimation were conducted in R using own code based on the package `glmnet` (Friedman, Hastie, and Tibshirani, 2010).

	(1) Onset[1]	(2) Onset[1]	(3) Onset[1]	(4) Onset[1]	(5) Onset[1]
Spatial Lag		0.516**	0.395**	0.316	0.724
		(0.238)	(0.199)	(0.220)	(0.839)
Economic growth, t	-0.722	-0.339	-0.456	-0.131	-0.0739
(predicted)	(0.681)	(0.564)	(0.619)	(0.745)	(0.624)
Commodity price growth, $t-1$	-0.0397	-0.0209	-0.0102	-0.0229	-0.0281
(labor-intensive goods)	(0.0243)	(0.0260)	(0.0285)	(0.0273)	(0.0289)
Commodity price growth, $t-1$	0.0429**	0.0448**	0.0627***	0.0489**	0.0537**
(capital intensive)	(0.0200)	(0.0194)	(0.0207)	(0.0197)	(0.0241)
Spatial Matrix		Country	Neighbour	Ethnic	Distance
Kleibergen-Paap Wald F-test	6.039	3.079	3.880	2.308	3.582
SW F-test: Growth		3.615	4.467	2.862	4.115
SW F-test: Spatial lag		5.992	12.14	2.974	7.697
Hansen-Sargan (Df.)	3	6	6	6	6
Hansen-Sargan (p-val.)	0.570	0.358	0.935	0.884	0.844
Observations	9727	9727	9727	9727	9727

Note: All models include fixed effects, year effects and country-specific time trends. Standard errors are in parentheses. Standard errors are robust to both arbitrary heteroskedasticity and arbitrary within-region correlation.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: *GMM estimation with predefined instruments.*

apply. Thirdly, the Hansen-Sargan test inspects the validity of moment conditions.

Model 1 in Table 2.5 is non-spatial, while subsequent models include spatial conflict lags. Economic growth predicted using nighttime lights is not significant in any of the specifications. The coefficient on the country weights matrix is 0.516 and significant at the 5% level. While the country weights matrix captures only within-country spill-over effects, the neighbour matrix also captures spill-over effects across country borders. The spatial lag on the neighbour matrix is also significant, but smaller, indicating that, as expected, spill-over effects are stronger within than across country borders. For comparison, spill-over effects across ethnic ties are insignificant and weaker with a coefficient of 0.316. The spatial autoregressive coefficient in Model 5, which is based on the inverse distance matrix, is also not significant.

The estimation results in Table 2.5 suggest that lagged commodity price growth of capital-intensive commodities raises the risk of conflict outbreak. This finding supports the rapacity mechanism and is in line with, e.g., Dube and Vargas (2013), who find that a rise in oil prices increases the risk of violence, but is in contrast to the state capacity mechanism—which is empirically supported by Bazzi and Blattman (2014). Labour-intensive commodities, on the other hand, do not seem to affect conflict outbreak. Furthermore, Tables B.1 and B.2 in the Appendix show that lagged economic growth and contemporaneous commodity price changes are not associated with conflict onset.

However, the results of the GMM estimation should be treated with caution for at least two reasons. First, the results rely on a predefined set of instruments and, given the large number of potentially relevant and valid instruments, are susceptible to the chosen set of instruments. Second, as indicated by the weak identification tests, the instruments identify the endogenous regressors only weakly, indicating that the estimation results may suffer from a non-negligible estimation bias in finite samples.

2.6.3 Data-driven instruments

Table 2.6 shows results when, following Belloni et al. (2012) and Belloni et al. (2016), the Post-Lasso estimator is used to obtain optimal instruments for both economic growth and the spatial conflict lag.[22] The large set of potential instruments, from which the optimal instruments are generated, account for parameter heterogeneity across countries, non-linearities and extreme weather conditions.

The weak identification tests indicate that the Post-Lasso estimator addresses the weak identification problem successfully. For instance, all Kleibergen-Paap rank test statistics exceed 40, compared to 2.3-6.0 when predefined instruments are employed. Across all model specifications there is again no indication that economic growth determines the risk of conflict outbreak. Moreover, lagged commodity price growth of capital-intensive goods has a significant effect on the risk of conflict outbreak, with coefficient estimates varying between 0.039 and 0.068. There is no robust statistical evidence that commodity price growth of labour-intensive goods is related to civil conflict, although the negative signs on the coefficients are consistent with the opportunity cost mechanism. In contrast to the GMM results above, the results suggest that there are significant spill-over effects along ethnic ties, which is in accordance with Buhaug and Gleditsch (2008). Spatial autoregressive coefficients are typically bounded between -1 and 1, whereas the coefficient on the inverse distance

[22]Note that the Hansen-Sargan tests for the IV Post-Lasso estimations are not reported, as the models are exactly identified.

	(1) Onset[1]	(2) Onset[1]	(3) Onset[1]	(4) Onset[1]	(5) Onset[1]
Spatial Lag		0.738***	0.478***	0.566***	1.375***
		(0.100)	(0.0959)	(0.115)	(0.378)
Economic growth, t	-0.225	0.136	-0.152	-0.0583	0.0629
(predicted)	(0.239)	(0.230)	(0.235)	(0.230)	(0.246)
Commodity price growth, $t-1$	-0.0448*	-0.0177	-0.00793	-0.00651	-0.0173
(labor-intensive)	(0.0243)	(0.0248)	(0.0265)	(0.0268)	(0.0255)
Commodity price growth, $t-1$	0.0390**	0.0484**	0.0643***	0.0572***	0.0677***
(capital-intensive)	(0.0196)	(0.0196)	(0.0202)	(0.0199)	(0.0202)
Spatial Matrix		Country	Neighbour	Ethnic	Distance
Kleibergen-Paap Wald F-test	87.64	42.42	43.49	43.40	40.33
SW F-test: Growth		84.70	87.09	87.71	81.33
SW F-test: Spatial lag		326.4	1123.3	395.0	494.1
Observations	9727	9727	9727	9727	9727

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: *IV Post-Lasso with data-driven instruments.*

matrix exceeds 1, presumably because the inverse distance matrix fails to capture the role of distance in diffusion processes appropriately.

Further estimation results are presented in Table 2.7. Models 1-3 disaggregate conflicts according to the three UCDP conflict types: state-based, non-state and one-sided conflicts. Model 4 and 5 code conflicts as active if a battle-related casualty threshold of 25 and 50 is exceeded, respectively. Model 6 takes conflict ending, as opposed to conflict onset, as the dependent variable.[23] To ease comparison, all models are based on the country weights matrix. Models 1-3 in Table 2.6 reveal that the significant effect of capital-intensive commodities is mainly driven by state-based conflicts, which is in line with the notion that increasing state revenues raise the expected profits from insurgence. The role of commodity prices is also not reflected in larger scale conflicts (Model 4 and 5) and commodity prices do not seem to affect conflict ending (Model 6). Spill-over effects are again significant across all specifications. Note that the spatial autoregressive coefficient is negative in Model 6, implying that conflict events in other areas decrease the likelihood of conflict ending. Tables B.3-B.6 in the Appendix consider contemporaneous commodity price changes and lagged economic growth as regressors and show that there is no statistical association of these variables to civil conflict.

[23] The conflict ending dummy is equal to one if there is an active conflict in year $t-1$, but no conflict in year t. If there is a conflict in both year t and $t-1$, the dummy variable is set to zero.

	(1) Onset[1] State	(2) Onset[1] Non-state	(3) Onset[1] One-sided	(4) Onset[25] All	(5) Onset[50] All	(6) Ending[1] All
Spatial Lag	0.683***	0.709***	0.863***	0.365*	0.477***	-0.910***
	(0.0960)	(0.233)	(0.141)	(0.198)	(0.129)	(0.190)
Economic growth, t	0.156	-0.00276	0.0283	0.0225	-0.122	0.963
(predicted)	(0.230)	(0.140)	(0.0659)	(0.129)	(0.114)	(1.230)
Commodity price growth, $t-1$	-0.00616	-0.00757	0.0146	-0.0142	-0.0199	-0.0361
(labor-intensive)	(0.0239)	(0.0200)	(0.0143)	(0.0202)	(0.0180)	(0.0976)
Commodity price growth, $t-1$	0.0629***	0.00908	0.00675	0.0193	0.0172	0.0945
(capital-intensive)	(0.0197)	(0.0144)	(0.0109)	(0.0128)	(0.0125)	(0.108)
Spatial Matrix	Country	Country	Country	Country	Country	Country
Kleibergen-Paap Wald F-test	42.12	51.48	43.19	43.41	43.06	2.895
SW F-test: Growth	83.89	93.11	86.14	87.46	85.04	5.842
SW F-test: Spatial lag	368.5	131.4	577.7	418.6	409.0	72.65
Observations	9816	10873	11178	10593	10827	1619

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: *IV Post-Lasso and alternative dependent variables.*

2.6.4 Extension: Spatial heterogeneity

The analysis in the previous section assumes that the effect of economic growth on conflict is the same across the African continent. However, as causes of conflicts are diverse and complex, the role of economic growth shocks is likely to vary substantially across the continent. Presumably, one reason for the missing link between economic growth and civil conflict is that the effect of economic growth varies across space, while the estimation method only identifies the average effect. For instance, some areas may be dominated by the opportunity cost mechanisms, whereas others by the rapacity mechanism.

In this section, I explore parameter heterogeneity across space using Geographically Weighted Regression (GWR) due to McMillen (1996) and Brunsdon, Fotheringham, and Charlton (1996). In order to approximate the local effect of economic shocks on the risk of civil conflict in area i, the reference model in Table 2.6, Model 1, which uses a country weights matrix, is estimated with the Gaussian weighting function

$$a_{ij} = \exp\left(-0.5(d_{ij}/b)^2\right) \quad \text{for } i, j = 1, \ldots, N, \tag{2.8}$$

where b is a bandwidth parameter and d_{ij} is the distance between the centroids of area i and area j in kilometres. Thus, areas farther away from unit i receive smaller weights and have less impact on the local coefficient estimate.[24]

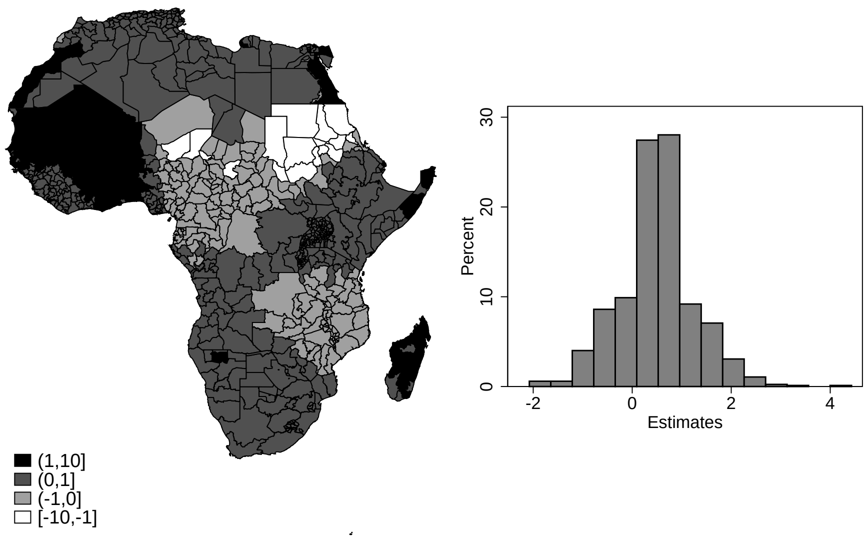
Figure 2.3 displays the coefficient estimates of the effect of economic growth on civil conflict for bandwidths of $b = 600$ and 800. Black and dark gray areas indicate a positive effect of economic growth on conflict, while white and light gray areas indicate a negative effect. Hence, the approach allows to determine which areas are dominated by the rapacity mechanism (black and dark gray) and which areas are dominated by the opportunity cost mechanism (white and light gray). At a bandwidth of $b = 600$, parameter heterogeneity is substantial. As the bandwidth increases, spatial parameter heterogeneity gradually declines. At a bandwidth of $b = 800$, Western Africa stands clearly out as a region that is governed by the rapacity mechanism. It is interesting to note that, for both bandwidth sizes, the distribution of estimates is centred slightly above zero as shown in the histograms, providing some support for the rapacity mechanism. However, the variation is large, with coefficient estimates varying between approximately -2 and 4.

2.7 Conclusion

Previous studies exploiting climate variables as instruments have suffered from at least two problems. First, it is *a priori* not evident how climate instruments should be specified in order to capture the effect on the economy, while at the same time accounting for non-linearities and heterogeneity across climate regions. Second, the link between climate and economic growth is only weak in many studies, inducing an estimation bias. The approach in this study uses the Lasso estimator to generate approximately optimal instruments from a large set of putative instruments and addresses both issues successfully as indicated by weak identification tests. The application of the Lasso estimator also demonstrates how modern variable selection techniques can enhance empirical research of conflicts. In particular, the data-driven selection and construction of instruments is transparent, leaves less room for misspecification and, thereby, could improve the robustness of estimation results (see discussion in Hegre and Sambanis, 2006).

[24] Note that the weight for observation i is $a_{ii} = 1$.

(a) *Bandwidth:* $b = 600$



(b) *Bandwidth:* $b = 800$

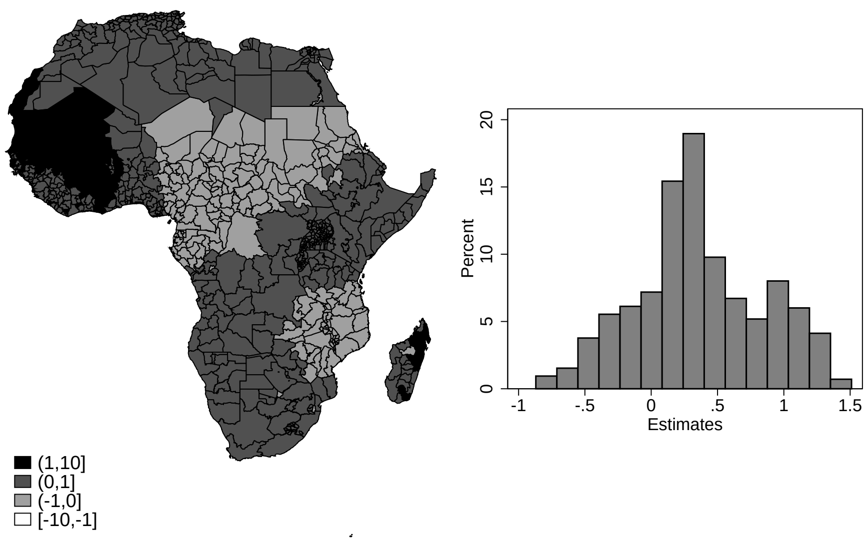


Figure 2.3: *Exploring spatial heterogeneity using Geographically Weighted Regression. The left-hand side figures show the parameter estimates using bandwidth parameters of* $b = 600$ *and* $b = 800$*, respectively. The right-hand side histograms show the distribution of coefficient estimates. The bandwidth parameter determines how quickly the least square weights decrease as distance increases.*

With respect to the data set, this article demonstrates that the use of nighttime light data from satellites provides the opportunity to examine political and economic processes at a disaggregate level. I show that civil conflicts are significantly reflected in the satellite data, which highlights that nighttime lights contain relevant information for empirical studies of conflicts and may be exploited in future studies.

For the considered sample of African sub-national areas in 1992-2010, estimation results yield no evidence that economic growth shocks have a significant causal effect on civil conflicts. One reason for the missing statistical association between economic growth and conflict onset may be that the estimation method only identifies the average effect and ignores parameter heterogeneity across space. Specifically, some areas may be predominantly governed by the rapacity mechanism, while others are predominantly governed by opportunity costs. This view is supported by results from GWR which reveals substantial spatial heterogeneity.

Furthermore, conflict risk seems to be increasing in prices of capital-intensive commodities, which suggests that higher resource rents provide an incentive for violent appropriation and that the state capacity mechanism is of less importance. Labour-intensive commodities, however, are not statistically associated with the onset of civil conflicts. Lastly, consistent with previous studies accounting for spatial effects (e.g. Buhaug and Gleditsch, 2008; Jensen and Gleditsch, 2009), spill-over effects are overall the most striking and robust empirical pattern in the data. This result stresses the need of controlling for spatial effects—even if the primary focus is not on understanding spatial diffusion processes.

Appendix of Chapter 2

B.1 Additional regression results

	(1) Onset[1]	(2) Onset[1]	(3) Onset[1]	(4) Onset[1]	(5) Onset[1]
Spatial Lag		0.547**	0.232	0.182	0.728
		(0.214)	(0.157)	(0.180)	(0.712)
Economic growth, t	-0.752	-0.143	-0.850	-0.415	-0.0472
(predicted)	(0.634)	(0.508)	(0.519)	(0.587)	(0.602)
Commodity price growth, t	-0.0203	-0.00960	-0.0172	-0.0111	-0.00555
(labor-intensive goods)	(0.0250)	(0.0230)	(0.0249)	(0.0252)	(0.0264)
Commodity price growth, t	-0.0361	-0.0115	-0.0172	-0.0323	-0.0250
(capital intensive)	(0.0247)	(0.0218)	(0.0257)	(0.0236)	(0.0243)
Spatial Matrix		Country	Neighbour	Ethnic	Distance
Kleibergen-Paap Wald F-test	5.674	3.824	4.492	4.658	3.565
SW F-test: Growth		4.477	5.271	5.576	4.087
SW F-test: Spatial lag		9.036	23.45	6.864	7.901
Hansen-Sargan (Df.)	3	6	6	6	6
Hansen-Sargan (p-val.)	0.296	0.271	0.446	0.561	0.593
Observations	10209	10209	10209	10209	10209

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.1: *GMM estimation with contemporaneous commodity prices.*

	(1) Onset[1]	(2) Onset[1]	(3) Onset[1]	(4) Onset[1]	(5) Onset[1]
Spatial Lag		0.565***	0.321***	0.332**	0.733
		(0.150)	(0.113)	(0.131)	(0.487)
Economic growth, $t-1$	-0.568	-0.255	-0.617*	-0.292	-0.246
(predicted)	(0.424)	(0.334)	(0.361)	(0.367)	(0.315)
Commodity price growth, $t-1$	-0.0483*	-0.0291	-0.0307	-0.0299	-0.0314
(labor-intensive goods)	(0.0254)	(0.0255)	(0.0269)	(0.0271)	(0.0266)
Commodity price growth, $t-1$	0.0391*	0.0317	0.0566***	0.0394*	0.0450**
(capital intensive)	(0.0216)	(0.0201)	(0.0212)	(0.0205)	(0.0213)
Spatial Matrix		Country	Neighbour	Ethnic	Distance
Kleibergen-Paap Wald F-test	4.740	3.787	3.403	4.757	4.550
SW F-test: Growth		4.067	3.663	4.845	5.061
SW F-test: Spatial lag		10.22	33.21	8.402	17.22
Hansen-Sargan (Df.)	7	14	14	14	14
Hansen-Sargan (p-val.)	0.600	0.696	0.670	0.531	0.802
Observations	9603	9603	9603	9603	9603

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: *GMM estimation with lagged economic growth.*

	(1) Onset[1]	(2) Onset[1]	(3) Onset[1]	(4) Onset[1]	(5) Onset[1]
Spatial Lag		0.663***	0.466***	0.527***	1.259***
		(0.104)	(0.0999)	(0.119)	(0.351)
Economic growth, t	-0.242	0.118	-0.132	-0.0830	0.107
(predicted)	(0.221)	(0.218)	(0.220)	(0.220)	(0.234)
Commodity price growth, t	-0.0136	-0.00536	0.00740	0.000420	0.00695
(labor-intensive)	(0.0235)	(0.0226)	(0.0242)	(0.0236)	(0.0234)
Commodity price growth, t	-0.0404*	-0.0110	-0.0138	-0.0222	-0.0174
(capital-intensive)	(0.0234)	(0.0211)	(0.0241)	(0.0226)	(0.0235)
Spatial Matrix		Country	Neighbour	Ethnic	Distance
Kleibergen-Paap Wald F-test	87.27	41.88	42.98	43.10	36.19
SW F-test: Growth		83.69	86.05	87.61	72.77
SW F-test: Spatial lag		307.3	1094.8	354.3	458.3
Observations	10209	10209	10209	10209	10209

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: *IV Post-Lasso with contemporaneous commodity prices.*

	(1) Onset[1] State	(2) Onset[1] Non-state	(3) Onset[1] One-sided	(4) Onset[25] All	(5) Onset[50] All	(6) Ending[1] All
Spatial Lag	0.606***	0.721***	0.869***	0.389***	0.481***	-1.054***
	(0.102)	(0.184)	(0.132)	(0.126)	(0.116)	(0.203)
Economic growth, t	0.145	-0.0165	0.0165	0.0709	-0.0869	0.465
(predicted)	(0.220)	(0.126)	(0.0680)	(0.123)	(0.105)	(1.085)
Commodity price growth, t	0.00750	-0.00573	-0.00510	-0.00556	0.00405	0.0178
(labor-intensive)	(0.0214)	(0.0191)	(0.0144)	(0.0186)	(0.0176)	(0.0985)
Commodity price growth, t	-0.0215	-0.000452	-0.00780	-0.0135	-0.0178	0.127
(capital-intensive)	(0.0205)	(0.0195)	(0.00959)	(0.0158)	(0.0130)	(0.115)
Spatial Matrix	Country	Country	Country	Country	Country	Country
Kleibergen-Paap Wald F-test	41.87	48.12	46.49	44.31	44.38	3.761
SW F-test: Growth	83.60	93.53	93.28	89.06	88.56	7.526
SW F-test: Spatial lag	364.8	285.7	609.4	733.2	616.5	50.98
Observations	10305	11419	11749	11130	11371	1706

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: *IV Post-Lasso with contemporaneous commodity prices (part 2).*

	(1) Onset[1]	(2) Onset[1]	(3) Onset[1]	(4) Onset[1]	(5) Onset[1]
Spatial Lag		0.698***	0.478***	0.528***	1.126***
		(0.0984)	(0.0940)	(0.116)	(0.380)
Economic growth, $t-1$	-0.361	-0.328	-0.307	-0.253	-0.293
(predicted)	(0.301)	(0.295)	(0.293)	(0.295)	(0.297)
Commodity price growth, $t-1$	-0.0484*	-0.0216	-0.0132	-0.0128	-0.0252
(labor-intensive)	(0.0252)	(0.0256)	(0.0272)	(0.0276)	(0.0265)
Commodity price growth, $t-1$	0.0379*	0.0435**	0.0602***	0.0507**	0.0589***
(capital-intensive)	(0.0210)	(0.0207)	(0.0213)	(0.0208)	(0.0209)
Spatial Matrix		Country	Neighbour	Ethnic	Distance
Kleibergen-Paap Wald F-test	72.63	36.22	36.15	36.27	35.93
SW F-test: Growth		72.74	72.71	71.43	71.81
SW F-test: Spatial lag		286.1	1063.8	333.1	612.7
Observations	9603	9603	9603	9603	9603

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: *IV Post-Lasso with lagged economic growth.*

	(1) Onset[1] State	(2) Onset[1] Non-state	(3) Onset[1] One-sided	(4) Onset[25] All	(5) Onset[50] All	(6) Ending[1] All
Spatial Lag	0.651***	0.712***	0.843***	0.362*	0.456***	-1.029***
	(0.0923)	(0.229)	(0.141)	(0.197)	(0.126)	(0.218)
Economic growth, $t-1$	-0.303	-0.0280	0.00469	0.117	0.112	-1.552
(predicted)	(0.294)	(0.124)	(0.103)	(0.136)	(0.128)	(1.832)
Commodity price growth, $t-1$	-0.0101	-0.00895	0.0194	-0.00491	-0.0102	-0.0707
(labor-intensive)	(0.0247)	(0.0207)	(0.0150)	(0.0214)	(0.0187)	(0.104)
Commodity price growth, $t-1$	0.0579***	0.00949	0.00550	0.0187	0.0170	0.0391
(capital-intensive)	(0.0210)	(0.0141)	(0.0112)	(0.0133)	(0.0125)	(0.132)
Spatial Matrix	Country	Country	Country	Country	Country	Country
Kleibergen-Paap Wald F-test	36.20	38.73	40.19	37.26	40.94	3.153
SW F-test: Growth	73.26	76.41	82.24	74.37	82.41	6.603
SW F-test: Spatial lag	309.2	124.7	668.0	427.1	534.6	83.85
Observations	9692	10766	11064	10481	10713	1627

See notes in Table 2.5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.6: *IV Post-Lasso with lagged economic growth (part 2).*

Chapter 3

Two-Step Lasso Estimation of the Spatial Weights Matrix[1]

Learn how to see. Realize that everything connects to everything else.

– Leonardo da Vinci

The spatial econometric application presented in Chapter 2 relies on specifying the spatial weights matrix using observable distance measures. Four different specifications were considered the previous chapter, but there exists an indefinite number of distinct specifications and theory provides only little guidance as to which spatial weights matrix is most suitable. As in conflict research application, researchers often select between standard specifications such as the binary contiguity matrix, inverse distance matrix or other matrices based on some observable notion of distance. The arbitrary choice of spatial weights has been a focus of criticism of spatial econometric methods, since estimation results highly depend on the researcher's specification of the spatial weights matrix (Arbia and Fingleton, 2008; Harris, Moffat, and Kravtsova, 2011; Corrado and Fingleton, 2012). Furthermore, a pre-defined weights matrix does not provide insights into the drivers of socio-economic interactions and general equilibrium effects in a network, but only allows for measuring the general strength of interactions, which is reflected in the size of the spatial autoregressive coefficient.

[1]An earlier version of this chapter has been published jointly with Arnab Bhattacharjee in *Econometrics*, 2015, vol. 3, pp. 128–155. See Ahrens and Bhattacharjee, 2015.

This study proposes an estimation method, based on the Lasso estimator, for an approximately sparse spatial weights matrix in a large T setting. The vast majority of spatial econometric research relies on the assumption that the spatial weights matrix, $\mathbf{W}$, which measures the strength of interactions between units, is known *a priori*.

The shortcomings of employing pre-specified spatial weights are well known. Pinkse, Slade, and Brett (2002) is one of the first attempts to conduct inferences in a setting where the spatial weights matrix is not known *a priori*. The authors propose a semi-parametric estimator which relies on observable distance measures. Bhattacharjee and Jensen-Butler (2013) consider estimation of the spatial weights matrix from the spatial autocovariance matrix in spatial panel models, and show that $\mathbf{W}$ is only partially identified. Intuitively, the main issue is that, in contrast to autocovariances, spatial weights reflect the direction and strength of causation between spatial units. Since there are twice as many spatial weights as there are autocovariances, further assumptions are required for identification. Bhattacharjee and Jensen-Butler (2013) propose an estimator that provides exact identification under the assumption that the spatial weights matrix is symmetric and n is fixed.[2] Estimation of the spatial weights matrix in a small n panel, under different structural assumptions on the autocovariances or using moment conditions is discussed in Bhattacharjee and Holly (2011) and Bhattacharjee and Holly (2013).

The aforementioned literature focuses on a low-dimensional setting where typically $n \ll T$. In contrast, Bailey, Holly, and Pesaran (2016) consider sparsity of the spatial weights matrix as an identification assumption in a spatial error panel model, where n can be large. They apply a multiple testing procedure to the matrix of spatial autocorrelation coefficients in order to identify the non-zero interactions, and place weights of $+1$, -1 or zero, depending on whether the autocorrelations are significantly positive, significantly negative or insignificant, respectively. There are also a few previous studies which apply Lasso-type estimators to high-dimensional spatial panel models and assume sparsity. Manresa (2013) considers a non-autoregressive

[2]See Beenstock and Felsenstein (2012) for a similar approach.

panel model with spatially lagged exogenous regressors. Hence, the model does not suffer from simultaneity and the Lasso estimator can be used for dimensionality reduction. Souza (2012) and Lam and Souza (2013) consider a spatial autoregressive model with additional spatial lags on exogenous regressors. Souza (2012) discusses several exclusion restrictions that allow for identification, but require prior knowledge about the network structure. Lam and Souza (2013) propose to employ the Adaptive Lasso estimator to estimate the spatial weights matrix under the assumption that the error variance decays to zero as T increases, which may be a strong assumption in some applications. By contrast, the method proposed here does not require prior knowledge about the network structure and does not rely on variance decay, but instead exploits exogenous regressors as instruments.

This study explores the estimation of the spatial weights matrix in a panel data setting where T, the number of time periods, is large. The spatial autoregressive or spatial lag model is given by

$$y_{it} = \sum_{j=1}^{n} w_{ij} y_{jt} + \mathbf{x}_{it}' \boldsymbol{\beta}_i + e_{it}, \quad t = 1, \dots, T, \;\; i = 1, \dots, n, \tag{3.1}$$

where y_{it} is the response variable, $\mathbf{x}_{it} = (x_{1,it}, x_{2,it}, \dots, x_{K,it})'$ is the vector of exogenous regressors and $\boldsymbol{\beta}_i$ is the $K \times 1$ parameter vector with $K \geq 1$. The error term is assumed to be independently distributed, but allowed to be heteroskedastic and non-Gaussian. w_{ij} is the (i, j)th element of the $n \times n$ spatial weights matrix, denoted by $\mathbf{W}$, and measures the strength of spill-over effects from unit j to unit i. The spatial weights matrix has zeros on the diagonal, i.e., $w_{ii} = 0$ for all i.[3] The first term on the right-hand side is often referred to as the spatial lag, analogous to a temporal lag in time-series models. The spatial autoregressive panel model is a natural extension to cross-sectional spatial autoregressive models introduced by Cliff and Ord (1973) and Anselin (1988). Spatio-temporal panel models, such as the spatial autoregressive model in (3.1), have recently attracted much attention; see

[3] We implicitly set the spatial autoregressive parameter, which is commonly employed in spatial models, equal to one, since w_{ij} and the spatial autoregressive parameter are not separately identified (Bhattacharjee and Jensen-Butler, 2013).

e.g. Kapoor, Kelejian, and Prucha (2007), Lee and Yu (2010c), Lee and Yu (2010b), Lee and Yu (2014) and Mutl and Pfaffermayr (2011).[4]

Estimation of the above model poses two major challenges when **W** is treated as unknown. First, the model suffers from reverse causality as the response variable appears both on the left and right-hand side of the equation. It is well known that Ordinary Least Squares (OLS) is inconsistent in the presence of endogeneity. Second, the model is not identified unless the number of parameters, $p := n(n-1) + Kn$, is smaller than the number of observations, nT, or further assumptions are made. The identification assumption considered here is sparsity of the weights matrix which requires that each unit is affected by only a limited number of other units. Specifically, the number of units affecting a specific unit i is assumed to be much smaller than T, but I explicitly allow for $p \gg nT$.

The proposed estimation method is a two-step procedure based on the Lasso estimator introduced by Tibshirani (1996). The Lasso is a regularization technique which can, under the sparsity assumption, deal with high-dimensional settings where the number of exogenous regressors is large relative to the number of observations. The ℓ_1-penalization employed by the Lasso sets some of the coefficient estimates to exactly zero, making the Lasso estimator attractive for model selection. The ℓ_1-penalization behaves similarly to the ℓ_0-penalty, as used in the Akaike information criterion and Bayesian information criterion (Akaike, 1974; Schwarz, 1978), but is computationally more attractive due to its convex form. The Lasso is a popular and well-established technique, but its theoretical properties have only recently been better understood. Recent theoretical contributions include Bickel, Ritov, and Tsybakov (2009), Bühlmann and Van de Geer (2011), Zhao and Yu (2006), Belloni et al. (2012) and Wainwright (2009).

Conceptually, identification of a spatial weights matrix requires suitably dealing with the endogeneity inherent in model (3.1). Lam and Souza (2013) address this issue by assuming that the error variance asymptotically decays to zero. By

[4] See Elhorst (2014) for an overview.

contrast, the proposed method addresses endogeneity using instruments. The estimation methodology proceeds in two steps. In the first step, relevant instruments are identified by the Lasso and predictions for $y_{1t}, \ldots, y_{nt}$ are obtained. In the second step, the regression model in (3.1) is estimated, but the spatial lag on the right-hand side is replaced with predictions from the first step. That is, the second-step Lasso selects the neighbours affecting y_{it}. The procedure is conceptually based on Two-stage Least Squares (2SLS), but employs the Lasso for selecting relevant instruments in the first step and for selecting relevant spatial lags in the second step. Figure 3.1 visualizes the spatial autoregressive model in (3.1) for $n = 2$ and motivates the choice of instruments exploited to identify the spatial weights. In the regression equation with y_{1t} as the dependent variable, $\mathbf{x}_{2t}$ can be exploited as instrumental variables for y_{2t} and vice versa.

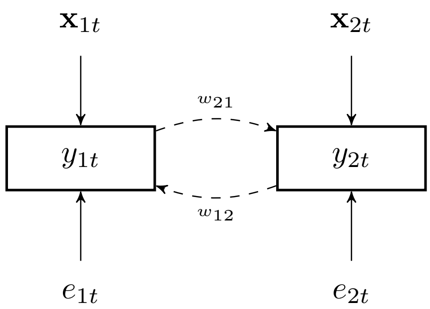


Figure 3.1: *The spatial autoregressive model for $n = 2$. Dashed arrows (‘$\dashrightarrow$’) indicate spatial effects, while non-spatial effects are denoted by ‘$\longrightarrow$’.*

I also consider the Post-Lasso OLS estimator due to Belloni and Chernozhukov (2013), which applies OLS to the model selected by the Lasso and aims at reducing the Lasso shrinkage bias. Although the estimation methodology relies on large T asymptotics, Monte Carlo results suggest that the two-step Lasso estimator is able to recover the spatial network structure if T is reasonably small.

Finally, this study is also related to the emerging literature on high-dimensional methods which allow the number of endogenous regressors to be larger than the sample size. The Self-Tuning Instrumental Variable (STIV) due to Gautier and Tsybakov (2014) is a generalization of the Dantzig estimator (Candes and Tao, 2007) and allows for many endogenous regressors. The Focused Generalised Methods of

Moments (FGMM) developed in Fan and Liao (2014) extends shrinkage Generalised Methods of Moments (GMM) estimators as in Caner (2009) to high-dimensional settings. The two-step Lasso estimator considered in this study is conceptually similar to Lin, Feng, and Li (2014) who apply two-step penalized least squares to genetic data. We improve upon Lin, Feng, and Li (2014) in that our approach allows for approximate sparsity, non-Gaussian errors and uses the sharper penalty level proposed by Belloni et al. (2012). However, the main contribution in this chapter is to point out that a simple two-step Lasso estimation method can be employed to estimate the spatial weights matrix. The approach does not require any prior knowledge about the network structure, except for the sparsity assumption and a set of exogenous regressors.

This chapter is organized as follows. In Section 3.1, I consider a general setting where the number of endogenous regressors and the number of instruments is allowed to be larger than the number of observations. Thus, I show that the two-step estimator may be of more general interest for applications with endogeneity in high-dimensions. Section 3.2 applies the proposed two-step estimator to estimate the spatial autoregressive model in (3.1). In Section 3.3, I present Monte Carlo results to demonstrate the performance of the two-step Lasso for estimating the spatial weights matrix. Finally, Section 3.4 concludes the chapter.

3.1 Two-step Lasso Estimator

In this section, I develop a two-step estimation procedure that allows the number of possibly endogenous regressors as well as the number of instruments to be larger than the sample size. The identifying assumption is approximate sparsity. Section 3.2 presents the spatial autoregressive panel model as an application to this setting. The two-step estimator may be of interest in, for example, cross-country growth regressions where the number of regressors is large relative to the number of countries and endogeneity is a potential issue. Furthermore, endogeneity in high dimensions may arise when the aim is to find a sparse linear approximation to a complex nonparametric data-generating process; see, e.g., earning regressions in Belloni and Chernozhukov (2011).

The structural equation and first-step equations are given by

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}^\star + e_t, \tag{3.2}$$

$$x_{tj} = \mathbf{z}_t'\boldsymbol{\pi}_j^\star + u_{tj}, \qquad j = 1, \ldots, p. \tag{3.3}$$

y_t is the outcome variable and $\mathbf{x}_t$ is a p-dimensional vector of regressors. For notational consistency with Section 3.2, I use $t = 1, \ldots, T$ to denote distinct units or repeated observations over time. Without loss of generality, I assume that the first $\bar{p}$ regressors are endogenous, i.e., $\mathrm{E}[e_t|x_{tj}] \neq 0$ for $j = 1, \ldots, \bar{p}$ with $\bar{p} \in \{1, \ldots, p\}$. The remaining $p - \bar{p}$ regressors are exogenous. Hence, I allow the set of exogenous regressors to be empty. I assume the existence of $L \geq p$ instruments, $\mathbf{z}_t$, which satisfy the exclusion restriction $\mathrm{E}[e_t|\mathbf{z}_t] = \mathrm{E}[u_{tj}|\mathbf{z}_t] = 0$ for $j = 1, \ldots, p$. If a regressors x_{tj} is exogenous, it serves as an instrument for itself. Hence, $z_{tj} = x_{tj}$ for $j > \bar{p}$. The error terms e_t and u_{tj} are independently distributed, but possibly heteroskedastic and non-Gaussian. The interest lies in obtaining a sparse approximation of $\boldsymbol{\beta}^\star$. While the model in (3.2)–(3.3) assumes that the conditional expectation functions are linear, the framework may be easily generalized to a non-parametric data-generating process as in Bickel, Ritov, and Tsybakov (2009).

3.1.1 First-step estimation

The aim of the first step is to estimate the conditional expectation function $x_{tj}^\star := \mathrm{E}[x_{tj}|\mathbf{z}_t] = \mathbf{z}_t'\boldsymbol{\pi}_j^\star$ for $j = 1, \ldots, \bar{p}$ where $x_{tj}^\star$ represents the optimal instrument. Note that $x_{tj}^\star = x_{tj}$ if x_{tj} is exogenous, which corresponds to $j = \bar{p} + 1, \ldots, p$.

If $L > T$, OLS estimation of the first-step equations in (3.3) is not feasible as the Gram matrix $T^{-1}\mathbf{Z}'\mathbf{Z}$ with $\mathbf{Z} = ((z_{tj}))$ is singular. The Lasso can achieve consistency in a high-dimensional setting where $L > T$ under the assumption of sparsity and further regularity conditions stated below. Exact sparsity requires that the number of nonzero elements in $\boldsymbol{\pi}_j^\star$, i.e., $\|\boldsymbol{\pi}_j^\star\|_0$, is small relative to the sample size. This assumption seems too restrictive as $\boldsymbol{\pi}_j^\star$ may have elements that are, although negligible, not exactly zero. Instead, I assume the existence of a sparse parameter vector $\boldsymbol{\pi}_j^0$ that approximates the true parameter vector $\boldsymbol{\pi}_j^\star$ sufficiently well. Specifically, as in Belloni et al. (2012), I assume that for each endogenous regressor j the number of instruments necessary for approximating the conditional expectation function is smaller than the sample size and the associated approximation error $a_{tj}(\mathbf{z}_t) = x_{tj}^\star - \mathbf{z}_t'\boldsymbol{\pi}_j^0$ converges as specified below.[5]

Assumption 3.1. *Consider the model in* (3.3)*. There exists a parameter vector* $\boldsymbol{\pi}_j^0$ *for all* $j = 1, \ldots, \bar{p}$ *such that*

$$\mathrm{E}[x_{tj}|\mathbf{z}_t] = \mathbf{z}_t'\boldsymbol{\pi}_j^0 + a_{tj}(\mathbf{z}_t), \;\; s_1 := \max_{1\le j\le\bar{p}} \|\boldsymbol{\pi}_j^0\|_0 \ll T, \;\; A_{s_1} := \max_{1\le j\le\bar{p}} \sqrt{\frac{1}{T}\sum_{t=1}^T a_{tj}^2} \lesssim_{\mathrm{P}} \sqrt{\frac{s_1}{T}}. \tag{3.4}$$

Remark 3.1. The sparse target parameter $\boldsymbol{\pi}_j^0$ can be motivated as the solution to the infeasible oracle program that penalises the number of non-zero parameters (Belloni and Chernozhukov, 2013). Under homoskedasticity, the oracle objective function can be written as

$$\min_{\boldsymbol{\pi}_j} \frac{1}{T}\|\mathbf{X}_{\bullet j} - \mathbf{Z}\boldsymbol{\pi}_j\|_2^2 + \frac{\sigma^2}{T}\|\boldsymbol{\pi}_j\|_0, \tag{3.5}$$

where $\mathbf{X}_{\bullet j}$ is the jth column of the matrix $\mathbf{X} = ((x_{tj}))$. The second term represents the noise level and $\sqrt{s_1/T}$ is the convergence rate of the oracle which knows the true model.

The first-step Lasso estimator for endogenous regressor j is defined as

$$\hat{\boldsymbol{\pi}}_j = \arg\min_{\boldsymbol{\pi}_j} \frac{1}{T}\|\mathbf{X}_{\bullet j} - \mathbf{Z}\boldsymbol{\pi}_j\|_2^2 + \frac{\lambda_1}{T}\|\boldsymbol{\Upsilon}_{1j}\boldsymbol{\pi}_j\|_1. \tag{3.6}$$

The first term is the residual sum of squares and the second term imposes a penalty on the absolute size of the parameters which is increasing in the penalty level λ_1.

[5]The subscripts '1' and '2' indicate, where appropriate, that the corresponding terms refer to the first and second step, respectively.

The Lasso nests OLS with $\lambda_1 = 0$, while $\lambda_1 = \infty$ yields a null model. $\boldsymbol{\Upsilon}_{1j}$ is a diagonal matrix of penalty loadings which account for heteroskedasticity and may be set to the identity matrix under homoskedasticity (Belloni et al., 2012). The Lasso predictions $\hat{\mathbf{X}}_{\bullet j} := \mathbf{Z}\hat{\boldsymbol{\pi}}_j$ replace $\mathbf{X}_{\bullet j}$ in the second step to address endogeneity. For the exogenous regressors, I set $\hat{\mathbf{X}}_{\bullet j} = \mathbf{X}_{\bullet j}$.

The penalty level λ_1 may be selected by cross-validation in order to minimize the prediction error as originally suggested by Tibshirani (1996). Since the primary purpose of our study is not prediction, but recovery of the spatial network structure, I follow an alternative approach that originates from Bickel, Ritov, and Tsybakov (2009). The penalty level is chosen as the smallest value that, with a high probability, overrules the random part of the data-generating process, which is represented by the score vector $\mathbf{S}_{1j} = -\frac{2}{T}\boldsymbol{\Upsilon}_{1j}^{-1}\mathbf{Z}'\mathbf{u}_j$, i.e.,

$$\frac{\lambda_1}{T} \geq c \max_{1 \leq j \leq \bar{p}} \|\mathbf{S}_{1j}\|_\infty \quad \text{with } c > 1. \tag{3.7}$$

The event in (3.7) plays a crucial role in the derivation of non-asymptotic bounds and convergence rates. Belloni et al. (2012) show with the use of moderate deviation theory in Jing, Shao, and Wang (2003) that setting

$$\lambda_1 = 2c\sqrt{T}\Phi^{-1}(1 - \alpha/(2L\bar{p})) \quad \text{with} \quad \log(1/\alpha) \lesssim \log(\max(L\bar{p}, T)) \tag{3.8}$$

guarantees

$$P\left(c \max_{1 \leq j \leq \bar{p}} \|\mathbf{S}_{1j}\|_\infty > \lambda_1/T\right) = o_{\mathrm{P}}(1) \quad \text{as} \quad T \to \infty.$$

under possibly non-Gaussian and heteroskedastic errors. Note that the term $\bar{p}$ in (3.8) accounts for the number of Lasso regressions in the first step and L is the number of instruments. c is a constant greater than, but close to 1. In applied work, Belloni, Chernozhukov, and Hansen (2014b) suggest setting $c = 1.1$ and $\alpha = \min(1/T, 0.05)$.

The optimal penalty loadings for the first-step equation are given by

$$\boldsymbol{\Upsilon}_{1j}^0 = \text{diag}(\gamma_{1j,1}, \ldots, \gamma_{1j,l}, \ldots, \gamma_{1j,L}), \quad \gamma_{1j,l} = \sqrt{\frac{1}{T}\sum_t z_{tl}^2 u_{tj}^2}, \tag{3.9}$$

but are infeasible as u_{tj} is unobserved. Belloni et al. (2012) propose an algorithm for constructing asymptotically valid penalty loadings, $\hat{\boldsymbol{\Upsilon}}_{1j}$, that are in the probability limit as least as large as the optimal penalty loadings (see Appendix C.2).

Definition 3.1. *The matrix of estimated penalty loadings $\hat{\boldsymbol{\Upsilon}}$ are said to be asymptotically valid for the optimal infeasible penalty loadings $\boldsymbol{\Upsilon}^0$ if $l\boldsymbol{\Upsilon}^0 \leq \hat{\boldsymbol{\Upsilon}} \leq u\boldsymbol{\Upsilon}^0$ where $0 < l \leq 1 \leq u$ and $l \to_{\text{P}} 1$ and $u \to_{\text{P}} u'$ with $u' \geq 1$ (Belloni et al., 2012).*

The properties of the Lasso estimator depend crucially on the Gram matrix $T^{-1}\mathbf{Z}'\mathbf{Z}$. As stated above, OLS is not feasible if $L > T$ as the Gram matrix is singular, which implies that the minimum eigenvalue is zero,

$$\min_{\boldsymbol{\delta}\neq\mathbf{0}} \frac{\|\mathbf{Z}\boldsymbol{\delta}\|_2}{\sqrt{T}\,\|\boldsymbol{\delta}\|_2} = 0. \tag{3.10}$$

Bickel, Ritov, and Tsybakov (2009) introduce the restricted eigenvalue,

$$\min_{\boldsymbol{\delta}\neq\mathbf{0}} \frac{\|\mathbf{Z}\boldsymbol{\delta}\|_2}{\sqrt{T}\left\|\boldsymbol{\delta}_{\Omega_{1j}}\right\|_2} \quad \text{subject to} \quad \|\boldsymbol{\delta}_{\overline{\Omega}_{1j}}\|_1 \leq C\|\boldsymbol{\delta}_{\Omega_{1j}}\|_1, \tag{3.11}$$

which is defined as the minimum over the restricted set $\|\boldsymbol{\delta}_{\overline{\Omega}_{1j}}\|_1 \leq C\|\boldsymbol{\delta}_{\Omega_{1j}}\|_1$, where $\Omega_{1j} = \text{supp}(\boldsymbol{\pi}_j^0)$ and C is a positive constant. The condition $\|\boldsymbol{\delta}_{\overline{\Omega}_{1j}}\|_1 \leq C\|\boldsymbol{\delta}_{\Omega_{1j}}\|_1$ holds with high probability and, when it does not hold, it is not required to bound the prediction error norm (see Appendix C.1).

Definition 3.2. *Let C and $\bar{\kappa}$ be positive constants and Ω denote the active set. The restricted eigenvalue condition holds for $\mathbf{M}$, if as $T \to \infty$*

$$\kappa_C(\mathbf{M}) := \min_{\|\boldsymbol{\delta}_{\overline{\Omega}}\|_1 \leq C\|\boldsymbol{\delta}_\Omega\|_1,\, \boldsymbol{\delta}\neq\mathbf{0}} \frac{\sqrt{s}}{\sqrt{T}} \frac{\|\mathbf{M}\boldsymbol{\delta}\|_2}{\|\boldsymbol{\delta}_\Omega\|_1} \geq \bar{\kappa} > 0, \quad s := \|\boldsymbol{\delta}\|_0. \tag{3.12}$$

In the above definition of the restricted eigenvalue the ℓ_2-norm in the denominator is replaced with the ℓ_1-norm using the Cauchy-Schwarz inequality, which allows to relate the ℓ_1-parameter norm to the ℓ_2-prediction norm. The restricted eigenvalue is closely related to the compatibility constant (Van de Geer and Bühlmann, 2009). Bühlmann and Van de Geer (2011) provide an extensive overview of related conditions and their relationship. The restricted eigenvalue conditions hold under general conditions; see, among others, Raskutti, Wainwright, and Yu (2010), Belloni and Chernozhukov (2013) and Bickel, Ritov, and Tsybakov (2009). One sufficient condition for the restricted eigenvalue is the restricted sparse eigenvalue condition which requires that any appropriate sub-matrix of the Gram matrix has positive and finite eigenvalues (Bickel, Ritov, and Tsybakov, 2009).

To accommodate heteroskedasticity, I also define the weighted restricted eigenvalue condition (Belloni et al., 2012).

Definition 3.3. *Let C and $\bar{\kappa}$ be positive constants and Ω denote the active set. The weighted restricted eigenvalue condition holds for $\mathbf{M}$, if as $T \to \infty$*

$$\kappa_C^{\omega}(\mathbf{M}) := \min_{\|\boldsymbol{\Upsilon}^0\boldsymbol{\delta}_{\overline{\Omega}}\|_1 \leq C\|\boldsymbol{\Upsilon}^0\boldsymbol{\delta}_{\Omega}\|_1,\, \boldsymbol{\delta}\neq \mathbf{0}} \frac{\sqrt{s}}{\sqrt{T}} \frac{\|\mathbf{M}\boldsymbol{\delta}\|_2}{\|\boldsymbol{\Upsilon}^0\boldsymbol{\delta}_{\Omega}\|_1} \geq \bar{\kappa} > 0, \quad s := \|\boldsymbol{\delta}\|_0, \tag{3.13}$$

where $\boldsymbol{\Upsilon}^0$ are the optimal penalty loadings.

If the restricted eigenvalue condition holds, the weighted restricted eigenvalue condition is also satisfied as long as the optimal penalty loadings are bounded away from zero and bounded from above, which I maintain in the following. With respect to the first-step equations in (3.3), I explicitly state the restricted eigenvalue condition as follows:

Assumption 3.2. *The restricted eigenvalue and the weighted restricted eigenvalue condition hold for $\mathbf{Z}$ in equation* (3.3).

Under Assumption 3.1-3.2, using the penalty level as in (3.8) and assuming the penalty loadings $\hat{\boldsymbol{\Upsilon}}_{1j}$ are asymptotically valid, then by Theorem 1 in Belloni et al. (2012), the ℓ_2-prediction error norm of the Lasso estimator has the following rate of convergence

$$\max_{1\le j\le \bar{p}} \frac{1}{\sqrt{T}} \left\| \mathbf{Z}\hat{\boldsymbol{\pi}}_j - \mathbf{Z}\boldsymbol{\pi}_j^{\star} \right\|_2 \lesssim_{\mathrm{P}} \sqrt{\frac{s_1 \log(\max(L\bar{p}, T))}{T}}. \tag{3.14}$$

The proof of Theorem 1 in Belloni et al. (2012) is not reproduced here. However, the main result of this chapter is a generalization in that it accounts for the prediction error that arises from the first-step Lasso estimation. Note that the convergence rate in (3.14) is slower than the oracle rate of $\sqrt{s_1/T}$ by a factor of $\sqrt{\log(\max(L\bar{p}, T))}$, which can be interpreted as the cost of not knowing the active set of $\boldsymbol{\pi}_j^0$.

3.1.2 Second-step estimation

Since the second step is not feasible by OLS if $p > T$, the assumption of approximate sparsity and the restricted eigenvalue condition are required, as in the first step, to guarantee identification.

Assumption 3.3. *Consider the model in* (3.2)*. There exists a parameter vector* $\boldsymbol{\beta}^0$ *such that*

$$\mathrm{E}[y_t|\mathbf{z}_t] = \mathbf{x}_t^{\star\prime}\boldsymbol{\beta}^0 + r_t(\mathbf{z}_t), \qquad s_2 := \|\boldsymbol{\beta}^0\|_0 \ll T, \qquad \|\boldsymbol{\beta}^0\|_2 \lesssim s_2, \tag{3.15}$$

$$R_{s_2} := \sqrt{\frac{1}{T}\sum_{t=1}^{T} r_t(\mathbf{z}_t)^2} \lesssim_{\mathrm{P}} \sqrt{\frac{s_2}{T}}. \tag{3.16}$$

Assumption 3.4. *The restricted eigenvalue and weighted restricted condition holds for* $\hat{\mathbf{X}}$ *in equation* (3.2)*.*

Assumption 3.3 is similar to Assumption 3.1, but assumes $\|\boldsymbol{\beta}^0\|_2 \lesssim s_2$, which allows to simplify the expression for the convergence rates. Assumption 3.4 could also be written in terms of the optimal instrument matrix $\mathbf{X}^{\star}$. Specifically, Assumption 3.4 holds if the restricted eigenvalue holds for $\mathbf{X}^{\star}$ and $\|\hat{\mathbf{X}}'\hat{\mathbf{X}} - \mathbf{X}^{\star\prime}\mathbf{X}^{\star}\|_\infty$ is small as discussed in Appendix C.4.

For identification of $\boldsymbol{\beta}^0$ I also require, as standard in the IV/GMM literature, that the matrix $\boldsymbol{\Pi}^0 = (\boldsymbol{\pi}_1^0, \ldots, \boldsymbol{\pi}_p^0)$ is full column rank.

Assumption 3.5. *The matrix* $\mathbf{\Pi}^0$ *is full column rank, i.e,* $\mathrm{rank}(\mathbf{\Pi}^0) = p$.

The second-step Lasso estimator uses the predictions $\hat{\mathbf{X}}$ as regressors and is defined as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{T} \left\| \mathbf{y} - \hat{\mathbf{X}}\boldsymbol{\beta} \right\|_2^2 + \frac{\lambda_2}{T} \left\| \mathbf{\Upsilon}_2 \boldsymbol{\beta} \right\|_1, \tag{3.17}$$

where the penalty level is set to

$$\lambda_2 = 2c\sqrt{T}\Phi^{-1}(1-\alpha/(2p)) \quad \text{with} \quad \log(1/\alpha) \lesssim \log(\max(L\bar{p}, T)) \tag{3.18}$$

and the penalty loadings are estimated using the algorithm in Appendix C.2.

The crucial difference to the first-step Lasso estimation is that $\mathbf{X}^\star$ is unobservable and, thus, replaced with $\hat{\mathbf{X}}$, which is an estimate that in general deviates from the optimal instrument $\mathbf{X}^\star$. For the two-step Lasso estimator, I consider the prediction bound $1/\sqrt{T}\|\hat{\mathbf{X}}\hat{\boldsymbol{\beta}} - \mathbf{X}^\star\boldsymbol{\beta}^\star\|_2$ where predictions obtained using the unknown optimal instrument and the unknown true parameter vector $\boldsymbol{\beta}^\star$ serve as a reference point. Note that, by triangle inequality

$$\frac{1}{\sqrt{T}} \left\| \hat{\mathbf{X}}\hat{\boldsymbol{\beta}} - \mathbf{X}^\star\boldsymbol{\beta}^\star \right\|_2 = \frac{1}{\sqrt{T}} \left\| (\hat{\mathbf{X}}\hat{\boldsymbol{\beta}} - \hat{\mathbf{X}}\boldsymbol{\beta}^0) + (\hat{\mathbf{X}}\boldsymbol{\beta}^0 - \mathbf{X}^\star\boldsymbol{\beta}^0) + (\mathbf{X}^\star\boldsymbol{\beta}^0 - \mathbf{X}^\star\boldsymbol{\beta}^\star) \right\|_2 \tag{3.19}$$

$$\leq \frac{1}{\sqrt{T}} \left\| \hat{\mathbf{X}}\hat{\boldsymbol{\beta}} - \hat{\mathbf{X}}\boldsymbol{\beta}^0 \right\|_2 + \frac{1}{\sqrt{T}} \left\| \hat{\mathbf{V}}\boldsymbol{\beta}^0 \right\|_2 + R_{s_2}, \tag{3.20}$$

where I define $\hat{\mathbf{V}} = \hat{\mathbf{X}} - \mathbf{X}^\star$ which has the typical element $\hat{v}_{jt} = \mathbf{z}_t'\hat{\boldsymbol{\pi}}_j - \mathbf{z}_t'\boldsymbol{\pi}_j^\star$. The bound for the third term is stated in Assumption 3.3. The convergence rate for the second term follows from prediction norm rate of the first-step Lasso in (3.14). The bound for the first term is derived in Appendix C.1. Combining the three bounds yields the following result.

Theorem 3.1. *Consider the model in* (3.2)–(3.3). *Suppose Assumptions 3.1–3.5 hold. Suppose asymptotically valid penalty loadings are used and the penalty levels*

λ_1 and λ_2 are set as in (3.8) *and* (3.18). *Then,*

$$\frac{1}{\sqrt{T}} \left\| \hat{\mathbf{X}} \hat{\boldsymbol{\beta}} - \mathbf{X}^{\star} \boldsymbol{\beta}^{\star} \right\|_2 \lesssim_{\mathrm{P}} s_2^2 \sqrt{\frac{s_1 \log(\max(L\bar{p}, T))}{T}}. \tag{3.21}$$

Furthermore, if s_1, s_2, L and $\bar{p}$ do not depend on T, then

$$\frac{1}{\sqrt{T}} \left\| \hat{\mathbf{X}} \hat{\boldsymbol{\beta}} - \mathbf{X}^{\star} \boldsymbol{\beta}^{\star} \right\|_2 \lesssim_{\mathrm{P}} \sqrt{\frac{\log(T)}{T}}, \tag{3.22}$$

$$\left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{0} \right\|_1 \lesssim_{\mathrm{P}} \sqrt{\frac{\log(T)}{T}}. \tag{3.23}$$

The proof is provided in Appendix C.1. As expected, the convergence rates of the ℓ_2-prediction norm depend on the degree of sparsity in the first-step and second-step equation. The second part of the theorem is relevant for the spatial panel model in the next section where the sparsity parameters (i.e., s_1 and s_2) and the dimension of the problem (L and $\bar{p}$) depend on the number of units (n), but may not depend on the time dimension (T).

3.2 The spatial autoregressive model

This section applies the proposed two-step Lasso procedure to the spatial lag model in (3.1). In Section 3.2.2, I discuss two extensions to the two-step Lasso estimator; namely, the Post-Lasso and thresholded Post-Lasso.

3.2.1 Two-step Lasso

The structural and reduced form equations can be written as

$$\mathbf{y}_i = \sum_{j=1}^{n} w_{ij}^{\star} \mathbf{y}_j + \mathbf{X}_i \boldsymbol{\beta}_i^{\star} + \mathbf{e}_i, \tag{3.24}$$

$$\mathbf{y}_j = \sum_{s=1}^{n} \mathbf{X}_s \boldsymbol{\pi}_{j,s}^{\star} + \mathbf{u}_j, \qquad i, j = 1, \ldots, n, \tag{3.25}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $w_{jj}^{\star} = 0$, and $\mathbf{X}_i = (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}')'$ is the $T \times K$ matrix of exogenous regressors. We assume that e_{it} is independently distributed across t, i.e., $E[e_{it}e_{is}] = 0$ for $t \neq s$. The '$\star$'-superscripts indicate that the parameters are interpreted as the true values, while the target parameters are marked with '0'-superscripts.

It is evident that the spatial model in (3.24)–(3.25) is an application of the more general model in (3.2)–(3.3). Specifically, the right-hand side regressors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ correspond to endogenous regressors and $\mathbf{X}_i$ corresponds to the exogenous regressors in (3.2). Furthermore, the set of exogenous instruments is given by $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

The choice of instruments is closely related to Kelejian and Prucha (1998). To identify the spatial autoregressive parameter, the authors suggest the use of first and higher order spatial lags of exogenous regressors as instruments for the endogenous spatial lag. As discussed in the Introduction, I use the columns of $\mathbf{X}_j$ as instruments in order to identify $w^\star_{ij}$, which represents the causal impact of $\mathbf{y}_j$ on $\mathbf{y}_i$. Therefore, for identification, I require contemporaneous exogeneity across space:

Assumption 3.6. $\mathrm{E}[e_{it}|\mathbf{x}_{jt}] = 0$ *for all* $i, j = 1, \ldots, n$ *and* $t = 1, \ldots, T$.

In many applications, estimation of (3.24)–(3.25) by 2SLS is not feasible as there appear $n - 1 + K$ and nK regressors on the right-hand side, respectively, which are both potentially larger than T. In order to exploit the Lasso estimator, I require sparseness as in Section 3.1.

Assumption 3.7. *(a) Consider the model in* (3.24). *There exists a parameter vector* $\mathbf{w}^0_i = (w^0_{i1}, \ldots, w^0_{in})$ *for all* $i = 1, \ldots, n$ *with* $w^0_{ii} = 0$ *such that*

$$\mathrm{E}[y_{it}|\mathbf{x}_{1t}, \ldots, \mathbf{x}_{nt}] = \sum_{j=1}^{n} w^0_{ij} y^\star_{jt} + \mathbf{x}'_{it}\boldsymbol{\beta}^0_i + a_{it}, \tag{3.26}$$

$$s_2 + K \ll T, \qquad A_{s_2} := \max_{1 \le i \le n} \sqrt{\frac{1}{T}\sum_{t=1}^{T} a^2_{it}} \lesssim_{\mathrm{P}} \sqrt{\frac{s_2}{T}}, \tag{3.27}$$

where $y^\star_{jt} = \mathrm{E}[y_{jt}|\mathbf{x}_{1t}, \ldots, \mathbf{x}_{nt}]$, $s_2 := \max_{1 \le i \le n} \|\mathbf{w}^0_i\|_0$ *and* $K = \|\boldsymbol{\beta}^0_i\|_0$.

(b) Consider the model in (3.25). *There exists a parameter vector* $\boldsymbol{\pi}^0_{i,j}$ *for all*

$i = 1, \ldots, n$ such that

$$\mathrm{E}[y_{it}|\mathbf{x}_{1t}, \ldots, \mathbf{x}_{nt}] = \sum_{j\neq i} \mathbf{x}'_{jt}\boldsymbol{\pi}^0_{i,j} + \mathbf{x}'_{it}\boldsymbol{\pi}^0_{i,i} + r_{it}, \tag{3.28}$$

$$s_1 + K \ll T, \qquad R_{s_1} := \max_{1\leq i\leq n} \sqrt{\frac{1}{T}\sum_{t=1}^{T} r^2_{it}} \lesssim_{\mathrm{P}} \sqrt{\frac{s_1}{T}}, \tag{3.29}$$

where $s_1 := \max_{1\leq i\leq n} \sum_{j\neq i} \|\boldsymbol{\pi}^0_{i,j}\|_0$ and $K = \|\boldsymbol{\pi}^0_{i,i}\|_0$.

To simplify the exposition and without loss of generality, I assume that $\boldsymbol{\beta}^0_i$ does not include any zero elements, implying that all regressors in $\mathbf{x}_{it}$ are relevant determinants of the dependent variable y_{it}. The assumption also guarantees identification as long as $K \geq 1$.

Remark 3.2. The sparsity assumption Assumption 3.7 (b) implies 3.7 (a). To see this, consider the case where $n = 2$. Assuming $|w_{12}| < 1$ and $|w_{21}| < 1$, the reduced form equations can be written as

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\pi}_{1,1} + \mathbf{X}_2\boldsymbol{\pi}_{1,2} + \mathbf{u}_1 \tag{3.30}$$

$$\mathbf{y}_2 = \mathbf{X}_1\boldsymbol{\pi}_{2,1} + \mathbf{X}_2\boldsymbol{\pi}_{2,2} + \mathbf{u}_2, \tag{3.31}$$

with $\boldsymbol{\pi}_{2,1} = \frac{w_{21}}{1-w_{12}w_{21}}\boldsymbol{\beta}_1$, $\boldsymbol{\pi}_{1,2} = \frac{w_{12}}{1-w_{12}w_{21}}\boldsymbol{\beta}_2$, $\boldsymbol{\pi}_{1,1} = \frac{1}{1-w_{12}w_{21}}\boldsymbol{\beta}_1$ and $\boldsymbol{\pi}_{2,2} = \frac{1}{1-w_{12}w_{21}}\boldsymbol{\beta}_2$. Suppose $\boldsymbol{\pi}_{1,2} = \mathbf{0}$, then $w_{12} = 0$ must hold given that $\boldsymbol{\beta}$ only contains non-zero elements, $|w_{12}| < 1$ and $|w_{21}| < 1$. That is, sparseness of the reduced form parameter vectors as specified in Assumption 3.7 (b) implies sparseness of the $\mathbf{W}$ matrix in Assumption 3.7 (a). Appendix C.3 shows that the results also holds for $n \geq 2$.

I maintain the following basic assumptions regarding the spatial weights matrix.

Assumption 3.8. *(a) The spatial weights matrix, $\mathbf{W}^0 = ((w^0_{ij}))$, is $n \times n$ with zeros on the diagonal, $w_{ii} = 0$. (b) The spatial weights matrix is time-invariant. (c) The row sums are bounded in absolute value, i.e., $\max_i \sum_j |w_{ij}| < 1$.*

Assumption 3.8 (a) is standard. Assumption 3.8 (b) is required as the identification strategy exploits variation over time to identify the weights matrix and is common in the spatial panel econometrics literature (see, e.g., Lee and Yu, 2010b). The assumption corresponds to parameter stability over time in time-series models. Assumption 3.8 (c) controls the degree of spatial dependence. Assumption (a) and (c) ensure that $\mathbf{I}_n - \mathbf{W}^0$ is invertible, where $\mathbf{I}_n$ is the identity matrix of dimension n. Invertibility of $\mathbf{I}_n - \mathbf{W}^0$ is required to derive the reduced form equations in (3.25).

The assumptions differ from standard assumptions in the spatial econometrics literature in two points (c.f., Kelejian and Prucha, 1998; Kelejian and Prucha, 1999). First, I do not make use of the spatial autoregressive coefficient, since the spatial autoregressive coefficient and the spatial weights are not separately identified. Second, I do not apply the row-standardisation as commonly employed or any other form of normalisation. Furthermore, it should be stressed that Assumption 3.8 does not impose any structure on the spatial weights matrix such as symmetry and, in particular, the interactions effects are allowed to be positive and negative. Recent evidence suggest that negative spatial weights are more common in practice than previously expected; see Bhattacharjee and Holly (2011, 2013) and Bailey, Holly, and Pesaran (2016).

In order to write the first and second-step Lasso estimator compactly, I introduce some additional notation. Let $\bar{\mathbf{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\boldsymbol{\pi}'_i = (\boldsymbol{\pi}'_{i,1}, \ldots, \boldsymbol{\pi}'_{i,n})$ is the corresponding parameter vector. The first-step Lasso estimator solves

$$\min \left\| \mathbf{y}_i - \bar{\mathbf{X}}\boldsymbol{\pi}_i \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\Upsilon}_{1,i}\boldsymbol{\pi}_i \right\|_1 . \tag{3.32}$$

Furthermore, let $\hat{\mathbf{y}}_i$ denote the first-step predictions and let $\hat{\mathbf{Y}} = (\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_n)$. I also define $\mathbf{w}_i = (w_{i1}, \ldots, w_{in})$ with $w_{ii} = 0$, which is the ith row of the spatial weights matrix. This allows us to write the second-step matrix of regressors as $\hat{\mathbf{G}}_i = (\hat{\mathbf{Y}}, \mathbf{X}_i)$ and define the corresponding high-dimensional parameter vector $\boldsymbol{\theta}'_i = (\mathbf{w}_i, \boldsymbol{\beta}'_i)$. The second-step Lasso solves

$$\min \left\| \mathbf{y}_i - \hat{\mathbf{G}}_i\boldsymbol{\theta}_i \right\|_2^2 + \lambda_2 \left\| \boldsymbol{\Upsilon}_{2,i}\boldsymbol{\theta}_i \right\|_1 . \tag{3.33}$$

I require both $\bar{\mathbf{X}}$ and $\hat{\mathbf{G}}_i$ to be well-behaved as stated in Assumption 3.9.[6]

ASSUMPTION 3.9. *The restricted eigenvalue and weighted restricted eigenvalue condition holds for* $\bar{\mathbf{X}}$ *and* $\hat{\mathbf{G}}_i$ *for all* $i = 1, \dots, n$ *in equations* (3.24)-(3.25).

The penalty levels are set to

$$\lambda_1 = 2c\sqrt{T}\Phi^{-1}(1 - \alpha/(2n^2K)), \tag{3.34}$$

$$\lambda_2 = 2c\sqrt{T}\Phi^{-1}(1 - \alpha/(2n(n-1+K))) \tag{3.35}$$

with $\log(1/\alpha) \lesssim \log(\max(n^2K, T))$. Note there are nK and $n-1+K$ penalized regressors in the first and second step, respectively, and n Lasso regressions in each step. The penalty loadings are again estimated using Algorithm C.2.

The convergence rates of the two-step Lasso estimator follow from Theorem 3.1. However, while the general setting in Section 3.1 allows s_1, s_2 and the number of first and second-step variables to depend on T, it is in spatial setting reasonable to assume that s_1, s_2 and n are independent of T. Therefore, the following results holds.

COROLLARY 3.1. *Consider the model in* (3.24)–(3.25). *Suppose Assumptions 3.6–3.9 hold, asymptotically valid penalty loadings are used and the penalty levels are set as in* (3.34) *and* (3.35). *Suppose that the parameters* s_1, s_2, K *and* n *do not depend on* T. *Then,*

$$\max_i \frac{1}{\sqrt{T}} \left\| \hat{\mathbf{G}}_i \hat{\boldsymbol{\theta}}_i - \mathbf{G}^\star \boldsymbol{\theta}_i^\star \right\|_2 \lesssim_{\mathrm{P}} \sqrt{\frac{\log(T)}{T}}, \tag{3.36}$$

$$\max_i \left\| \hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i^0 \right\|_1 \lesssim_{\mathrm{P}} \sqrt{\frac{\log(T)}{T}}. \tag{3.37}$$

[6]To simplify the exposition, the first and second-step Lasso also applies a penalty to $\boldsymbol{\beta}_i$ and $\boldsymbol{\pi}_{i,i}$, although I assume $\|\boldsymbol{\pi}_{i,i}\|_0 = \|\boldsymbol{\beta}_i\|_0 = K$ for identification. For better performance in finite samples, I recommend that the coefficients $\boldsymbol{\beta}_i$ and $\boldsymbol{\pi}_{i,i}$ are not penalised.

3.2.2 Post-Lasso and Thresholded Post-Lasso

The shrinkage of the Lasso estimator induces a downward bias which can be addressed by the Post-Lasso estimator. The Post-Lasso estimator treats the Lasso as a genuine model selector and applies OLS to the set of regressors for which the Lasso coefficient estimate is non-zero. In other words, Post-Lasso is OLS applied to the model selected by the Lasso. Formally, the first and second-step Post-Lasso estimator of the spatial autoregressive model are defined as

$$\tilde{\boldsymbol{\pi}}_i = \arg\min_{\boldsymbol{\pi}_i} \left\| \mathbf{y}_i - \bar{\mathbf{X}}\boldsymbol{\pi}_i \right\|_2^2 \quad \text{s.t.} \quad \text{supp}(\boldsymbol{\pi}_i) \subseteq \text{supp}(\hat{\boldsymbol{\pi}}_i), \tag{3.38}$$

$$\tilde{\boldsymbol{\theta}}_i = \arg\min_{\boldsymbol{\theta}_i} \left\| \mathbf{y}_i - \hat{\mathbf{G}}_i\boldsymbol{\theta}_i \right\|_2^2 \quad \text{s.t.} \quad \text{supp}(\boldsymbol{\theta}_i) \subseteq \text{supp}(\hat{\boldsymbol{\theta}}_i). \tag{3.39}$$

The thresholded Post-Lasso addresses the issue that the Lasso estimator often selects too many variables and that, despite the ℓ_1-penalization, many coefficient estimates are very small, but not exactly zero. The thresholded Post-Lasso applies OLS to all spatial lags for which the Post-Lasso estimate is larger than a pre-defined threshold τ.[7] While it is in general difficult to select and justify a specific threshold, in the spatial autoregressive model one can use the knowledge that $-1 < w_{ij} < 1$ and assume interaction effects that are smaller than, for example, 0.05 are negligible. For formal results on the Post-Lasso and thresholded Lasso, see Belloni and Chernozhukov (2013).

3.3 Monte Carlo simulations

This Monte Carlo study[8] explores the finite sample performance of the proposed Two-step Lasso estimator for estimating the spatial autoregressive model

$$y_{it} = \sum_{\substack{j=1 \\ j\neq i}}^{n} w_{ij} y_{jt} + \eta_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \quad t = 1, \ldots, T, \ \ i = 1, \ldots, n. \tag{3.40}$$

[7] The thresholded Lasso estimators considered in Belloni and Chernozhukov (2013) apply the threshold to the Lasso estimates whereas I apply the threshold to the Post-Lasso estimates.

[8] I am grateful to two anonymous referees who suggested useful extensions to the Monte Carlo simulations.

I consider two different spatial weights matrices. Specification 1 is given by

$$w_{ij} = \begin{cases} 1 & \text{if } |j-i| = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i,j = 1,\dots,n, \tag{3.41}$$

and specification 2 is given by

$$w_{ij} = \begin{cases} 1 & \text{if } j-i = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i,j = 1,\dots,n. \tag{3.42}$$

Subsequently, a row-standardisation is applied such that the row sum is equal to $\bar{w}$. The row-standardisation ensures that the strength of spill-over effects is constant across i. The strength of spatial interactions is determined by $\bar{w}$, which corresponds to the spatial autoregressive coefficient.

The spatial weights matrix in specification 1 has non-zeros on the sub-diagonal and super-diagonal. Thus, the weights matrix is symmetric and the number of non-zero elements is $2(n-1)$. In specification 2, only the super-diagonal elements are non-zero, implying $n-1$ non-zero elements. The structure in (3.42) corresponds to the extreme case where there are only one-way spatial effects. Specification 2 is more challenging than specification 1 as the triangular structure makes it difficult to identify the direction of causal effects. Note that, the spatial weights matrix is in principle identified if the spatial weights matrix is known to be triangular or symmetric. However, the challenge here is to estimate the spatial weights matrix without any prior knowledge. I stress that the estimation strategy does not depend on any particular structure of the spatial weights matrix, but only requires sparsity.

The parameter vector $\boldsymbol{\beta}$ is a K-dimensional vector of ones. Hence, $\boldsymbol{\beta}$ is constant across i, although the estimation method allows for spatial heterogeneity of $\boldsymbol{\beta}$. The exogenous regressors and the spatial fixed effect η_i are drawn from the standard normal distribution, i.e, $\eta_{it} \sim \mathcal{N}(0,1)$, $x_{k,it} \sim \mathcal{N}(0,1)$ for $k = 1,\dots,K$. The idiosyncratic error is drawn as $\varepsilon_{it} \sim N(0, \sigma_{it}^2)$ where

$$\sigma_{it}^2 = \frac{(1 + \mathbf{x}_{it}'\boldsymbol{\beta})^2}{\frac{1}{nT}\sum_{i,t}(1 + \mathbf{x}_{it}'\boldsymbol{\beta})^2}, \tag{3.43}$$

which induces conditional heteroskedasticity.

I consider four estimators:[9]

(a) The Two-step Lasso introduced in Section 3.2.1.

(b) Two-step Post-Lasso from Section 3.2.2.

(c) Two-step Thresholded Post-Lasso with a threshold of $\tau = 0.05$.

(d) The oracle estimator.

The oracle estimator has full knowledge about the network structure and applies 2SLS to the true model. However, the oracle estimator is infeasible as the true model is in general unknown and only serves as a benchmark. The penalty levels are defined as in (3.34)–(3.35) with $c = 1.1$ and $\alpha = \min(1/T, 0.05)$.[10] The penalty loadings are estimated by Algorithm C.2.

I present results for a range of different settings. Specifically, $n = \{30, 50, 70\}$, $T = \{50, 100, 500\}$, $K = 1$ and $\bar{w} = \{0.5, 0.7, 0.9\}$. Tables 3.1 and 3.2 report the following statistics to assess the performance of the estimators. 'False negative' is the average percentage of non-zero elements falsely identified as zero. 'False positive' is the average percentage of zero elements falsely identified as non-zero. Furthermore, let $\widehat{\mathbf{W}}_{(i)}$ be the estimate of the spatial weights matrix from the ith Monte Carlo iteration. The bias is defined as

$$\widehat{\text{bias}}_{(i)} = \frac{1}{n(n-1)} \left\| \widehat{\mathbf{W}}_{(i)} - \mathbf{W} \right\|_1, \tag{3.44}$$

where $\|\cdot\|_1$ denotes the entry-wise ℓ_1-norm and $\mathbf{W}$ is the true weights matrix. Average and median bias across iterations are reported. Note that the false negative

[9] The Lasso estimations were conducted in R using the package `glmnet` by Friedman, Hastie, and Tibshirani (2010).

[10] I have also considered, among others, $c = 1.01$ and $\alpha = 0.05/\log(T)$ and did not find significant performance differences.

and false positive rate are 0% for the oracle estimator by construction.

3.3.1 Specification 1: Symmetric matrix

The first specification in (3.41) defines a sparse, symmetric matrix. As can be seen from Table 3.1 the performance of the Two-step Lasso improves across all n and $\bar{w}$ in terms of false negative rate and bias as T increases. For example, if $n = 70$, $T = 50$ and $\bar{w} = 0.7$, in which case the model cannot be estimated by 2SLS, on average more than 88.0% of the non-zero spatial weights are identified by the Lasso. When $\bar{w} = 0.9$, this rate increases to 99.4%. However, the false positive rate of the Lasso estimator is high at approximately 10%–25% and remains high as T increases. This is in line with the known phenomenon that the Lasso estimator often selects too many variables (Bühlmann and Van de Geer, 2011).

The Two-step Post-Lasso estimator shows substantial performance improvement over the Two-step Lasso. The bias is smaller across all T and n, suggesting that Post-Lasso OLS estimation successfully addresses the shrinkage bias arising from ℓ_1-penalization. Moreover, the Two-step Post-Lasso also dominates the Two-step Lasso in terms of false negative and false positive rate. This is consistent with Belloni et al. (2012) and Belloni and Chernozhukov (2013), who show that the Post-Lasso often performs as least as good as the Lasso. However, the false positive rate is still relatively high at 5%–13% and does not seem to decrease with T. The Thresholded Post-Lasso, which sets Post-Lasso estimates below 0.05 equal to zero, improves upon the Post-Lasso in that it shows a lower false positive rate. While I refrain from recommending $\tau = 0.05$ as a general threshold, the Thresholded Post-Lasso reveals that many 'falsely positive' Post-Lasso estimates are close to zero, but not exactly zero, which explains the high false positive rate. As expected, the infeasible oracle estimator which knows the true model, exhibits the lowest bias across all n and T.

Notice that both false negative as well as false positive rate decrease with n. The decrease in the false positive rate is because the number of zero weights increases

(a) *Two-step Lasso*

$\bar{w}$	n	T	False neg.	False pos.	bias mean	bias median
0.70	30	50	19.23	19.01	0.04335	0.04266
0.70	30	100	14.62	21.22	0.03812	0.03747
0.70	30	500	5.30	24.82	0.02310	0.02305
0.70	50	50	13.96	15.13	0.02855	0.02826
0.70	50	100	8.86	17.58	0.02653	0.02626
0.70	50	500	3.14	22.13	0.01725	0.01718
0.70	70	50	11.93	12.35	0.02132	0.02130
0.70	70	100	6.09	15.36	0.02071	0.02055
0.70	70	500	1.89	20.39	0.01454	0.01446
0.90	30	50	2.07	14.99	0.02360	0.02346
0.90	30	100	0.97	14.92	0.02002	0.01992
0.90	30	500	0.18	18.35	0.01372	0.01365
0.90	50	50	1.03	12.69	0.01495	0.01489
0.90	50	100	0.29	13.11	0.01354	0.01354
0.90	50	500	0.03	15.24	0.00935	0.00931
0.90	70	50	0.52	10.37	0.01048	0.01044
0.90	70	100	0.15	12.17	0.01058	0.01054
0.90	70	500	0.01	13.60	0.00759	0.00758

(b) *Two-step Post-Lasso*

False neg.	False pos.	bias mean	bias median
8.83	11.55	0.03649	0.03606
5.64	11.89	0.03090	0.03083
2.18	12.69	0.02029	0.02022
4.81	9.38	0.02453	0.02432
1.90	10.12	0.02187	0.02185
0.58	10.80	0.01519	0.01519
3.84	7.65	0.01829	0.01828
0.87	8.96	0.01746	0.01741
0.17	9.78	0.01277	0.01273
0.76	9.83	0.01823	0.01806
0.25	7.93	0.01267	0.01254
0.12	7.62	0.00794	0.00787
0.29	9.27	0.01364	0.01349
0.04	7.94	0.00947	0.00944
0.01	6.07	0.00513	0.00512
0.17	7.26	0.00931	0.00928
0.02	8.26	0.00854	0.00852
0.00	5.50	0.00409	0.00408

(c) *Thresholded Post-Lasso with* $\tau = 0.05$

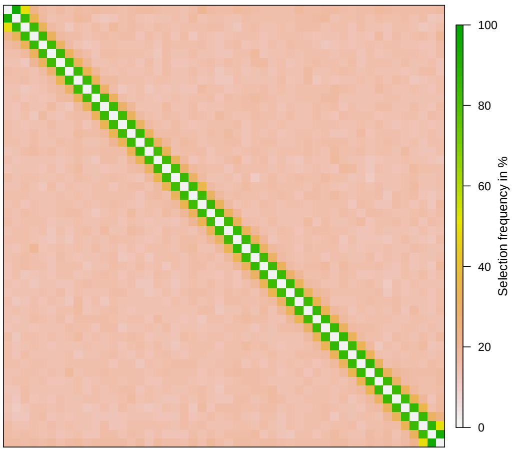
$\bar{w}$	n	T	False neg.	False pos.	bias mean	bias median
0.70	30	50	10.04	6.36	0.02561	0.02523
0.70	30	100	6.40	6.44	0.02165	0.02145
0.70	30	500	2.36	6.60	0.01368	0.01360
0.70	50	50	5.71	4.85	0.01621	0.01609
0.70	50	100	2.28	5.21	0.01434	0.01430
0.70	50	500	0.65	5.36	0.00968	0.00969
0.70	70	50	4.59	3.84	0.01200	0.01197
0.70	70	100	1.13	4.48	0.01114	0.01110
0.70	70	500	0.20	4.69	0.00791	0.00790
0.90	30	50	1.57	5.00	0.01316	0.01298
0.90	30	100	0.42	3.54	0.00923	0.00915
0.90	30	500	0.13	2.35	0.00528	0.00525
0.90	50	50	1.06	4.30	0.00877	0.00871
0.90	50	100	0.12	3.25	0.00620	0.00615
0.90	50	500	0.01	1.54	0.00310	0.00308
0.90	70	50	0.56	3.14	0.00589	0.00584
0.90	70	100	0.09	3.24	0.00513	0.00510
0.90	70	500	0.00	1.21	0.00230	0.00229

(d) *Oracle estimator*

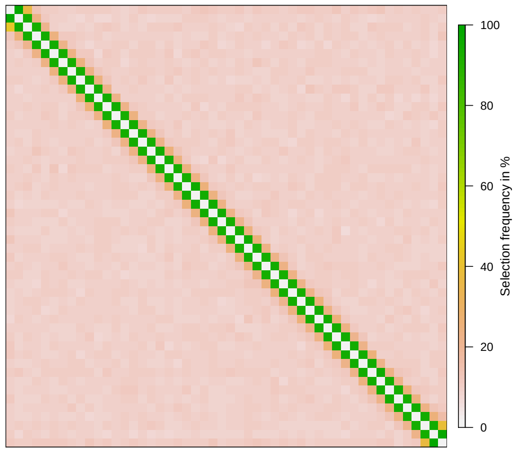
False neg.	False pos.	bias mean	bias median
–	–	0.02742	0.02444
–	–	0.01915	0.01779
–	–	0.00762	0.00753
–	–	0.01533	0.01462
–	–	0.01135	0.01057
–	–	0.00455	0.00453
–	–	0.01114	0.01055
–	–	0.00816	0.00771
–	–	0.00325	0.00323
–	–	0.02576	0.02358
–	–	0.02039	0.01887
–	–	0.01250	0.01237
–	–	0.01562	0.01418
–	–	0.01216	0.01132
–	–	0.00751	0.00744
–	–	0.01087	0.01020
–	–	0.00860	0.00816
–	–	0.00533	0.00532

Table 3.1: *Monte Carlo results: Specification 1. 'False neg.' denotes false negative rate in %. 'False pos.' denotes false positive rate in %. The bias is defined in* (3.44)*. The false negative and false positive rate is 0% for the oracle estimator by construction. Note that the oracle estimator is infeasible in practice and serves only as a reference point. The number of replications is 1,000. The number of exogenous regressors is* $K = 1$.

(a) *Two-step Lasso*



(b) *Two-step Post-Lasso*



(c) *Thresholded Post-Lasso with* $\tau = 0.05$

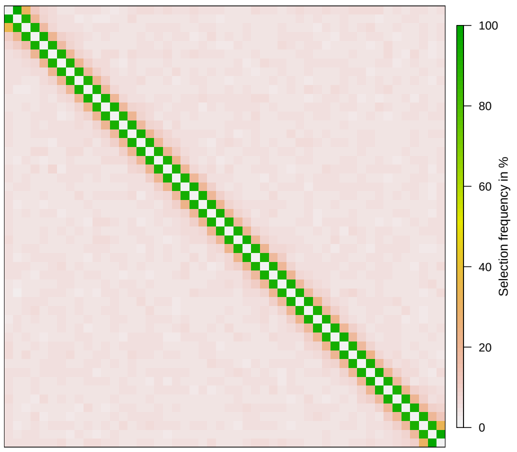


Figure 3.2: *Recovery of spatial weight matrix with* $n = 50$ *and* $T = 50$*: Specification 1. The colour indicates the frequency at which each spatial weight is identified as being non-zero.*

(a) *Two-step Lasso*

$\bar{w}$	n	T	False neg.	False pos.	bias mean	bias median
0.50	30	50	41.60	19.36	0.05258	0.04915
0.50	30	100	32.82	20.11	0.04119	0.03952
0.50	30	500	7.80	18.61	0.01904	0.01877
0.50	50	50	39.98	15.89	0.03820	0.03512
0.50	50	100	30.54	16.98	0.03055	0.02976
0.50	50	500	7.26	17.07	0.01497	0.01483
0.50	70	50	40.38	12.76	0.02885	0.02730
0.50	70	100	28.60	14.98	0.02477	0.02447
0.50	70	500	6.90	15.95	0.01290	0.01284

(b) *Two-step Post-Lasso*

False neg.	False pos.	bias mean	bias median
36.30	13.00	0.05679	0.05388
25.70	13.08	0.04681	0.04572
4.46	13.10	0.02784	0.02779
32.20	10.46	0.03943	0.03689
20.85	10.80	0.03353	0.03326
3.22	11.66	0.02276	0.02274
32.13	8.41	0.02944	0.02814
17.51	9.37	0.02653	0.02642
2.32	10.59	0.01958	0.01956

(c) *Thresholded Post-Lasso with* $\tau = 0.05$

$\bar{w}$	n	T	False neg.	False pos.	bias mean	bias median
0.50	30	50	37.53	7.81	0.03843	0.03722
0.50	30	100	26.45	8.03	0.03247	0.03218
0.50	30	500	4.54	7.72	0.01811	0.01806
0.50	50	50	33.53	6.00	0.02573	0.02498
0.50	50	100	21.68	6.36	0.02257	0.02252
0.50	50	500	3.31	6.72	0.01441	0.01438
0.50	70	50	33.43	4.72	0.01946	0.01899
0.50	70	100	18.29	5.40	0.01762	0.01758
0.50	70	500	2.40	5.99	0.01224	0.01223

(d) *Oracle estimator*

False neg.	False pos.	bias mean	bias median
–	–	0.01335	0.01186
–	–	0.00934	0.00861
–	–	0.00351	0.00348
–	–	0.00794	0.00724
–	–	0.00559	0.00524
–	–	0.00210	0.00208
–	–	0.00561	0.00512
–	–	0.00388	0.00371
–	–	0.00149	0.00149

Table 3.2: *Monte Carlo results: Specification 2. See notes in Table 3.1.*

with n as a proportion of the total number of off-diagonal elements in $\mathbf{W}$. The same situation holds in many real spatial applications where the number of neighbours of a region are bounded. In turn such boundedness is a necessity for spatial stationarity; see Assumption 3.8 and the spatial granularity condition in Chudik, Pesaran, and Tosetti (2011). In the large n setting, standard least squares methods would not be feasible because of high-dimensionality, which underlines the important advantage of the Lasso-based methods proposed in this chapter.

Figure 3.2 shows how often each w_{ij} is identified as being non-zero by the estimators for $n = T = 50$. It can be seen that the two-step procedures successfully recover the spatial structure in (3.41). Note that weights to the left of the sub-diagonal and to the right of the super-diagonal (i.e., $w_{13}, w_{24}, w_{31}, \dots,$ *etc.*) are falsely selected slightly more often relative to other weights. This is likely due to indirect effects. For example, w_{13} is selected slightly more often relative to other zero elements as y_{3t} has an indirect effect on y_{1t} through y_{2t}.

3.3.2 Specification 2: Triangular matrix

As expected, the performance under specification 2 is not as satisfactory as for specification 1. Table 3.2 shows that false negative rate and bias decrease in T

for all three Lasso-based estimators. As in specification 1, the Two-step Post-Lasso outperforms the two-step Lasso in terms of the false negative rate. The Thresholded Lasso mainly differs from the Two-step Post-Lasso in that the false positive rate is lower. Figure 3.3 shows the selection frequency for $n = T = 50$ and $n = 50, T = 500$. For $T = 50$ (*left*), it can clearly be seen that the elements in the sub-diagonal are selected more often relative to other non-zero elements, stressing the difficulty of identifying the direction of the effects in small samples. This problem reduces with T and is negligible for $T = 500$ (*right*).

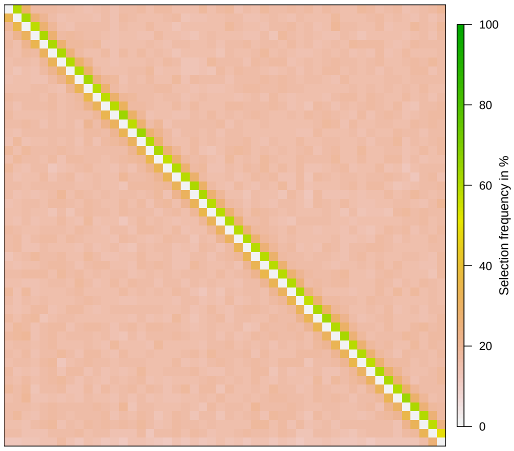
Overall, the Two-step Lasso performs well in recovering the network structure, even for the more challenging specification 2. However, we observe that the Two-step Lasso selects too many spatial lags in small samples, although the performance improves substantially with T in terms of bias and false negative rate. The Two-step Post-Lasso outperforms the Two-step Lasso in terms of bias and selection performance. The thresholded Lasso demonstrates that many falsely identified weights are associated with small coefficient estimates.

3.4 Conclusion

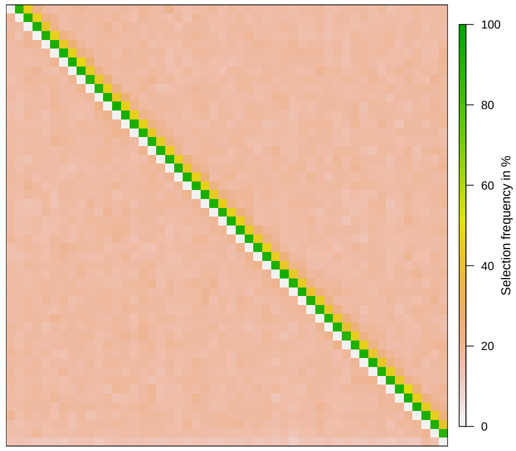
The identification of interaction effects is crucial for the understanding of how individuals, firms and regions interact. However, to date there is a lack of methods that allow the estimation of interaction effects, particularly when the spatial dimension is large. Thus, most applied spatial econometric research uses *ad hoc* specifications to incorporate interaction effects. The lack of estimation strategies may also explain why interaction effects in socio-economic processes are often ignored.

I propose a two-step procedure based on the Lasso estimator that accounts for reverse causality and allows estimating interaction effects between units in a spatial autoregressive panel model without requiring any prior knowledge about the network structure. The identifying assumption is sparsity. The two-step estimator can be implemented based on fast algorithms available for the Lasso estimator (e.g. Friedman, Hastie, and Tibshirani, 2010). The estimation methodology is attrac-

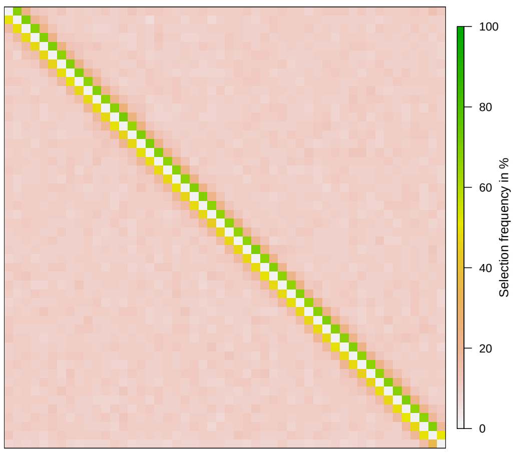
(a) *Two-step Lasso with* $T = 50$



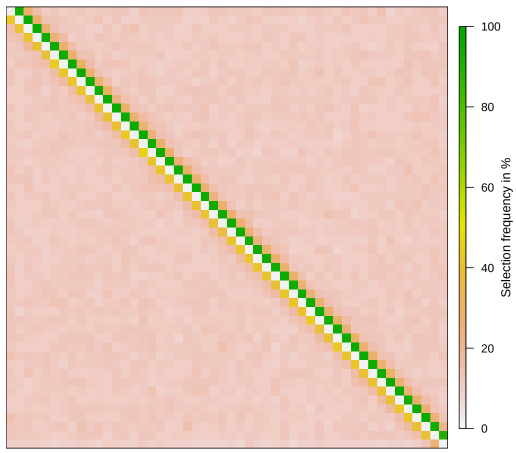
(b) *Two-step Lasso with* $T = 500$



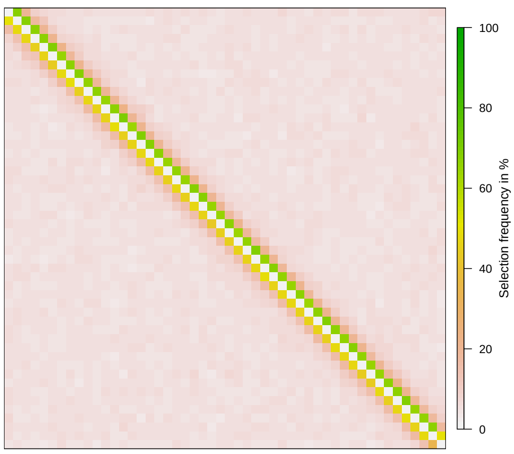
(c) *Two-step Post-Lasso with* $T = 50$



(d) *Two-step Post-Lasso with* $T = 500$



(e) *Thresholded Post-Lasso with* $T = 50$



(f) *Thresholded Post-Lasso with* $T = 500$

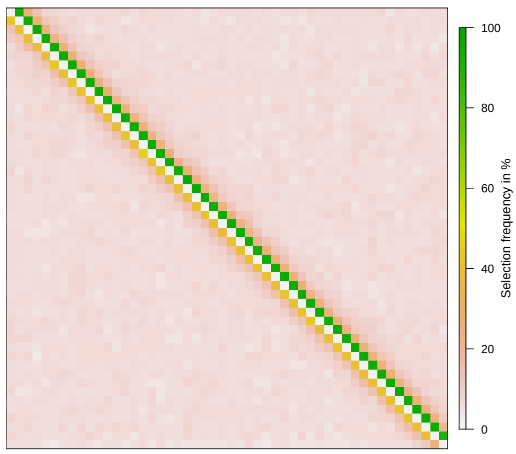


Figure 3.3: *Recovery of spatial weight matrix with* $n = 50$ *and* $T = \{50, 500\}$*: Specification 2. The Thresholded Post-Lasso uses* $\tau = 0.05$*. See also notes in Figure 3.2.*

tive for applied research as the Lasso estimator also serves as a model selector and, hence, is relatively robust to misspecification.

I derive convergence rates for a general Two-step Lasso estimator which allows for the number of endogenous regressors and the number of instruments to be larger than the sample size. I then apply the two-step estimator to the spatial autoregressive panel model. Monte Carlo results confirm that the estimation method recovers the structure of the spatial weights matrix. However, Monte Carlo results also show a tendency for over-selection of spatial weights. The Two-step Post-Lasso estimator, which in each step applies OLS to the model selected by the Lasso, outperforms the Two-step Lasso in terms of bias, false positive and false negative rate.

The use of the Two-step Lasso raises several issues shared with other Lasso-type estimators. Controlling uncertainty and conducting inference with the Lasso is challenging and remains an area of ongoing research. A recent important contribution is the Lasso significance tests due to Lockhart et al. (2014).

In addition, the choice of an optimal penalty level is an important issue. Penalized estimators typically select the penalty level oriented towards optimising predictive performance, which may not be appropriate if the purpose is structure recovery. The optimal penalty used here is not based on cross-validation or other model selection criteria commonly employed and is therefore not directly subject to this criticism. Specifically, we follow Bickel, Ritov, and Tsybakov (2009) and Belloni et al. (2012) in choosing the smallest penalty level that dominates the noise of the problem. Our Monte Carlo results show that the proposed method works well in the structure discovery context.

This work suggests several lines of future research. First, given that the Two-step Post-Lasso outperforms the Two-step Lasso, formal results for the Two-step Post-Lasso are required. Secondly, the methodology can be extended to the Square-root Lasso and Square-root Post-Lasso. The main advantage of the Square-root Lasso is that the optimal penalty level does not depend on the unknown error variance

(Belloni, Chernozhukov, and Wang, 2011; Belloni, Chernozhukov, and Wang, 2014). Hence, further performance improvements seem possible. Thirdly, instead of relying on a Two-step Lasso estimation method, an alternative estimation strategy may be based on the recent work by Fan and Liao (2014) or Gautier and Tsybakov (2014) who allow for endogeneity in high dimensions. These one-step procedures potentially lead to better performances and may facilitate accounting for uncertainty in model selection and estimation. In this respect, I emphasise that the primary contribution of this chapter is to suggest that statistical methods allowing for endogeneity in a high-dimensional setting can be exploited to identify the spatial weights matrix. Lastly, the estimation methodology could also be combined with established large T panel estimators. In particular, it would be of interest to develop a more general framework that allows for both spatial effects and common factors as in Holly, Pesaran, and Yamagata (2010) as well as temporal dynamics. However, these ideas are retained for the future.

Appendix of Chapter 3

C.1 Proof of Theorem 3.1

Setting. At first, I summarise the setting and introduce some notation. The model in (3.2)–(3.3) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^\star + \mathbf{e}, \quad \mathbf{X} = \mathbf{Z}\boldsymbol{\Pi}^\star + \mathbf{U}. \tag{C.1}$$

Thus, the reduced form equation for $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\Pi}^\star\boldsymbol{\beta}^\star + \mathbf{U}\boldsymbol{\beta}^\star + \mathbf{e}. \tag{C.2}$$

Using Assumption 3.3,

$$\mathbf{y} = \mathbf{X}^\star\boldsymbol{\beta}^0 + \mathbf{X}^\star(\boldsymbol{\beta}^\star - \boldsymbol{\beta}^0) + \boldsymbol{\varepsilon} = \mathbf{X}^\star\boldsymbol{\beta}^0 + \mathbf{r} + \boldsymbol{\varepsilon}, \tag{C.3}$$

where $\mathbf{X}^\star = \mathbf{Z}\boldsymbol{\Pi}^\star$, $\boldsymbol{\varepsilon} = \mathbf{U}\boldsymbol{\beta}^\star + \mathbf{e}$, $\mathbf{r} = \mathbf{X}^\star(\boldsymbol{\beta}^\star - \boldsymbol{\beta}^0)$ with $1/\sqrt{T}\,\|\mathbf{r}\|_2 = R_{s_2}$ and $\boldsymbol{\beta}^0$ is the target parameter vector. As $\mathbf{X}^\star$ is unknown, the second step uses $\hat{\mathbf{X}} = \mathbf{X}^\star + \hat{\mathbf{V}}$ instead,

$$\mathbf{y} = (\hat{\mathbf{X}} - \hat{\mathbf{V}})\boldsymbol{\beta}^0 + \mathbf{r} + \boldsymbol{\varepsilon} = \hat{\mathbf{X}}\boldsymbol{\beta}^0 - \hat{\mathbf{V}}\boldsymbol{\beta}^0 + \mathbf{r} + \boldsymbol{\varepsilon} = \hat{\mathbf{X}}\boldsymbol{\beta}^0 + \mathbf{r} + \mathbf{m} + \boldsymbol{\varepsilon}, \tag{C.4}$$

where $\mathbf{m} := -\hat{\mathbf{V}}\boldsymbol{\beta}^0$ is the matrix of prediction errors from the first step weighted by the target parameter vector. Recall, the second-step Lasso estimator solves

$$\min \frac{1}{T}\left\|\mathbf{y} - \hat{\mathbf{X}}\boldsymbol{\beta}\right\|_2^2 + \frac{\lambda_2}{T}\left\|\hat{\boldsymbol{\Upsilon}}_2\boldsymbol{\beta}\right\|_2. \tag{C.5}$$

Let $Q(\boldsymbol{\beta}) = \frac{1}{T}\|\mathbf{y} - \hat{\mathbf{X}}\boldsymbol{\beta}\|_2^2$ and $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$. Lastly, I define the active set $\Omega_2 = \text{supp}(\boldsymbol{\beta}^0)$ and its cardinality $|\Omega_2| = s_2$.

The general approach in the following steps is based on Belloni et al. (2012) and Bickel, Ritov, and Tsybakov (2009), but accounts for the prediction error from the first step, $1/\sqrt{T}\,\|\mathbf{m}\|_2$.

Non-asymptotic ℓ_2-prediction norm bound. In this step, I bound $1/\sqrt{T}\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\|_2$ and treat $1/\sqrt{T}\,\|\mathbf{m}\|_2$ as given. The convergence rate of $1/\sqrt{T}\,\|\mathbf{m}\|_2$ will be derived in the next step.

By optimality of the Lasso estimate $\hat{\boldsymbol{\beta}}$,

$$Q(\hat{\boldsymbol{\beta}}) + \frac{\lambda_2}{T}\left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\beta}}\right\|_1 \leq Q(\boldsymbol{\beta}^0) + \frac{\lambda_2}{T}\left\|\hat{\boldsymbol{\Upsilon}}_2\boldsymbol{\beta}^0\right\|_1 \tag{C.6}$$

$$Q(\hat{\boldsymbol{\beta}}) - Q(\boldsymbol{\beta}^0) \leq \frac{\lambda_2}{T}\left(\left\|\hat{\boldsymbol{\Upsilon}}_2\boldsymbol{\beta}^0\right\|_1 - \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\beta}}\right\|_1\right). \tag{C.7}$$

That is, the Lasso objective function evaluated at $\hat{\boldsymbol{\beta}}$ is at least as small as the objective function evaluated at $\boldsymbol{\beta}^0$. With regard to the right-hand side,

$$\left\|\hat{\boldsymbol{\Upsilon}}_2\boldsymbol{\beta}^0\right\|_1 - \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\beta}}\right\|_1 = \left\|\hat{\boldsymbol{\Upsilon}}_2\boldsymbol{\beta}^0_{\Omega_2}\right\|_1 - \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\beta}}_{\Omega_2}\right\|_1 - \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\beta}}_{\overline{\Omega}_2}\right\|_1 \tag{C.8}$$

$$\leq \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\Omega_2}\right\|_1 - \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1 \tag{C.9}$$

using $\boldsymbol{\beta}^0 = \boldsymbol{\beta}^0_{\Omega_2}$, $\hat{\boldsymbol{\beta}}_{\overline{\Omega}_2} = \hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}$ and, by reverse triangle inequality, $\left\|\hat{\boldsymbol{\Upsilon}}_2\boldsymbol{\beta}^0_{\Omega_2}\right\|_1 - \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\beta}}_{\Omega_2}\right\|_1 \leq \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\beta}}_{\Omega_2} - \hat{\boldsymbol{\Upsilon}}_2\boldsymbol{\beta}^0_{\Omega_2}\right\|_1$. With regard to the left-hand side,

$$TQ(\hat{\boldsymbol{\beta}}) - TQ(\boldsymbol{\beta}^0) = \left\|\mathbf{y} - \hat{\mathbf{X}}\hat{\boldsymbol{\beta}}\right\|_2^2 - \left\|\mathbf{y} - \hat{\mathbf{X}}\boldsymbol{\beta}^0\right\|_2^2 \tag{C.10}$$

$$= -2\mathbf{y}'\hat{\mathbf{X}}\hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\beta}}'\hat{\mathbf{X}}'\hat{\mathbf{X}}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{0\prime}\hat{\mathbf{X}}'\hat{\mathbf{X}}\boldsymbol{\beta}^0. \tag{C.11}$$

Substracting $\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2^2$ from both sides gives

$$Q(\hat{\boldsymbol{\beta}}) - Q(\boldsymbol{\beta}^0) - \frac{1}{T}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2^2 = -\frac{2}{T}\boldsymbol{\varepsilon}'\hat{\mathbf{X}}\hat{\boldsymbol{\delta}} - \frac{2}{T}\mathbf{r}'\hat{\mathbf{X}}\hat{\boldsymbol{\delta}} - \frac{2}{T}\mathbf{m}'\hat{\mathbf{X}}\hat{\boldsymbol{\delta}} \tag{C.12}$$

$$= -\frac{2}{T}\boldsymbol{\varepsilon}'\hat{\mathbf{X}}(\hat{\boldsymbol{\Upsilon}}_2)^{-1}\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}} - \frac{2}{T}\mathbf{r}'\hat{\mathbf{X}}\hat{\boldsymbol{\delta}} - \frac{2}{T}\mathbf{m}'\hat{\mathbf{X}}\hat{\boldsymbol{\delta}} \tag{C.13}$$

$$\geq_{(i)} -\boldsymbol{S}_2'\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}} - 2R_{s_2}\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 - \frac{2}{T}\|\mathbf{m}\|_2\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 \tag{C.14}$$

$$\geq_{(ii)} -\|\boldsymbol{S}_2\|_\infty\left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}\right\|_1 - 2R_{s_2}\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 - \frac{2}{T}\|\mathbf{m}\|_2\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 \tag{C.15}$$

$$\geq_{(iii)} -\frac{\lambda_2}{cT}\left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}\right\|_1 - 2R_{s_2}\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 - \frac{2}{T}\|\mathbf{m}\|_2\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2. \tag{C.16}$$

Here, (i) uses the Cauchy-Schwarz inequality and the definitions $R_{s_2} = \frac{1}{\sqrt{T}}\|\mathbf{r}\|_2$ and $\boldsymbol{S}_2 := \frac{2}{T}(\hat{\boldsymbol{\Upsilon}}_2)^{-1}\hat{\mathbf{X}}'\boldsymbol{\varepsilon}$. (ii) uses the Hölder inequality. (iii) uses $\lambda_2 \geq cT\|\boldsymbol{S}_2\|_\infty$ which holds as $T \to \infty$. Note that, by substituting for $\|\boldsymbol{S}_2\|_\infty$, the random component of the equation is eliminated. Combining (C.7), (C.9) and (C.16) yields

$$\begin{aligned}
\frac{1}{T}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2^2 &\leq 2R_{s_2}\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + \frac{2}{T}\|\mathbf{m}\|_2\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + \frac{\lambda_2}{cT}\left(\left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\Omega_2}\right\|_1 + \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1\right) \\
&\quad + \frac{\lambda_2}{T}\left(\left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\Omega_2}\right\|_1 - \left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1\right) &\text{(C.17)} \\
&\leq 2R_{s_2}\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + \frac{2}{T}\|\mathbf{m}\|_2\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + \left(1+\frac{1}{c}\right)\frac{\lambda_2}{T}\left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\Omega_2}\right\|_1 \\
&\quad - \left(1-\frac{1}{c}\right)\frac{\lambda_2}{T}\left\|\hat{\boldsymbol{\Upsilon}}_2\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1 &\text{(C.18)} \\
&\leq \left(2R_{s_2} + \frac{2}{\sqrt{T}}\|\mathbf{m}\|_2\right)\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + \left(1+\frac{1}{c}\right)\frac{\lambda_2}{T}u\left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\Omega_2}\right\|_1 \\
&\quad - \left(1-\frac{1}{c}\right)\frac{\lambda_2}{T}l\left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1 &\text{(C.19)}
\end{aligned}$$

with $0 < l \leq 1 \leq u$. The last step assumes that $\hat{\boldsymbol{\Upsilon}}_2$ is asymptotically valid. Specifically, there are two constants u and l such that $l\boldsymbol{\Upsilon}_2^0 \leq \hat{\boldsymbol{\Upsilon}}_2 \leq u\boldsymbol{\Upsilon}_2^0$ where $l \to_{\mathrm{P}} 1$ and $u \to_{\mathrm{P}} \bar{u}$ with $\bar{u} \geq 1$ (Belloni et al., 2012).

I distinguish between two cases. Case A: If $1/\sqrt{T}\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\|_2 \leq 2R_{s_2} + 2/\sqrt{T}\|\mathbf{m}\|_2$, the bound is established by assumption. Case B: If $1/\sqrt{T}\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\|_2 > 2R_{s_2} + 2/\sqrt{T}\|\mathbf{m}\|_2$, the above equation yields

$$\left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1 \leq c_0\left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\Omega_2}\right\|_1, \tag{C.20}$$

where $c_0 = u(c+1)/(l(c-1))$ which allows to invoke the weighted restricted eigenvalue condition,

$$\frac{1}{\sqrt{T}} \left\|\hat{\mathbf{X}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\right\|_2 \leq 2R_{s_2} + \frac{2}{\sqrt{T}} \|\mathbf{m}\|_2 + \left(1 + \frac{1}{c}\right) u \frac{\lambda_2}{T} \frac{\sqrt{s_2}}{\kappa^{\omega}_{c_0}(\hat{\mathbf{X}})}. \tag{C.21}$$

This establishes the non-asymptotic ℓ_2-prediction norm bound, but takes the prediction error $1/\sqrt{T}\,\|\mathbf{m}\|_2$ from the first step as given. Note that setting $\mathbf{m} = \mathbf{0}$ yields the bound in Lemma 6 in Belloni et al. (2012).

Convergence rate of $1/\sqrt{T}\|\mathbf{m}\|_2$. In this step, I derive the convergence rate for $1/\sqrt{T}\|\mathbf{m}\|_2$. Notice that

$$\|\mathbf{m}\|_2 = \left\|\hat{\mathbf{V}}\boldsymbol{\beta}^0\right\|_2 = \left\|\hat{\mathbf{V}}_{\Omega_2}\boldsymbol{\beta}^0_{\Omega_2}\right\|_2 \leq \left\|\hat{\mathbf{V}}_{\Omega_2}\right\|_F \left\|\boldsymbol{\beta}^0\right\|_2 = \left\|\boldsymbol{\beta}^0\right\|_2 \left(\sum_{j\in\Omega_2}\sum_{i}\hat{v}_{ij}^2\right)^{1/2} \tag{C.22}$$

$$\leq \left\|\boldsymbol{\beta}^0\right\|_2 \sum_{j\in\Omega_2}\left(\sum_{i}\hat{v}_{ij}^2\right)^{1/2} = \left\|\boldsymbol{\beta}^0\right\|_2 \sum_{j\in\Omega_2}\left\|\hat{\mathbf{V}}_j\right\|_2 \tag{C.23}$$

$$\leq \left\|\boldsymbol{\beta}^0\right\|_2 s_2 \max_{1\leq j\leq\bar{p}}\left\|\hat{\mathbf{V}}_j\right\|_2, \tag{C.24}$$

where $\hat{\mathbf{V}}_j = \mathbf{Z}\hat{\boldsymbol{\pi}}_j - \mathbf{Z}\boldsymbol{\pi}^\star_j$. By Theorem 1 in Belloni et al. (2012),

$$\max_{1\leq j\leq\bar{p}} \frac{1}{\sqrt{T}}\left\|\hat{\mathbf{V}}_j\right\|_2 \lesssim_{\mathrm{P}} \sqrt{\frac{s_1\log(\max(L\bar{p},T))}{T}}. \tag{C.25}$$

Substituting (C.25) into (C.24) and assuming $\|\boldsymbol{\beta}^0\|_2 \lesssim s_2$,

$$\frac{1}{\sqrt{T}}\|\mathbf{m}\|_2 = \frac{1}{\sqrt{T}}\left\|\hat{\mathbf{V}}\boldsymbol{\beta}^0\right\|_2 \lesssim_{\mathrm{P}} s_2^2\sqrt{\frac{s_1\log(\max(L\bar{p},T))}{T}}. \tag{C.26}$$

Convergence rate of ℓ_2-prediction norm bound. The non-asymptotic ℓ_2-prediction bound and the convergence rate for $1/\sqrt{T}\,\|\mathbf{m}\|_2$ allow to derive the ℓ_2-prediction norm convergence rate. Note that $\lambda_2 \lesssim \sqrt{T\log(L\bar{p}/\alpha)}$ and $R_{s_2} \lesssim_{\mathrm{P}} \sqrt{s_2/T}$

by assumption. By (C.21) and substituting the convergence rate of $1/\sqrt{T}\|\hat{\mathbf{V}}\boldsymbol{\beta}^0\|_2$,

$$\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)\right\|_2 \lesssim_{\mathrm{P}} \sqrt{\frac{s_2}{T}} + s_2^2\sqrt{\frac{s_1\log(\max(L\bar{p},T))}{T}} + \sqrt{\frac{s_2\log(\max(L\bar{p},T))}{T}} \tag{C.27}$$

$$\lesssim_{\mathrm{P}} s_2^2\sqrt{\frac{s_1\log(\max(L\bar{p},T))}{T}}. \tag{C.28}$$

However, the aim is to bound the deviations from $\hat{\mathbf{X}}\hat{\boldsymbol{\beta}}$ to $\mathbf{X}^{\star}\boldsymbol{\beta}^{\star}$. Hence, I apply the triangle inequality

$$\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\beta}}-\mathbf{X}^{\star}\boldsymbol{\beta}^{\star}\right\|_2 = \left\|(\hat{\mathbf{X}}\hat{\boldsymbol{\beta}}-\hat{\mathbf{X}}\boldsymbol{\beta}^0)+(\hat{\mathbf{X}}\boldsymbol{\beta}^0-\mathbf{X}^{\star}\boldsymbol{\beta}^0)+(\mathbf{X}^{\star}\boldsymbol{\beta}^0-\mathbf{X}^{\star}\boldsymbol{\beta}^{\star})\right\|_2 \tag{C.29}$$

$$\leq \frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\beta}}-\hat{\mathbf{X}}\boldsymbol{\beta}^0\right\|_2 + \frac{1}{\sqrt{T}}\left\|\hat{\mathbf{V}}\boldsymbol{\beta}^0\right\|_2 + R_{s_2} \tag{C.30}$$

$$\lesssim_{\mathrm{P}} s_2^2\sqrt{\frac{s_1\log(\max(L\bar{p},T))}{T}}. \tag{C.31}$$

Non-asymptotic ℓ_1-parameter norm bound. Again, I distinguish between two cases. Case A: $\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\|_1 \leq 2c_0\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\Omega_2}\|_1$. Then, I can use the definition of the weighted restricted eigenvalue

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 \leq (1+2c_0)\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\Omega_2}\|_1 \leq (1+2c_0)\frac{\sqrt{s_2}}{\kappa_{2c_0}^{\omega}(\hat{\mathbf{X}})\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2. \tag{C.32}$$

Case B: If $\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\|_1 > 2c_0\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\Omega_2}\|_1$, then by (C.19) $2R_{s_2}+2/\sqrt{T}\|\mathbf{m}\|_2 \geq \frac{1}{\sqrt{T}}\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\|_2$ must hold. Also, from (C.19)

$$\left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1 \leq \left(2R_{s_2}+\frac{2}{\sqrt{T}}\|\mathbf{m}\|_2-\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2\right)\frac{T}{\lambda_2}\frac{1}{\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2\left(\frac{c}{l(c-1)}\right) + c_0\left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\Omega_2}\right\|_1 \tag{C.33}$$

$$\leq \left(R_{s_2}+\frac{1}{\sqrt{T}}\|\mathbf{m}\|_2\right)^2\frac{T}{\lambda_2}\left(\frac{c}{l(c-1)}\right)+\frac{1}{2}\left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\right\|_1 \tag{C.34}$$

$$\leq 2\left(R_{s_2}+\frac{1}{\sqrt{T}}\|\mathbf{m}\|_2\right)^2\frac{T}{\lambda_2}\left(\frac{c}{l(c-1)}\right) \tag{C.35}$$

where the second step uses $\max_{x\geq 0} x(2a-x) \leq a^2$. Next, by Case B assumption,

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\Omega_2}\|_1 < \frac{1}{2c_0}\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\|_1 \tag{C.36}$$

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 < \left(1+\frac{1}{2c_0}\right)\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}_{\overline{\Omega}_2}\|_1 \tag{C.37}$$

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 < \left(1+\frac{1}{2c_0}\right)2\left(R_{s_2}+\frac{1}{\sqrt{T}}\|\mathbf{m}\|_2\right)^2\frac{T}{\lambda_2}\left(\frac{c}{l(c-1)}\right) \tag{C.38}$$

Combining (C.32) and (C.38),

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 \leq (1+2c_0)\frac{\sqrt{s_2}}{\kappa_{2c_0}^{\omega}(\hat{\mathbf{X}})\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + \left(1+\frac{1}{2c_0}\right)2\left(R_{s_2}+\frac{1}{\sqrt{T}}\|\mathbf{m}\|_2\right)^2 \times\frac{T}{\lambda_2}\left(\frac{c}{l(c-1)}\right) \tag{C.39}$$

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 \leq 3c_0\frac{\sqrt{s_2}}{\kappa_{2c_0}^{\omega}(\hat{\mathbf{X}})\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + 3\frac{c_0T}{\lambda_2}\left(R_{s_2}+\frac{1}{\sqrt{T}}\|\mathbf{m}\|_2\right)^2 \tag{C.40}$$

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 \leq 3c_0\frac{\sqrt{s_2}}{\kappa_{2c_0}^{\omega}(\hat{\mathbf{X}})\sqrt{T}}\left\|\hat{\mathbf{X}}\hat{\boldsymbol{\delta}}\right\|_2 + 3\frac{c_0T}{\lambda_2}\left(R_{s_2}^2+\frac{1}{T}\|\mathbf{m}\|_2^2+2R_{s_2}\frac{1}{\sqrt{T}}\|\mathbf{m}\|_2\right), \tag{C.41}$$

where I use that $c/(l(c-1)) \leq c_0$ and $1+1/(2c_0) \leq 3/2$. Again, setting $\mathbf{m} = \mathbf{0}$ yields the result in Belloni et al. (2012, Lemma 6).

ℓ_1-parameter norm convergence rate. In the last step, I derive the ℓ_1-convergence rates. I assume, as stated in the Theorem, that s_1 and s_2 do not depend on T. This assumption may be strong in general, but reasonable in the spatial autoregressive panel model where s_1 and s_2 are likely to depend on n, but not on T.

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 \lesssim_{\mathrm{P}} \sqrt{s_1s_2}s_2^2\sqrt{\frac{\log(\max(L\bar{p},T))}{T}} + \frac{s_2}{\sqrt{T\log\left(\max(L\bar{p},T)\right)}} \tag{C.42}$$

$$+ s_2^4s_1\sqrt{\frac{\log(\max(L\bar{p},T))}{T}} + s_2^2\sqrt{\frac{s_1s_2}{T}} \tag{C.43}$$

$$\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\|_1 \lesssim_{\mathrm{P}} \sqrt{\frac{\log(\max(L\bar{p},T))}{T}}. \tag{C.44}$$

Lastly,

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right\|_1 = \left\|(\boldsymbol{\Upsilon}_2^0)^{-1}\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\right\|_1 \leq \left\|(\boldsymbol{\Upsilon}_2^0)^{-1}\right\|_\infty \left\|\boldsymbol{\Upsilon}_2^0\hat{\boldsymbol{\delta}}\right\|_1 \tag{C.45}$$

$$\lesssim_{\mathrm{P}} \sqrt{\frac{\log(\max(L\bar{p}, T))}{T}}. \tag{C.46}$$

C.2 Algorithm for estimating penalty loadings

The algorithm is reproduced from Algorithm A.1 in Belloni et al. (2012).

Algorithm C.1. *Consider the model* $\mathrm{E}[y_t|\mathbf{x}_t] = \mathbf{x}_t'\boldsymbol{\beta}^0$ *for* $t = 1, \ldots, T$ *where* $\mathbf{x}_t$ *is a* p*-dimensional vector and* $\boldsymbol{\beta}^0$ *is the target value. The initial and refined penalty loadings are given by*

$$\textit{initial:} \quad \hat{\gamma}_j = \sqrt{\frac{1}{T}\sum_{t=1}^{T} x_{tj}^2 (y_t - \bar{y})^2} \qquad\qquad \textit{refined:} \quad \hat{\gamma}_j = \sqrt{\frac{1}{T}\sum_{t=1}^{T} x_{tj}^2 \hat{e}_t} \tag{C.47}$$

where $\bar{y} = T^{-1}\sum_t y_t$*. Specify the number of iterations* Q*. Proceed as follows: (1) Obtain the Lasso or Post-Lasso estimate* $\hat{\boldsymbol{\beta}}$ *using the initial penalty loadings and the optimal penalty level. (2) Obtain the corresponding residuals* $\hat{e}_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}$ *and update the Lasso or Post-Lasso estimate* $\hat{\boldsymbol{\beta}}$ *using the refined penalty loadings. (3) Repeat the second step* Q *times, or until convergence.*

C.3 A Note on the sparsity parameters s_1 and s_2

This brief note proofs the following proposition.

Proposition C.1. *Consider the model in* (3.24)-(3.25) *and Assumption 3.7 3.8. The first-step sparsity parameter,* s_1*, is at least as large as the second-step sparsity parameter* s_2*.*

Rewrite the spatial autoregressive panel model as

$$\mathbf{y}_t = \mathbf{W}\mathbf{y}_t + \mathbf{c}_t + \mathbf{e}_t \tag{C.48}$$

$$\mathbf{y}_t = (\mathbf{I}_n - \mathbf{W})^{-1}\mathbf{c}_t + (\mathbf{I}_n - \mathbf{W})^{-1}\mathbf{e}_t \tag{C.49}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{nt})'$, $\mathbf{c}_t = (\mathbf{x}_{1t}'\boldsymbol{\beta}_1, \ldots, \mathbf{x}_{nt}'\boldsymbol{\beta}_n)'$. Note that s_2 is the maximum number of non-zero row-elements of $\mathbf{W}$. Since $\boldsymbol{\beta}_i$ is non-zero for all i, s_1 is the maximum number of non-zero off-diagonal row-elements of $(\mathbf{I}_n - \mathbf{W})^{-1}$. By Neumann series,

$$(\mathbf{I}_n - \mathbf{W})^{-1} = \mathbf{I}_n + \mathbf{W} + \mathbf{W}^2 + \mathbf{W}^3 + \ldots \tag{C.50}$$

From the equation follows that the sparsity of $(\mathbf{I}_n - \mathbf{W})^{-1}$ must be at least as large as sparsity of $\mathbf{W}$, i.e., $s_1 \geq s_2$.

C.4 A note on the restricted eigenvalue condition

The argument is based on Bühlmann and Van de Geer (see 2011, Lemma 6.17 and Corollary 6.8). A similar argument can be made for the weighted restricted eigenvalue condition.

LEMMA C.1. *Suppose there exists a constant $\xi < 1$, that the restricted eigenvalue condition holds for $\mathbf{X}^\star$ and*

$$(1+C)^2 \frac{s}{T} \frac{\gamma_{max}}{\kappa_C(\mathbf{X}^\star)^2} \leq \xi \tag{C.51}$$

where $\|\hat{\mathbf{X}}'\hat{\mathbf{X}} - \mathbf{X}^{\star\prime}\mathbf{X}^\star\|_\infty \leq \gamma_{max} < \infty$ and $\|\cdot\|_\infty$ denotes the supremum absolute distance. Then, the restricted eigenvalue condition holds also for $C > 0$ and $\hat{\mathbf{X}}$ with

$$(1-\xi)\kappa_C(\mathbf{X}^\star) \leq \kappa_C(\hat{\mathbf{X}}) \leq (1+\xi)\kappa_C(\mathbf{X}^\star). \tag{C.52}$$

First, note $|\boldsymbol{\delta}'(\hat{\mathbf{X}}'\hat{\mathbf{X}} - \mathbf{X}^{\star\prime}\mathbf{X}^\star)\boldsymbol{\delta}| \leq \gamma_{\max}\|\boldsymbol{\delta}\|_1^2$. Secondly, since $\|\boldsymbol{\delta}_{\overline{\Omega}}\|_1 \leq C\|\boldsymbol{\delta}_\Omega\|$, using the restricted eigenvalue condition,

$$\|\boldsymbol{\delta}\|_1 \leq (1+C)\|\boldsymbol{\delta}_\Omega\|_1 \leq (1+C)\frac{\sqrt{s}}{\sqrt{T}} \frac{\|\mathbf{X}^\star\boldsymbol{\delta}\|_2}{\kappa_C(\mathbf{X}^\star)}. \tag{C.53}$$

Thus, $|\boldsymbol{\delta}'(\hat{\mathbf{X}}'\hat{\mathbf{X}} - \mathbf{X}^{\star\prime}\mathbf{X}^\star)\boldsymbol{\delta}| \leq \gamma_{\max}(1+C)^2 \frac{s}{T} \frac{\|\mathbf{X}^\star\boldsymbol{\delta}\|_2^2}{\kappa_C(\mathbf{X}^\star)^2}$. Substituting condition (C.51) gives $|\boldsymbol{\delta}'(\hat{\mathbf{X}}'\hat{\mathbf{X}} - \mathbf{X}^{\star\prime}\mathbf{X}^\star)\boldsymbol{\delta}| \leq \xi\|\mathbf{X}^\star\boldsymbol{\delta}\|_2^2$ which in turn yields the result.

Chapter 4

Spatial Analysis of the US Housing Market

> *Although we certainly cannot rule out home price declines, especially in some local markets, these declines, were they to occur, likely would not have substantial macroeconomic implications.*
>
> – Alan Greenspan, 9 June 2005[1]

The global financial crisis of 2007-2009 following the burst of the US subprime mortgage bubble has demonstrated the importance of housing markets for the health of the economy. The sheer size of the US housing market, estimated at $27 trillion[2] which even exceeds the US stock market, highlights the special role for the economy. Economic research has identified multiple channels through with the housing market is linked with the rest of the economy. For instance, Goodhart and Hofmann (2008) discuss the relationship with inflation, private credit, broad money and real Gross Domestic Product (GDP). Campbell and Cocco (2007) examine the effect of house price changes on consumption behaviour across age groups and attribute the connection to the wealth effect. The DSGE model in Iacoviello and Neri (2010) also suggests a link between the house market and consumption.

Given the adverse effects that sudden drops in house prices can have on the

[1] Testimony of Chairman Alan Greenspan to the Joint Economic Committee of the United States (US) Congress. Full statement is available at `http://www.federalreserve.gov/BOARDDOCS/TESTIMONY/2005/200506092/default.htm`, accessed on 31 August 2016.

[2] *The Economist*, 20 August 2016. Available at `http://www.economist.com/news/leaders/21705317-americas-housing-system-was-centre-last-crisis-it-has-still-not-been-properly`.

economy, there is a need to predict future developments and identify risks in the housing market. One of the central questions is if and when house prices show a 'bubble'-type behaviour[3], i.e., deviate from macroeconomic fundamentals such as income. For this reason, a number of empirical studies have examined the long-run relationship between income and house prices, and employed cointegration tests (e.g. Malpezzi, 1999; Meen, 2002; Gallin, 2006). The absence of a cointegration relationship between house prices and income can be interpreted as evidence that house prices deviate from the underlying economic fundamentals and, thus, as evidence for a bubble. On the other hand, the existence of a cointegration relationship suggests that house prices and income are bound to move together.

Since local housing markets are not isolated from each other, accounting for spatial effects is essential. It is well known that ignoring cross-section dependence may induce non-negligible estimation bias (LeSage and Pace, 2009; Chudik and Pesaran, 2013). From a theoretical point, there is reason to expect significant spatial effects due to spatial arbitrage and migration patterns (Meen, 1999). Indeed, the importance of the spatial dimension has been acknowledged in housing market research. For example, one strand of the empirical literature examines the 'ripple effect' for the United Kingdom (UK), which refers to the empirical phenomenon that house price shocks at one location diffuse over time to other regions (Drake, 1995; Meen, 1999).

There are two alternative approaches to cross-section dependence. First, the spatial econometric literature attempts to model dependence using a spatial weights matrix, which is a $n \times n$ matrix that determines the interaction structure between units. Numerous spatial models and estimation methods have been suggested in order to capture interactions between units (see, e.g., overview in Elhorst, 2010a). However, since the spatial dimension in most panel data sets is large relative to the time dimension, the spatial weights matrix is typically not identified due to the

[3]The definition of the term 'bubble' is disputed. For instance, Fama (2014) argues that the term is meaningless, since price declines that are associated with the burst of bubbles are not predictable *a priori*. Following Stiglitz (1990, p. 13) and Case and Shiller (2003, p. 299), I refer to a bubble loosely as a situation of temporarily elevated prices driven by price expectations as opposed to economic fundamentals.

large number of unknown parameters, i.e., $n(n-1)$. In practice, applied researchers often specify a spatial weights matrix on an *ad hoc* basis using observable distance measures, while theorists treat the spatial weights matrix as known (Kelejian and Prucha, 1998; Lee, 2004).

The second approach tends to view cross-section dependence as a nuisance. Cross-section dependence is viewed as arising from unobservable common factors in the error term—e.g. global economic shocks—that affect all units with heterogeneous coefficients. Pesaran (2006) proposes a method to control for cross-section dependence arising from unobservable factors. The Common Correlated Effects (CCE) estimator is least-squares applied to the regression model augmented by cross-sectional averages, which are shown to eliminate cross-section dependence asymptotically. The CCE is easy to implement and applicable in various settings, including non-stationary factors (Kapetanios, Pesaran, and Yamagata, 2011) and dynamic models (Chudik and Pesaran, 2015). Pesaran and Tosetti (2011) show that the CCE approach also deals with cross-section dependence arising from spatial processes. The seminal study by Holly, Pesaran, and Yamagata (2010, in the following HPY) unifies both perspectives on cross-section dependence. In an application to a US panel dataset of house prices, the authors first apply the CCE estimator to eliminate strong cross-section dependence and, in the second step, examine the de-factored residuals for spatial dependence.

In this chapter, I consider an alternative estimation method that allows to control for spatial effects in a panel model with spatially lagged exogenous regressors when the spatial weights matrix is unknown. Following LeSage and Pace (2009), I refer to the model as the SLX model. The SLX can be motivated as the reduced form of the spatial autoregressive or spatial lag model. The methodology relies on the Lasso introduced by Tibshirani (1996) and advanced by Bickel, Ritov, and Tsybakov (2009) and Belloni et al. (2012), among others. The Lasso is a well-established regularisation technique that penalises the absolute size of coefficients and can select relevant regressors from a large set of variables. Belloni, Chernozhukov, and Hansen (2014b) consider a regression model with low-dimensional regressors of primary interest and a large number of putative confounding factors, from which the Lasso selects appropriate control variables. The proposed Lasso-based estimator, referred to as Post-double-selection Estimator (PDSE), is consistent and asymptotically normal for the low-dimensional parameters, even if the total number of regressors exceeds the sample size. The identifying assumption is approximate sparsity, which requires that only a small set of variables is needed to account for confounding effects.

The application in this chapter examines the relationship between house prices and real income for a panel data set of US state-level house prices, covering the period 1975 to 2014. It is demonstrated that the PDSE can be employed to control for spatial effects in a large T panel model. The approach treats the spatial weights as high-dimensional parameters and does not rely on pre-specified spatial weights, thereby avoiding misspecification issues.

Estimation results establish a cointegration relationship between real house prices and real per capita income, providing evidence that house prices are largely driven by economic fundamentals. The PDSE is then applied to a spatial Error-correction Model (ECM). Remarkably, it is shown that the estimated spatial weights are associated with observable variables, including population size and geographic distance. The results also indicate that standard specifications of the spatial weights matrix, such as the binary contiguity matrix, would fail to capture the full complexity of spatial interactions.

The remainder of this chapter is organised as follows. The next section offers an overview of the empirical literature on house prices and their economic determinants, with a special focus on the issue of spatial dependence. Section 4.2 presents the Lasso-based estimation method. The application to the US housing market is presented in Section 4.3. Section 4.4 provides some concluding remarks.

4.1 Empirical literature on housing markets

The importance of spatial effects has been acknowledged in the house price literature since the early 1990s. Notably, a number of studies have considered the 'ripple effect' for the UK, which describes the transmission of price shocks emanating in the South-East and London to the rest of the UK (early works include Giussani and Hadjimatheou, 1991; Drake, 1995). Meen (1996) examines the issues of spatial heterogeneity, spatial dependence and convergence for a panel of UK regions, and find evidence for a cointegration relationships between each region and the South-East. The adjustment coefficients are shown to decrease with distance to the South-East, thus providing support for the ripple effect hypothesis. In contrast, Ashworth and Parker (1997) find no evidence for spatial correlation when testing for spatial dependence in the residuals from ECM estimations with UK regions. However, the contiguity matrix employed might fail to capture all spatial effects.

Meen (1996) also contributes to the literature on unknown spatial weights by estimating the interaction effects in a spatial error model. However, the OLS-based estimation method does not account for the endogeneity inherent in spatial interactions and is therefore inconsistent. Furthermore, the method is only feasible if the time dimension is sufficiently large (Bhattacharjee and Jensen-Butler, 2013). Despite this limitation, the method of Meen (1996) provides valuable insights, as it reveals complex interaction effects that would not have been captured by standard *ad hoc* specifications of the spatial weights matrix. Meen (2002) offers an empirical comparison of the US and UK housing market.[4] While the US market does not exhibit a 'ripple effect' behaviour, Meen (2002) argues that both markets show similar short and long-run dynamics as indicated by cointegration analysis. One central difference, however, is that the price elasticity of supply is significantly smaller in the UK market.

[4]While this overview focuses on the UK and especially the US, a noteworthy study for the Netherlands is given by Dijk et al. (2011), who focus on the topic of spatial heterogeneity in the Dutch housing market. They form two clusters of Dutch regions for which common slope coefficients are assumed, but allow for parameter heterogeneity across the clusters. Results indicate that house prices in both cluster are cointegrated with income, which also implies a cointegration relationship between the two clusters.

Malpezzi (1999) examines a panel data set of US Metropolitan Statistical Areas (MSA) over 18 sample years and identify a cointegration relationship between house prices and income significant at the 1% level. In contrast, using national-level and city-level US data, Gallin (2006) finds no indication that house prices are cointegrated with market fundamentals, but as pointed out by HPY the bootstap cointegration test employed by the author may lead to misleading results given the relatively small time dimension of 23 years. Brady (2011, 2014) estimates spatial autoregressive panel models for California counties and US States. The author obtains spatial impulse response functions using the local projection method introduced in Jorda (2005). Employing first and second-order contiguity weights matrices, estimation results suggest that local shocks have statistically significant and lasting effects on other areas.

One strand of the literature employs Vector Autoregressive (VAR) analysis. In a VAR model, each variable depends on its own past and previous realisations of other variables included in the system. For instance, Pollakowski and Ray (1997) estimate a VAR of house prices for the region of New York, New Jersey and Long Island. They conclude that house prices observed in one area determines house prices in other areas. Using US state-level data, Kuethe and Pede (2011) estimate the spatial VAR model proposed in Beenstock and Felsenstein (2007), which augments a classical VAR by spatially lagged variables. The authors consider a first and higher-order neighbour matrix and find spatial effects to exhibit relevant forecasting power using Granger causality tests. However, as in most spatial econometric applications, the spatial VAR approach assumes a known spatial weights matrix.

HPY and Holly, Pesaran, and Yamagata (2011) unify the spatial econometric and the common factor approach in two applications to the US and UK market. HPY first employ the CCE estimator to extract the effect of unobserved common factors and, in the second step, analyse the residuals for spatial dependence using a spatial contiguity matrix. The approach is based on Pesaran and Tosetti (2011) who demonstrate that the CCE estimator is consistent and asymptotically normal in the presence of weak and strong forms of cross-section dependence, including dependence arising from spatial processes. HPY find a cointegration relationship between real house prices and real per capita income. Baltagi and Li (2014) reproduce the estimation methodology in HPY using MSA data and confirm the existence of a cointegration relationship. Holly, Pesaran, and Yamagata (2011) demonstrate that, in contrast to the US market, the UK is characterised by one dominant region, London. Specifically, the authors show that other UK regions cointegrate with London house prices.

Bailey, Holly, and Pesaran (2016) develop an estimation method that allows for unknown spatial weights in a spatial error panel model. In the first step, the authors propose to apply, as in HPY, the CCE estimator to extract unobserved common factors. The de-factored residuals are examined for spatial dependence in the second step. The multiple testing procedure by Holm (1979) is employed to identify non-zero pairwise correlations between units, which in turn allows to construct a spatial weights matrix with entries -1 (significant negative correlation), +1 (significant positive correlation) and 0 (insignificant correlation). A central finding of the analysis is the existence of negative spatial effects, which are disregarded in many spatial econometric applications employing pre-specified weights.

The above overview suggest that there is overall strong evidence that spatial dependence and spatial heterogeneity are essential in the analysis of housing markets. However, spatial econometric analysis suffers from the limitations of pre-specified spatial weights. In particular, the specification of the weights matrix is often arbitrary and distorts estimation results. This highlights the importance of the work in Bailey, Holly, and Pesaran (2016). A potential limitation of the multiple testing approach is the underlying assumption of symmetry. The correlation-based method provides no insights into the direction of spatial effects. Thus, there is, complementary to the work of Bailey, Holly, and Pesaran (2016), a need for alternative estimation methods allowing for unknown weights. This article argues that the Lasso may provide the basis for such a method.

4.2 Estimation method

At first, I introduce the feasible Lasso and feasible Post-Lasso, which Belloni et al. (2012) propose in the context of Instrumental Variables (IV) regression and the issue of many instruments. The feasible Lasso and Post-Lasso form the basis for the PDSE due to Belloni, Chernozhukov, and Hansen (2014b), to be discussed in Section 4.2.2. The framework for applying the PDSE to the SLX panel model is presented in Section 4.2.3.

4.2.1 The feasible Lasso and Post-Lasso

Suppose the interest lies in predicting an outcome based on a set of predictors. Specifically, consider the general model

$$y_i = f(\mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{4.1}$$

where the dependent variable, y_i, is affected by a vector of observable variables, $\mathbf{z}_i$, through the unknown and potentially non-linear function $f(\cdot)$. The error ε_i is independently, but not necessarily identically distributed. A common approach is to approximate the function $f(\cdot)$ by linearisation (e.g. Bickel, Ritov, and Tsybakov, 2009),

$$y_i = \underbrace{\mathbf{x}_i'\boldsymbol{\delta}_0 + r_i}_{f(\mathbf{z}_i)} + \varepsilon_i, \tag{4.2}$$

where $r_i := f(\mathbf{z}_i) - \mathbf{x}_i'\boldsymbol{\delta}_0$ represents the error from approximation. The p-dimensional vector $\mathbf{x}$ may be set to $\mathbf{z}_i$ if linearity is assumed or may consists of transformations of $\mathbf{z}_i$ such as polynomials, interaction terms and dummy variables.

In most econometric applications, there is substantial uncertainty about the correct model specification and about which variables are to be included in the regression. To account for a large number of competing model specifications, p is allowed to be large and may even exceed the sample size. Belloni and Chernozhukov (2011) and Belloni, Chernozhukov, and Hansen (2014a) discuss prominent situations in economics and argue that the setting where p is large is common in applied econometrics, although usually not explicitly acknowledged. Typical examples include applications where there is a large number of control variables available (e.g. cross-country growth regressions) or where there exists a large number of instruments (e.g. wage regressions based on Angrist and Krueger, 1991). In these cases, researchers tend to consider only a small number of alternative specifications due to practical constraints and select variables on an *ad hoc* basis guided by economic considerations, implying a considerable risk of misspecification. Naturally, the composition of $\mathbf{x}_i$ determines the quality of the approximation and, hence, also leaves room for influencing estimation results. However, the large p setting is an improvement in that it can accommodate many alternative model specifications at the same time.

The model in (4.2) is not identified and estimation by OLS is not feasible if p is larger than n, unless further assumptions are made. Sparsity is an an attractive assumption due to its intuitive appeal and as it facilitates model interpretation (Hastie, Tibshirani, and Wainwright, 2015). The assumption of *exact* sparsity assumes $f(\mathbf{z}_i) = \mathbf{x}_i'\boldsymbol{\delta}_0$ and that the parameter vector $\boldsymbol{\delta}_0$ is composed of many zeros such that $\|\boldsymbol{\delta}_0\|_0 \ll n$, where $\|\cdot\|_0$ denotes the ℓ_0-norm. Thus, exact sparsity ensures that the true model is identified, while treating the support of $\boldsymbol{\delta}_0$ as unknown. However, the condition seems too restrictive if $\boldsymbol{\delta}_0$ has many non-zero elements that are negligible in size. A more suitable assumption is *approximate* sparsity, which relies on the existence of a sparse target vector $\boldsymbol{\delta}_0$, but allows for a non-zero approximation error r_i. The resulting approximation error is required to be sufficiently small. Formally, the condition of approximate sparsity assumes that there is a parameter vector $\boldsymbol{\delta}_0$ such that

$$s := \|\boldsymbol{\delta}_0\|_0 \ll n \quad \text{and} \quad \left(\frac{1}{n}\sum_{i=1}^{n} r_i^2\right)^{1/2} \lesssim \sqrt{\frac{s}{n}}, \tag{4.3}$$

where $\sqrt{s/n}$ can be motivated as the convergence rate if the true model is known (Belloni et al., 2012).

The Lasso is a popular regularisation method that allows for estimation in high-dimensional models under the sparsity assumption. The ℓ_1-penalisation employed by the Lasso is, in contrast to ℓ_0-penalisation, computational attractive due to its convex form, while yielding sparse solutions. The feasible Lasso estimator is

$$\hat{\boldsymbol{\delta}} = \arg\min \ \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\delta})^2 + \frac{\lambda}{n}\|\hat{\boldsymbol{\Upsilon}}^0\boldsymbol{\delta}\|_1. \tag{4.4}$$

The optimal penalty level proposed in Belloni et al. (2012) is $\lambda = 2c\sqrt{n}\Phi^{-1}(1 - \gamma/(2p))$ where $c > 1$ and γ meets $\log(1/\gamma) \lesssim \log(\max(p, n))$ as $n \to \infty$.[5] As in Bickel, Ritov, and Tsybakov (2009), the penalty level is chosen to dominate the error variability in order to avoid false inclusion of irrelevant variables. The estimated penalty loadings $\hat{\boldsymbol{\Upsilon}}^0 = \mathrm{diag}(\hat{l}_1, \ldots, \hat{l}_p)$ account for heteroskedastic errors and are obtained using the iterative algorithm in Appendix C.2, which ensures asymptotically ideal penalty loadings (Belloni et al., 2012). Under the assumption of homoskedasticity, $\hat{\boldsymbol{\Upsilon}}^0$ can be set to the identity matrix. Since the second term, $\lambda/n\|\hat{\boldsymbol{\Upsilon}}^0\boldsymbol{\beta}\|_1$, imposes a penalty on the absolute size of coefficient estimates, some estimates are set to exactly zero. The Lasso can thus operate as a variable selection method.

The ℓ_1-penalisation imposed by the Lasso introduces a shrinkage bias. To moderate the bias, the feasible Post-Lasso OLS estimator applies least squares to the model selected by the feasible Lasso, i.e.,

$$\min \ \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\delta})^2 \qquad \text{s.t.} \quad \delta_j = 0, \forall j \notin \hat{T}, \tag{4.5}$$

where $\hat{T} = \{j : \hat{\delta}_j \neq 0\}$ is the support selected by the Lasso. The Monte Carlo simulations in Belloni and Chernozhukov (2011) confirm that the Post-Lasso performs better than the Lasso in terms of finite sample bias.

Related to sparsity, the Lasso and Post-Lasso also rely on a central assumption with regard to the Gram matrix $\sum_i \mathbf{x}_i\mathbf{x}_i'/n$ (Belloni et al., 2012). While the OLS estimator requires the Gram matrix to be of full rank, this assumption is not appli-

[5] Belloni, Chernozhukov, and Hansen (2014b, fn. 10) recommend $c = 1.1$ and $\gamma = \min(1/n, 0.05)$.

cable in the high-dimensional setting where $p > n$ and the Gram matrix is singular by construction. The restricted sparse eigenvalue condition replaces the full rank condition. For the condition to be satisfied, the minimal and maximal χ-sparse eigenvalues, defined by

$$\phi_{\min(\chi)}(\mathbf{A}) := \min_{1\leq\|\boldsymbol{\theta}\|_0\leq\chi} \frac{\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|_2^2} \quad \text{and} \quad \phi_{\max(\chi)}(\mathbf{A}) := \max_{1\leq\|\boldsymbol{\theta}\|_0\leq\chi} \frac{\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|_2^2}, \tag{4.6}$$

should be bounded away from zero and bounded from above for the Gram matrix $\sum_i \mathbf{x}_i\mathbf{x}_i'/n$. The restricted sparse eigenvalue condition guarantees that any submatrix of appropriate size is positive definite and thus invertible.

4.2.2 Post-double-selection estimator

Belloni, Chernozhukov, and Hansen (2014b) consider the model

$$y_i = d_i\alpha_0 + g(\mathbf{z}_i) + \zeta_i, \qquad \mathrm{E}[\zeta_i|\mathbf{z}_i, d_i] = 0, \tag{4.7}$$

$$d_i = m(\mathbf{z}_i) + v_i, \qquad \mathrm{E}[v_i|\mathbf{z}_i] = 0, \tag{4.8}$$

where $(y_i, d_i, \mathbf{z}_i')$ are independent across i. The interest lies in the parameter α_0, which measures the effect of the exogenous regressor d_i on the dependent variable y_i. Following Belloni, Chernozhukov, and Hansen (2014b), I discuss the case where d_i is a scalar, although d_i can be generalised to be a vector of finite and fixed dimension (Belloni, Chernozhukov, and Hansen, 2014b, fn. 8). The vector $\mathbf{z}_i$ denotes confounding factors, which affect the response variable y_i and the regressor d_i through the unknown functions $g(\cdot)$ and $m(\cdot)$, respectively. Since both y_i and d_i depend on $\mathbf{z}_i$, an omitted variables bias arises unless the confounding factors are accounted for.

As in the previous section, the idea is to approximate the unknown functions $g(\cdot)$ and $m(\cdot)$ using the high-dimensional vector $\mathbf{x}_i$. Hence,

$$y_i = d_i\alpha_0 + \underbrace{\mathbf{x}_i'\boldsymbol{\beta}_{g,0} + r_{g,i}}_{g(\mathbf{z}_i)} + \zeta_i, \tag{4.9}$$

$$d_i = \underbrace{\mathbf{x}_i'\boldsymbol{\beta}_{m,0} + r_{m,i}}_{m(\mathbf{z}_i)} + v_i, \tag{4.10}$$

where $r_{g,i}$ and $r_{m,i}$ are the approximations errors. Note that under the assumption of linearity, $\mathbf{z}_i$ can be set to $\mathbf{x}_i$.

Identifying the appropriate set of controls is challenging if the size of $\mathbf{x}_i$, p, is large relative to the number of observations, n. In particular, the model is not identified without further assumptions if $1+p > n$. In this context, the assumption of approximate sparsity states that the parameter vectors $\boldsymbol{\beta}_{m,0}$ and $\boldsymbol{\beta}_{g,0}$ exist such that

$$g(\mathbf{z}_i) = \mathbf{x}_i'\boldsymbol{\beta}_{g,0} + r_{g,i}, \quad 1 \leq \|\boldsymbol{\beta}_{g,0}\|_0 \leq s \ll n, \quad \sqrt{\sum_{i=1}^{n} r_{g,i}^2} \lesssim_{\mathrm{P}} \sqrt{s/n}, \tag{4.11}$$

$$m(\mathbf{z}_i) = \mathbf{x}_i'\boldsymbol{\beta}_{m,0} + r_{m,i}, \quad 1 \leq \|\boldsymbol{\beta}_{m,0}\|_0 \leq s \ll n, \quad \sqrt{\sum_{i=1}^{n} r_{m,i}^2} \lesssim_{\mathrm{P}} \sqrt{s/n}, \tag{4.12}$$

and the sparsity index s should obey $s^2 \log^2(\max(n,p))/n \to 0$.

The reduced form equations corresponding to (4.9)-(4.10) are given by

$$y_i = \mathbf{x}_i'\bar{\boldsymbol{\beta}}_0 + \bar{r}_i + \bar{\zeta}_i \tag{4.13}$$

$$d_i = \mathbf{x}_i'\boldsymbol{\beta}_{m,0} + r_{m,i} + v_i \tag{4.14}$$

with $\bar{\boldsymbol{\beta}}_0 = \alpha_0\boldsymbol{\beta}_{m,0} + \boldsymbol{\beta}_{g,0}$, $\bar{r}_i = \alpha_0 r_{m,i} + r_{g,i}$ and $\bar{\zeta}_i = \alpha_0 v_i + \zeta_i$. The post-double-selection estimation method applies the feasible Lasso to both reduced form equations and identifies the appropriate set of controls as the union of variables selected. A *naive* single-selection approach based on (4.13) only is likely to ignore controls that have a substantial impact on d_i, but only a small effect on y_i, thus inducing an omitted variable bias (Belloni, Chernozhukov, and Hansen, 2014a; Belloni, Chernozhukov, and Hansen, 2014b). The PDSE proceeds in the following three steps.

STEP 1: The feasible Lasso estimator is applied to equation (4.14). The set of controls selected by the first step Lasso is $\hat{I}_1 = \text{support}(\hat{\boldsymbol{\beta}}_{m,0})$.

STEP 2: The feasible Lasso is applied to (4.13). The set of selected controls in the second step is denoted by $\hat{I}_2$.

STEP 3: The PDSE is the least squares estimator of regressing y_i against d_i and the union of control variables $\hat{I}_1$ and $\hat{I}_2$ as well as, optionally, additional control variables selected by the researcher, denoted by $\hat{I}_3$.

Thus, the PDSE is defined as

$$(\tilde{\alpha}, \tilde{\boldsymbol{\beta}}) = \arg\min \sum_{i=1}^{n} (y_i - d_i\alpha - \mathbf{x}_i'\boldsymbol{\beta})^2 \quad \text{s.t.} \quad \beta_j = 0, \forall j \notin \hat{I}, \tag{4.15}$$

where $\hat{I}$ is the union of $\hat{I}_1$, $\hat{I}_2$ and $\hat{I}_3$. The amelioration set $\hat{I}_3$ may be used by the analyst to improve the fit, but its cardinality should not be considerably larger than the size of $\hat{I}_1$ and $\hat{I}_2$.

Belloni, Chernozhukov, and Hansen (2014b, Theorem 1) show that the PDSE of α_0 is consistent and asymptotically normal. The central assumptions for the results are approximate sparsity as stated in (4.11)-(4.12), the restricted sparse eigenvalue condition for the Gram matrix $\sum_i \mathbf{x}_i\mathbf{x}_i'/n$ and further regularity conditions concerning the moments of the structural variables and errors, which allow to utilise the moderate deviation theory of Jing, Shao, and Wang (2003).[6]

Model (4.7)-(4.8) represents a common econometric setting, in which the analyst is faced with a large number of controls and the choice of variables affects the coefficient estimate of α_0. For instance, Belloni, Chernozhukov, and Hansen (2014b) revisit the influential crime study of Donohue and Levitt (2001), which establishes a causal relationship between abortion and crime rates. In this example, the parameter of interest, α_0, corresponds to the effect of abortion on a specific type of crime observed around two decades later. Belloni, Chernozhukov, and Hansen (2014a) find that the formal Lasso-based method for selecting state-level control variables alters the result relative to the standard approach of selecting a small set of controls

[6]The moderate deviation theory for self-normalised sums is used to derive the optimal penalty level under non-Gaussian and heteroskedastic errors. See conditions ATSE, SE and SM in Belloni, Chernozhukov, and Hansen (2014b) for further details.

on theoretical grounds. Specifically, the authors show that the causal relationship between abortion rates and crime disappears when the PDSE is employed.

4.2.3 Post-double-selection and the SLX panel model

This section discusses the application of the PDSE to the spatial panel model

$$y_{it} = x_{it}\beta_{0,i} + g_i(\tilde{\mathbf{x}}_{-i,t}) + \varepsilon_{it}, \tag{4.16}$$

$$x_{it} = m_i(\tilde{\mathbf{x}}_{-i,t}) + u_{it}, \qquad i = 1, \ldots, n, \ t = 1, \ldots T, \tag{4.17}$$

in which ε_{it} and u_{it} are independently distributed over i and t. The main focus of the analysis is on conducting inference on $\beta_{0,i}$ and its mean, while controlling for a large number of confounding spatial effects. The vector $\tilde{\mathbf{x}}'_{-i,t} = (x_{1t}, \ldots, x_{i-1,t}, x_{i+1,t}, \ldots, x_{nt})$ includes contemporaneous explanatory variables observed at locations other than i. Thus, the model allows the outcome at location i to be affected not only by the local regressor x_{it}, but also by exogenous regressors observed at other locations through the unknown function $g(\cdot)$. While the above model only allows for one local, low-dimensional regressors, the framework can be extended to include a finite and fixed number of local regressors. Also note that setting $g(\tilde{\mathbf{x}}_{-i,t}) = \sum_{i \neq j} w_{ij} x_{jt}$ yields the panel version of the SLX model in LeSage and Pace (2009).

In the context of the spatial panel model in (4.16)-(4.17), the identifying assumption of approximate sparsity requires the existence of spatial weights w^0_{ij} and m^0_{ij} such that for $i = 1, \ldots, n$

$$g_i(\tilde{\mathbf{x}}_{-i,t}) = \sum_{j \neq i} x_{jt} w^0_{ij} + r_{y,i}, \quad \sum_{j \neq i} \mathbb{1}\{w^0_{ij} \neq 0\} \leq s, \quad \sqrt{\sum_{t=1}^{T} r^2_{y,i}} \lesssim_{\mathrm{P}} \sqrt{s/T}, \tag{4.18}$$

$$m_i(\tilde{\mathbf{x}}_{-i,t}) = \sum_{j \neq i} x_{jt} m^0_{ij} + r_{x,i}, \quad \sum_{j \neq i} \mathbb{1}\{m^0_{ij} \neq 0\} \leq s, \quad \sqrt{\sum_{t=1}^{T} r^2_{x,i}} \lesssim_{\mathrm{P}} \sqrt{s/T}, \tag{4.19}$$

$$\text{with} \quad s^2 \log^2(\max(T, p))/T \to 0. \tag{4.20}$$

The second central assumption is that the restricted sparse eigenvalue condition holds for the Gram matrices $\sum_t \tilde{\mathbf{x}}_{-i,t}\tilde{\mathbf{x}}'_{-i,t}/T$ with $i = 1, \dots, n$.

The assumption of sparsity is especially plausible in the spatial context. In the spatial econometric literature, the row and column sums of the spatial weights matrix are assumed to be uniformly bounded in absolute value (Kelejian and Prucha, 1998; Kapoor, Kelejian, and Prucha, 2007), which guarantees weak cross-section dependence. The sparsity assumption is closely related to uniform boundedness in that it limits the number of connections for each row. Specifically, assuming that each spatial weight is bounded in absolute value ensures in conjunction with sparsity uniform boundedness of the absolute row and column sums.

The PDSE is implemented for each i separately.

STEP 1: The Lasso estimator regresses y_{it} against $\tilde{\mathbf{x}}_{-i,t}$. The set of controls selected in this first step is denoted by $\hat{I}_1$.

STEP 2: The Lasso is applied to regress x_{it} against $\tilde{\mathbf{x}}_{-i,t}$. The set of selected controls is denoted by $\hat{I}_2$.

STEP 3: The PDSE of $\beta_{0,i}$ is the least squares estimator of regressing y_{it} against x_{it} as well as the union of spatial lags in $\hat{I}_1$, $\hat{I}_2$ and, optionally, further controls selected by the analyst, i.e., $\hat{I}_3$.

Following the Common Correlated Effects Mean Group (CCEMG) estimator, the PDSE mean group estimator of β_0, which denotes the mean of $\beta_{0,i}$, is obtained as the average of the individual PDSE estimates $\hat{\beta}_{i,\text{PDSE}}$ and its standard error is obtained as $1/(n-1)\sum_i \hat{\beta}_{i,\text{PDSE}}$.

4.3 Application

The econometric analysis in this study uses a quarterly panel data set covering US States over the period 1975-2014. Table 4.1 provides an overview of the variables and their respective sources, frequencies and time frames. Table 4.2 reports summary statistics.

Variable	Time period	Sources / Notes
House prices	1975M1-2014M12	FreddieMac House Price Index. Normalised such that 1980Q1=1. Downloaded from `http://www.freddiemac.com/finance/fmhpi/archive.html`.
Personal income	1948Q1-2014Q3	US Bureau of Economic Analysis. Downloaded from `http://www.bea.gov/regional/`
Population	1969-2013	US Bureau of Economic Analysis. Retrieved from `http://www.bea.gov/itable/` and temporally disaggregated using total labour force from the Bureau of Labor Statistics as a predictor variable (see method due to Chow and Lin, 1971).
Consumer price index	1975M1-2014M12	US Bureau of Labor Statistics. CPI is not available at the state-level. Data from metropolitan areas and US regions is matched with states.
Unemployment rate	1976M1-2014M12	US Bureau of Labor Statistics, Local Area Unemployment Statistics. Averaged to obtain quarterly data.
Effective interest rates	1978-2012	Federal Housing Finance Agency. Annual values approximate Quarter 1 to 4. Retrieved from `http://www.fhfa.gov/DataTools/Downloads/Pages/Monthly-Interest-Rate-Data.aspx`.

Notes: The data was retrieved in February to March, 2015.

Table 4.1: *Data overview and sources.*

4.3.1 Data overview

The house price measure is the monthly, state-level FreddieMac House Price Index (FMHPI), which is constructed based on observations of repeated transactions for given properties (i.e., purchases or appraisal valuations for refinancing purposes). The series was aggregated to quarterly frequency by taking averages. Quarterly income and annual population data are downloaded from the US Bureau of Economic Analysis. The population data is converted from annual to quarterly frequency using the temporal disaggregation method due to Chow and Lin (1971) and utilising quarterly total labour force as a predictor variable. The method ensures that the predicted high-frequency series is consistent with the original series in terms of annual averages.[7] Consumer price indices and unemployment rates are from the US Bureau of Labor Statistics. Since inflation data is not available by state, consumer price indices of MSAs and US regions are matched with states to construct state-level proxy series of inflation as in HPY. Lastly, the effective interest rate is from the Federal Housing Finance Agency (FHFA) and annual values are used to approximate quarterly observations.

[7] See also the R package `tempdisagg` (Sax and Steiner, 2013).

	Description	*Obs.*	*Mean*	*St. Dev.*	*Min*	*Max*
y_{it}	Natural logarithm of real per capita income	7448	2.46	0.22	1.89	3.29
p_{it}	Natural logarithm of real house prices	7448	0.03	0.24	-0.64	0.97
η_{it}	Population growth, $\eta_{it} = \Delta \log(\text{Population}_{it})$	7399	0.00	0.01	-0.05	0.04
r_{it}	Real effective interest rate	6860	7.56	2.59	0.70	17.89

Table 4.2: *Summary statistics and notation.*

	New England	Mid-East	South-East	Great lakes	Plains	South-West	Rocky mountain	Far West
First difference of log of real house prices								
New England region	0.46	-	-	-	-	-	-	-
Mid-East	0.38	0.531	-	-	-	-	-	-
South-East	0.27	0.315	0.403	-	-	-	-	-
Great lakes	0.29	0.323	0.375	0.625	-	-	-	-
Plains region	0.17	0.232	0.269	0.351	0.364	-	-	-
South-West	0.18	0.259	0.347	0.329	0.271	0.586	-	-
Rocky mountain region	0.12	0.163	0.311	0.343	0.264	0.396	0.533	-
Far West	0.29	0.332	0.387	0.417	0.281	0.377	0.398	0.628
First difference of log of real per capital income								
New England region	0.74	-	-	-	-	-	-	-
Mid-East	0.53	0.609	-	-	-	-	-	-
South-East	0.48	0.523	0.701	-	-	-	-	-
Great lakes	0.50	0.507	0.601	0.783	-	-	-	-
Plains region	0.37	0.44	0.565	0.579	0.71	-	-	-
South-West	0.41	0.478	0.623	0.543	0.545	0.75	-	-
Rocky mountain region	0.37	0.486	0.603	0.501	0.548	0.653	0.737	-
Far West	0.48	0.526	0.623	0.574	0.521	0.614	0.635	0.709

Table 4.3: *Average correlation coefficients within and between US regions. The reported numbers are the average sample pair-wise correlation coefficients. Region definitions are listed in Table D.1. See Holly, Pesaran, and Yamagata (2010, Fig. 3-4) for comparison.*

4.3.2 Exploring spatial dependence

Table 4.3 shows the average pairwise correlation coefficients within and between regions for the log-difference of real house prices and real per capita income. It is expected that spatial dependence is greater within than across regions, as nearby states tend to share more common characteristics and typically exhibit a similar economic development. Indeed, the diagonal elements, which correspond to within-region dependence, are at least as large as the off-diagonal elements. Overall, spatial dependence appears strong, even between regions.

4.3.3 Panel unit root testing

In order to establish a cointegration relationship between house prices and income, the first step is to identify the order of integration for both variables. The so-called first generation of panel unit root tests, which includes Levin, Lin, and Chu (2002) and Im, Pesaran, and Shin (2003, IPS) among others, rely on the assumption of

Lags	*1*	*2*	*3*	*4*
	Intercept and trend		$a_{i0} \neq 0$, $a_{i1} \neq 0$	
p_{it}	-1.64	-1.65	-1.87	-2.18
y_{it}	-1.91	-1.89	-1.87	-1.99
η_{it}	-7.11***	-6.39***	-3.81***	-3.86***
r_{it}	-6.14***	-5.47***	-4.23***	-4.56***
	Intercept	$a_{i0} \neq 0$, $a_{i1} = 0$		
Δp_{it}	-8.44***	-6.18***	-4.73***	-3.87***
Δy_{it}	-8.88***	-7.13***	-5.63***	-5.30***
$\hat{u}_{it}$	-2.40***	-2.20**	-2.32***	-2.54***

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Obtained using the command `cipstest` from the `plm` package (Croissant and Millo, 2008). Critical values are given by -2.71 (1%), -2.61 (5%) and -2.55 (10%) if both intercept and trend are included. If only an intercept is included, critical values are -2.23 (1%), -2.12 (5%) and -2.05 (10%), respectively. All critical values correspond to $n = 50$ and $T = 100$ (Pesaran, 2007, Table II).

Table 4.4: *Results of the CIPS panel unit root test due to Pesaran (2007).*

independence across units.[8]

The IPS test considers the classical univariate Augmented Dickey-Fuller (ADF) regression equation (Dickey and Fuller, 1979), which is given by

$$\Delta\omega_{it} = \rho_i \omega_{i,t-1} + a_{i0} + a_{i1}t + \sum_{j=1}^{q} b_{ij}\Delta\omega_{i,t-j} + e_{it}, \quad i = 1, \ldots, n. \tag{4.21}$$

The augmented version of the Dickey-Fuller test extends the equation by temporal lags of $\Delta\omega_{it}$ to capture serial dependence. The IPS test statistic is defined as the average of the ADF t-statistics, which in turn are obtained from applying OLS to each i separately. A major advantage of treating the n equations separately is that the approach allows, in contrast to Levin, Lin, and Chu (2002), for heterogeneous ρ_i coefficients.

Under the null hypothesis of the IPS test, all n series in the panel contain a unit root, i.e.,

$$H_0\colon \rho_i = 0 \quad \text{for} \quad i = 1, \ldots, n, \tag{4.22}$$

while the alternative allows a fixed fraction $\delta = n_1/n$ to be stationary,

$$H_a\colon \rho_i < 0 \quad \text{for} \quad i = n_1 + 1, \ldots, n. \tag{4.23}$$

[8]For a survey of panel unit root tests, see Breitung and Pesaran (2008).

Pesaran (2007) introduces the Cross-sectionally Augmented IPS (CIPS) test that allows to test for unit roots in a panel in the presence of parameter heterogeneity as well as cross-section dependence. Similar to Im, Pesaran, and Shin (2003), the CIPS test builds on the univariate ADF test, but extends the individual ADF regressions by cross-sectional averages in order to capture contemporaneous dependence. Thus, the individual Cross-sectionally Augmented ADF (CADF) regression equations are given by

$$\Delta\omega_{it} = \rho_i\omega_{i,t-1} + a_{i0} + a_{i1}t + \sum_{j=1}^{q} b_{ij}\Delta\omega_{i,t-j} + c_i\bar{\omega}_{t-1} + \sum_{j=0}^{q} d_{ij}\Delta\bar{\omega}_{t-j} + e_{it}, \qquad (4.24)$$

where $\bar{\omega}_{t-1} = n^{-1}\sum_k \omega_{k,t-1}$ and $\Delta\bar{\omega}_t = n^{-1}\sum_k \Delta\omega_{kt}$. The CIPS test statistic is calculated as the average t-statistics from the CADF regressions and can be compared to the non-standard critical values provided in Pesaran (2007). As in the IPS test, all individual time-series are non-stationary under the null hypothesis, while the alternative allows a fraction of time series to be stationary.

Table 4.4 reports results for the CIPS applied to real house prices (p_{it}) and real per capita income (y_{it}) as well as their first differences. The tests fail to reject the null hypothesis of a unit root for both series in levels, while the CIPS tests are rejected for the first-differenced series at all considered significance levels, suggesting that both house prices and income per capita are integrated of order one. The CIPS test also provides strong evidence that population growth (η_{it}) and real interest rates (r_{it}) are stationary.

4.3.4 Long-run relationship

The long-run relationship between real house prices and real per capita income is

$$p_{it} = \alpha_i + \beta_i y_{it} + u_{it} \qquad (4.25)$$

$$u_{it} = \sum_{k=1}^{m} \gamma_{ik} f_{kt} + \epsilon_{it}. \qquad (4.26)$$

As in HPY, the error term u_{it} may exhibit strong cross-section dependence arising from a common factor model.

Similar to omitted variables, unobserved common factors may be correlated with the regressor of interest (here, y_{it}) and induce a bias if not accounted for. For example, an unobserved factor may represent economic shocks affecting all states with heterogeneous intensity. The CCE estimator controls for cross-section dependence in the error term and is consistent under general conditions. Kapetanios, Pesaran, and Yamagata (2011) show that the CCE estimator is also applicable if some of the unobserved factors are non-stationary. Pesaran and Tosetti (2011) demonstrate consistency and asymptotic normality in the presence of both a multifactor error structure and dependence arising from spatial processes. The estimator is implemented by augmenting the long-run equation with cross-section averages, i.e.,

$$p_{it} = \alpha_i + \beta_i y_{it} + \gamma_{p,i}\bar{p}_t + \gamma_{y,i}\bar{y}_t + e_{it}. \tag{4.27}$$

The underlying rationale is that the cross-section averages eliminate the influence of unobserved factors asymptotically and, thus, the CCE estimator resembles the CADF regressions employed for robust unit root testing.

There are two versions of the CCE estimator. The Common Correlated Effects Pooled (CCEP) estimator restricts the β_i parameters to be constant across i, while the CCEMG estimator allows for slope heterogeneity. The CCEMG estimate of β—i.e., the expected value of β_i—is the average of the OLS estimates obtained from individual regressions applied to (4.27) for each i. For comparison, I also consider the Mean Group (MG) estimator, which is the average β_i estimate from individual OLS regressions applied to (4.25) (Pesaran and Smith, 1995). Hence, the MG estimator accounts for slope heterogeneity, but ignores cross-section dependence.

Pesaran's CD test (2004, 2015) provides insights into the degree and type of cross-section error dependence. Let $\hat{\rho}_{ij}$ denote the sample cross-section correlation

	p_{it}		
	MG	CCEP	CCEMG
y_{it}	0.642***	1.512***	1.464***
	(0.073)	(0.143)	(0.154)
Constant	−1.596***		−0.119
	(0.171)		(0.163)
CD test statistic	227.48	-0.45	-0.47
Avg. correlation ($\hat{\bar{\rho}}$)	0.538	-0.001	-0.001
Observations	7,448	7,448	7,448
R^2	0.762	0.918	0.918
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 4.5: *Long-run relationship between real house prices (p_{it}) and real per capita income (y_{it}). The column 'MG' shows results for the mean group estimator due to Pesaran and Smith (1995), which is obtained by averaging OLS estimates from separate regressions. CCEP and CCEMG are the pooled and mean group version of the CCE estimator, respectively, which augments the regression equations by cross-sectional averages. $\hat{\bar{\rho}}$ denotes the average correlation coefficient across units.*

coefficient between the residuals of unit i and j,

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} \hat{u}_{jt}^2\right)^{1/2}}. \tag{4.28}$$

The estimate of the average correlation coefficient and the CD test statistic are

$$\hat{\bar{\rho}} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{\rho}_{ij} \quad \text{and} \quad CD = \left[\frac{Tn(n-1)}{2}\right]^{1/2} \hat{\bar{\rho}}, \tag{4.29}$$

respectively. Under the null hypothesis of weak cross-section dependence, CD follows a standard normal distribution. However, as shown in Pesaran (2015), the exact formal definition of the null hypothesis depends on the relative speed of convergence of n and T. Intuitively, the notion of weak dependence allows for dependence across units, but requires the degree of dependence to be bounded as $n \to \infty$. Spatial processes satisfy this condition if the spatial weights matrix is uniformly bounded, which is commonly assumed. In contrast, strong dependence arises from unobserved common factors, such as economy-wide shocks, which tend to affect all units independent of the size of the cross-section dimension.

Table 4.5 reports estimation results. The CD test statistics indicate that the inclusion of cross-section averages successfully addresses the issue of cross-section dependence. While the average cross-section correlation coefficient, denoted by $\hat{\bar{\rho}}$, is

0.538 for the residuals from MG estimation, the value drops to close to zero when the CCEP and CCEMG estimator are used. Similar to HPY, $\hat{\beta}$ is below unity for the MG estimator and larger when accounting for cross-section dependence using the CCE estimator. HPY (Table 7) report CCEMG point estimates (and standard errors) of 1.14 (0.20). Using MSA-level data, Baltagi and Li (2014) obtain 0.8219 (0.058). Thus, in contrast to HPY, I reject the null hypothesis of unit elasticity, which is in line with Baltagi and Li (2014).

4.3.5 Panel cointegration testing

In the next step, I test for a cointegration relationship between real house price and real per capita income. If the disequilibrium error u_{it} is stationary, then house prices and income per capita are bound to move together over the long run and are said to be cointegrated. The existence of a cointegration relationship is interpreted as evidence that house prices are driven by economic fundamentals.

The test results from applying the CIPS panel unit root test to the residuals $\hat{u}_{it} = p_{it} - \hat{\alpha}_i - \hat{\beta}_{\text{CCEMG}} y_{it}$ with $\hat{\beta}_{\text{CCEMG}} = 1.464$ are shown in Table 4.4. The relevant critical values for the 5% and 1% level are given by -2.12 and -2.23, respectively. The critical values correspond to the intercept version of the test and $n = 50$ and $T = 100$ (Pesaran, 2007, Table II). Thus, the null hypothesis of a unit root is rejected in favour of a cointegration relationship at the 5% for the considered lag lengths.

4.3.6 Panel error-correction model

Based on the existence of a cointegration relationship, this step examines the short-run temporal dynamics of house prices. A basic panel ECM involving real house prices and real per capita income can be written as

$$\Delta p_{it} = \alpha_i + \phi_i u_{i,t-1} + \delta_{1i} \Delta p_{i,t-1} + \delta_{2i} \Delta y_{it} + \nu_{it}. \tag{4.30}$$

The ECM consists of a short-run and a long-run component. The latter relates to the lagged disequilibrium error, $u_{i,t-1}$, and the speed of adjustment parameter, ϕ_i. For instance, a large positive value of $u_{it} = p_{it} - \alpha_i - \beta_i y_{it}$ reflects that p_{it} is large relative to y_{it}. For the system to return to its equilibrium where $u_{it} = 0$, p_{it} has to decrease or, in other words, Δp_{it} should be negative. Therefore, it is expected that $\phi_i < 0$. The half-life time, which measures how long it takes for a shock to reduce its impact by one half, can be approximated by $-\ln(2)/\ln(1 + \phi_i)$.

Model (1) and (3) in Table 4.6 report results from estimating (4.30) employing the MG and CCEMG estimator when the unobserved $u_{i,t-1}$ is replaced with $\hat{u}_{i,t-1}$. CCEP estimates are omitted for brevity. Model (2) and (4) also include lagged population growth, $\eta_{i,t-1}$, and the lagged real effective interest rate, $r_{i,t-1}$, as additional regressors. The inclusion of cross-sectional averages addresses the issue of strong dependence as indicated by the CD test statistics. Specifically, the CD test statistics, which follow a standard normal distribution under the null of weak dependence, drop from 114.45 and 123.61 to 1.57 and 3.69, respectively. The speed of adjustment coefficient estimates are around -0.08 and highly significant, implying a half-life time between 7.8 to 8.2 quarters.

As to be expected, the coefficient estimates on Δy_{it} are positive, suggesting that an increase in real income is associated with an immediate increase in real house prices. Similar to the results in HPY (Table 10), the estimates drop when cross-section dependence is accounted for using the CCEMG estimator. Also, the sign of the coefficient estimates on $\Delta p_{i,t-1}$ changes from positive to negative when the CCEMG estimator is employed, which suggests a bias in the MG estimates. The effect of population growth on house price changes is positive, while interest rates have a negative effect on house prices, which is consistent with economic theory.

4.3.7 Spatial panel error-correction model

While the CCE estimator successfully controls for cross-section dependence, it does offer only limited insights into the structure of spatial interactions between units.

	Dependent variable: Δp_{it}					
	MG	MG	CCEMG	CCEMG	PDSE	PDSE
	(1)	(2)	(3)	(4)	(5)	(6)
Δy_{it}	0.409***	0.386***	0.282***	0.284***	0.287***	0.270***
	(0.029)	(0.028)	(0.037)	(0.037)	(0.037)	(0.039)
$\Delta \hat{u}_{i,t-1}$	−0.084***	−0.085***	−0.081***	−0.087***	−0.078***	−0.086***
	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)	(0.013)
$\Delta p_{i,t-1}$	0.089**	0.116***	−0.050	−0.033	−0.100**	−0.098**
	(0.037)	(0.040)	(0.039)	(0.046)	(0.041)	(0.047)
$\eta_{i,t-1}$		0.203		0.987***		0.840***
		(0.150)		(0.242)		(0.222)
$r_{i,t-1}$		−0.001***		−0.003*		−0.002*
		(0.0002)		(0.001)		(0.001)
Half-life time	7.92	7.81	8.23	7.63	8.56	7.75
CD test statistic	114.46	123.61	1.57	3.69	1.07	3.08
Spatial lag	–	–	–	–	Δp_{it-1}	Δp_{it-1}
Cross-sectional avg.	No	No	Yes	Yes	Yes	Yes
Observations	7,350	6,860	7,350	6,860	7,350	6,860

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
The half-life time is approximated by $-\ln(2)/\ln(1+\hat{\phi}_i)$. The row 'Spatial lag' indicates which variable the spatial lag is applied to. Note that estimates for the spatial weights ($\hat{w}_{ij}$) are not shown in this table. Model (3) and (4) include cross-sectional averages of all dependent and independent variables. Model (5) includes the cross-sectional averages $\Delta\bar{p}_t$, $\Delta\bar{y}_t$ and $\Delta\bar{\hat{u}}_{t-1}$. Model (6) also adds $\bar{\eta}_{t-1}$ and $\bar{r}_{t-1}$ to the regression. Lasso regressions are based on the package `glmnet` (Friedman, Hastie, and Tibshirani, 2010).

Table 4.6: *Panel error-correction estimates. 'MG' denotes the mean-group estimator and the 'CCEMG' is the mean-group version of the common correlated effects estimator. The last two columns on the right-hand side report results from applying the post-double-selection estimator (PDSE) to the spatial error-correction model.*

To allow for short-run spatial effects, I consider the spatial single-equation ECM,

$$\Delta p_{it} = \alpha_i + \phi_i \hat{u}_{i,t-1} + \delta_{1i}\Delta p_{i,t-1} + \delta_{2i}\Delta y_{it} + \sum_{j\neq i} w_{ij}\Delta p_{j,t-1} + \nu_{it}, \tag{4.31}$$

which includes a spatial lag of $\Delta p_{i,t-1}$. In the spatial ECM, $\hat{u}_{i,t-1}$, $\Delta p_{i,t-1}$ and Δy_{it} represent the low-dimensional regressors and the spatial lag, $\sum_{j\neq i} w_{ij}\Delta p_{j,t-1}$, corresponds to the high-dimensional model component. It should be emphasises that, as shown by Belloni, Chernozhukov, and Hansen (2014b), valid inference can be conducted for the low-dimensional parameters based on the PDSE.

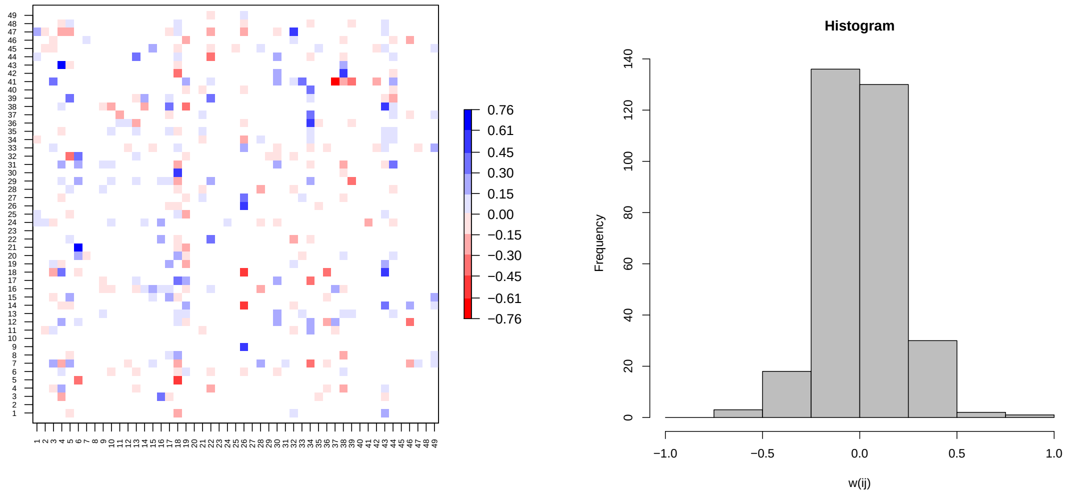
The post-double-selection estimation proceeds as follows for each i separately:

Step 1: The Lasso estimator regresses Δp_{it} against the spatial lag, i.e., $\Delta p_{1,t-1}, \dots, \Delta p_{i-1,t-1}, \Delta p_{i+1,t-1}, \dots, \Delta p_{n,t-1}$. The set of controls selected in this first step is denoted by $\hat{I}_{\Delta p_{it}}$.

Step 2: Similarly, the Lasso is applied to regress $\hat{u}_{i,t-1}$, $\Delta p_{i,t-1}$ and Δy_{it} against the spatial lag. The set of selected controls are denoted by $\hat{I}_{\hat{u}_{i,t-1}}$, $\hat{I}_{\Delta p_{i,t-1}}$ and $\hat{I}_{\Delta y_{it}}$, respectively.

Step 3: PDSE is the least squares estimator of regressing y_{it} against $\hat{u}_{i,t-1}$, $\Delta p_{i,t-1}$ and Δy_{it}, as well as the union of regressors in $\hat{I}_{\Delta p_{it}}$, $\hat{I}_{\hat{u}_{i,t-1}}$, $\hat{I}_{\Delta p_{i,t-1}}$ and $\hat{I}_{\Delta y_{it}}$. In addition, cross-sectional averages are included into the Lasso regression to capture strong cross-sectional dependence.

Model (5)-(6) in Table 4.6 show estimation results for the spatial ECM. The reported point estimates for the low-dimensional parameters are equal to the individual PDSE estimates averaged over i. Model (5) corresponds to equation (4.31) and Model (6) also includes lagged population growth, $\eta_{i,t-1}$, and lagged interest rates, $r_{i,t-1}$, as regressors. The results are qualitatively similar to the CCEMG estimates and largely confirm the results discussed above. Most importantly, the post-double-selection estimation also addresses the issue of strong cross-section dependence as indicated by the CD test statistics, which are slightly lower than the CD statistics from CCEMG estimation.



(a) *Estimated spatial weights* **(b)** *Histogram of spatial weights*

Figure 4.1: *Illustration of estimated spatial weights based on the spatial ECMs in* (4.31)*. The left-hand side plot shows the estimated spatial weights matrix. The numbers on the vertical and horizontal axes are state identifiers, which are listed in Table D.1. Colours represent the size of estimated weights, where blue corresponds to positive weights, red to negative weights and white to* $\hat{w}_{ij} = 0$*. The right-hand side presents the empirical distribution of the estimated weights.*

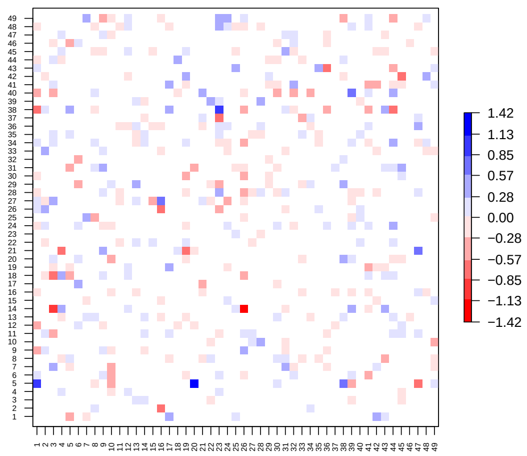
Model	*Avg.*	*St. dev.*	*Counts*		
			$\hat{w}_{ij}$<0	$\hat{w}_{ij}$=0	$\hat{w}_{ij}$>0
Model (5)	0.01703	0.19768	157	2032	163
Model (6)	0.01996	0.21121	153	2035	164

Table 4.7: *Summary statistics for estimated spatial weights. The model numbers in the first column refer to the regression results in Table 4.6. Sample averages ('Avg.') and standard deviations ('St. dev.') are based on non-zero weights only. The three column on the right-hand side show the number of estimated weights which are negative, zero and positive, respectively.*

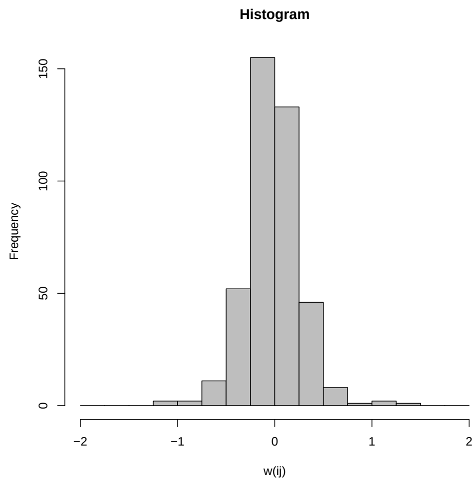
4.3.8 The spatial weights matrix

Estimation of the spatial single-equation spatial ECM in (4.31) allows to investigate short-run interaction effects between US States. In particular, the estimation of the spatial weights, w_{ij}, provides insights into the drivers of spatial effects such as geographic distance.

Figure 4.1a and 4.2a graph the estimated spatial weights matrix. The two figures are based on Models (4.31) and the extended model including $\eta_{i,t-1}$ and $r_{i,t-1}$. Each cell corresponds to one spatial weight. Blue cells indicate positive weights and red weights indicate negative spatial weights, while white cells show that $\hat{w}_{ij}$ =



(a) *Estimated spatial weights*



(b) *Histogram of spatial weights*

Figure 4.2: *Illustration of estimated spatial weights based on the extended spatial ECM in equation* (4.31) *with* $\eta_{i,t-1}$ *and* $r_{i,t-1}$ *added. See notes in Figure 4.1.*

0. Note that the diagonal elements are zero by construction. Furthermore, the histograms in Figure 4.1b and 4.2b show the empirical distribution of estimated spatial weights. The histograms and the summary statistics in Table 4.7 reveal that the estimated weights are centred around zero and that a large fraction of weights is negative. However, notice that the coefficient estimate on the local effect $\Delta p_{i,t-1}$ is also negative, indicating negative serial correlation. Hence, some of the negative spatial weights likely reflect negative serial correlation in house price changes.

Table 4.8 and 4.9 investigate to what extent spatial weights can be explained by observable variables. The dependent variable is a binary indicator which is set to unity if $\hat{w}_{ij}$ is positive or negative. If $\hat{w}_{ij} = 0$, the dependent variable is equal to zero. The estimation method is Probit regression and the table reports the average marginal effects for each regressor. The explanatory variables are the logarithm of geographic distance between state i and j (denoted by $\log(\text{Distance}_{ij})$), a binary contiguity indicator (i.e., Neighbour_{ij}) and the logarithm of state j's population (i.e., $\log(\text{Population}_j)$). In addition, the models include fixed effects for state j, state i or both. The index j corresponds to the origin of the spatial effect w_{ij}, while state i represents the recipient.

	$\mathbb{1}\{\hat{w}_{ij} \neq 0\}$			
	(1)	(2)	(3)	(4)
$\log(\text{Distance}_{ij})$	-0.059^{***}	-0.075^{***}	-0.078^{***}	-0.093^{***}
	(0.013)	(0.013)	(0.014)	(0.013)
Neighbour_{ij}	0.097^{***}	0.067^{**}	0.064^{*}	0.033
	(0.036)	(0.031)	(0.033)	(0.027)
$\log(\text{Population}_{j})$	0.029^{***}	0.032^{***}		
	(0.006)	(0.006)		
j fixed effects	No	No	Yes	Yes
i fixed effects	No	Yes	No	Yes
Observations	2,352	2,352	2,352	2,352
Akaike Inf. Crit.	1,769.781	1,527.144	1,755.154	1,487.759

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Standard errors are robust to heteroskedasticity. The table shows the average marginal effects from Probit estimation. Distance_{ij} denotes the geographic distance between the centroid of state i and state j. Neighbour_{ij} is a binary indicator that is equal to one if state i and j share a common border, zero otherwise.

Table 4.8: *Estimated spatial weights and geographic distance. The results are based on Model* (4.31)*. The dependent variable is coded as unity if the estimated spatial weight is positive or negative, zero otherwise. The table shows the average marginal effects from Probit regression.*

	$\mathbb{1}\{\hat{w}_{ij} \neq 0\}$			
	(1)	(2)	(3)	(4)
$\log(\text{Distance}_{ij})$	-0.054^{***}	-0.075^{***}	-0.066^{***}	-0.087^{***}
	(0.013)	(0.013)	(0.014)	(0.013)
Neighbour_{ij}	0.089^{**}	0.057^{*}	0.074^{**}	0.042
	(0.035)	(0.029)	(0.035)	(0.027)
$\log(\text{Population}_{j})$	0.027^{***}	0.030^{***}		
	(0.007)	(0.006)		
j fixed effects	No	No	Yes	Yes
i fixed effects	No	Yes	No	Yes
Observations	2,352	2,352	2,352	2,352
Akaike Inf. Crit.	1,774.531	1,479.733	1,780.673	1,464.986

See notes in Table 4.8.

Table 4.9: *Estimated spatial weights and geographic distance. The results are based on the extended spatial ECM in equation* (4.31) *with* $\eta_{i,t-1}$ *and* $r_{i,t-1}$ *added. See notes in Table 4.8.*

Table 4.8 uses estimated spatial weights from the estimation equation (4.31). The Probit estimation results suggest that the effect of geographic distance is highly significant. For example in Model (1), if the distance between i and j is doubled, the probability of a non-zero interaction effect between i and j decreases by $-0.059\log(2) = 4.1$ percentage points on average. The effect of first-order contiguity on the probability that $w_{ij} \neq 0$ is highly significant in Model (1) and amounts to 9.7 percentage points, but is less robust to the inclusion of fixed effects. Model (1) and (2) also provide strong evidence that population has a statistically significant impact, with coefficient estimates of 0.029 and 0.032, respectively. The results are largely confirmed in Table 4.9, which corresponds to the extended spatial ECM.

The bar graph in Figure 4.3 presents, based on the spatial ECM in (4.31), the absolute column and row averages for each state.[9] A large absolute column average

[9]The same bar graph corresponding to the extended spatial ECM including $\eta_{i,t-1}$ and $r_{i,t-1}$ is

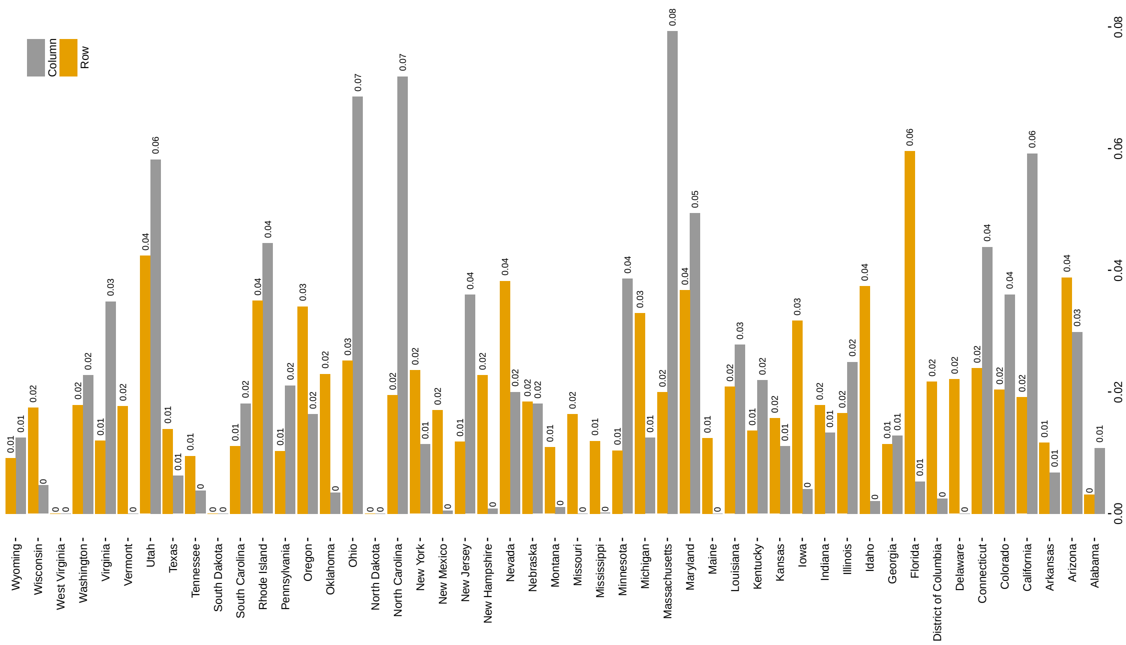


Figure 4.3: *Absolute column and row averages of the estimated spatial weights matrix for each US State. The results are based on the spatial ECM in equation* (4.31).

(in grey colour) suggest that a specific state has a profound 'influence on others', while a large row average (in orange) indicates that the state is susceptible to 'influence by other' US States (borrowing the terminology from Bhattacharjee and Holly, 2013). Thus, the bar graphs allow to identify states with a large leverage such as California (with an absolute column average of 0.059), Massachusetts (0.079), North Caroline (0.072), Ohio (0.068) and Utah (0.058). With an absolute row average of 0.06, Florida is an example of a state that is strongly affected by other states.

4.4 Conclusion

This chapter examines the relationship between house prices and income per capita for US States. At first, it should be noted that the results in HPY and Baltagi and Li (2014) are confirmed. There is robust evidence for a cointegration relationship between real house prices and real per capita income, indicating that house prices are driven by economic fundamentals. However, one limitation of the approach is that, due to the underlying assumption of parameter stability over time, temporary house price bubbles may not be detected. Also, as pointed out by HPY, local house

shown in Figure D.1.

price bubbles may exist, despite the existence of a cointegration relationship for US States on the whole.

In addition to the CCE estimator, a novel method based on the Lasso estimator is considered and applied to a spatial ECM. The proposed framework does control for spatial effects without relying on a pre-specified spatial weights matrix. It is demonstrated that the method also allows for examining the spatial network structure. Specifically, US States susceptible to 'influences by others' and those states which have an 'influence on others' can be identified. Furthermore, the estimated spatial weights are shown to be statistically associated with contiguity and geographic distance as well as the population size of the influencing state. In terms of future research, the framework may provide a basis for rich spatial impulse response functions and predictive models incorporating spatial effects.

Appendix of Chapter 4

D.1 Definition of US regions

Regions	*States*
New England region:	Connecticut (CT), Maine (ME), Massachusetts (MA), New Hampshire (NH), Rhode Island (RI), Vermont (VT)
Mid-East region:	Delaware (DE), District of Columbia (DC), Maryland (MD), New Jersey (NJ), New York (NY), Pennsylvania (PA)
South-East region:	Alabama (AL), Arkansas (AR), Floria (FL), Georgia (GA), Kentucky (KY), Louisiana (LA), Mississippi (MS), North Carolina (NC), South Carolina (SC), Tennessee (TN), Virginia (VA), West Virginia (WV)
Great lakes region:	Illinois (IL), Indiana (IN), Michigan (MI), Ohio (OH), Wisconsin (WI)
Plains region:	Iowa (IA), Kansas (KS), Minnesota (MN), Missouri (MO), Nebraska (NE), North Dakota (ND), South Dakota (SD)
South-West region:	Arizona (AZ), New Mexico (NM), Oklahoma (OK), Texas (TX)
Rocky mountain region:	Colorado (CO), Idaho (ID), Montana (MT), Utah (UT), Wyoming (WY)
Far West region:	Alaska (AL), California (CA), Nevada (NV), Oregon (OR), Washington (WA)

State IDs: AL (1), AR (2), AZ (3), CA (4), CO (5), CT (6), DC (7), DE (8), FL (9), GA (10), IA (11), ID (12), IL (13), IN (14), KS (15), KY (16), LA (17), MA (18), MD (19), ME (20), MI (21), MN (22), MO (23), MS (24), MT (25), NC (26), ND (27), NE (28), NH (29), NJ (30), NM (31), NV (32), NY (33), OH (34), OK (35), OR (36), PA (37), RI (38), SC (39), SD (40), TN (41), TX (42), UT (43), VA (44), VT (45), WA (46), WI (47), WV (48), WY (49).

Table D.1: *Definition of US regions used for Table 4.3. See Holly, Pesaran, and Yamagata (2010, Fig. 2) for comparison.*

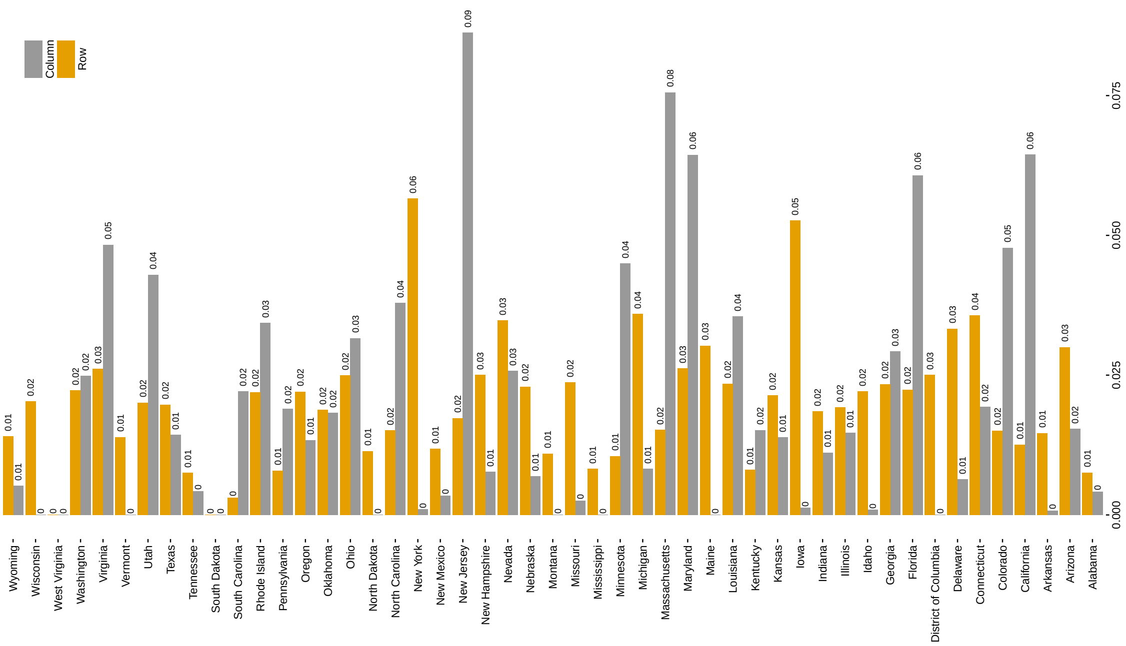


Figure D.1: *Absolute column and row averages of the estimated spatial weights matrix. The results are based on the extended spatial ECM with $\eta_{i,t-1}$ and $r_{i,t-1}$ added.*

Chapter 5

Conclusion

> *Since* estimating is what you do when you do not know*, it is inherent in a great many situations that after reading the estimate you still do not know.*
>
> – Thomas L. Hughes[1]

The above quote is not only true for any econometric estimation, but applies to the research process as whole and, thus, also to this thesis. Although the discussions and results in this thesis point to a promising connection between high-dimensional statistics and spatial econometrics, this work has also revealed further issues to be investigated. In this final chapter, I summarise the key results of this thesis, discuss their implications and point to future research opportunities. The chapter is divided into three topics: model selection, endogeneity in high dimensions and estimation of the spatial weights matrix.

5.1 Model selection

Following the seminal work of Miguel, Satyanath, and Sergenti (2004), a large number of studies have exploited rainfall and temperature variations as instruments for economic growth and other socio-economic variables. However, it is not clear how climate variables should be coded in order to capture the relevant exogenous variation. Various specifications have been considered in the literature and the study

[1]See Hughes, 2013, p. 259. Own emphasis.

by Miguel, Satyanath, and Sergenti (2004) has been criticised for the use of rainfall growth as opposed to, for example, rainfall levels (Ciccone, 2011; Ciccone, 2013).

There are at least three econometric issues characterising the relationship between economic growth and climate variables. First, the relationship is likely to be non-linear. For instance, although rainfall is an input to agricultural production, extreme precipitation levels may reflect adverse weather conditions. Secondly, the relationship is expected to vary substantially across different climate and producing regions. Thirdly, if instruments are not adequately specified, the link between endogenous regressors and instruments may only be weak, carrying the risk of a non-negligible estimation bias.

These issues make it difficult to identify an appropriate set of instruments in practice. The Lasso-based instrument selection developed in Belloni et al. (2012) provides a formal, automated approach to generate approximately optimal instruments from a large set of putative instruments. The compilation of a large set of potential instruments allows to account for non-linearity and spatial heterogeneity without being restricted to a single model specification. The estimation results in Chapter 2 suggest that the Lasso-based approach of generating instruments successfully addresses the discussed issues as indicated by weak identification tests.

The implications of this result are not only relevant for the large body of literature exploiting climate variables as instruments, but are in fact much more general. In most econometric applications, researchers are faced with substantial uncertainty about the form of the true model. To identify the most appropriate model specification, empirical analysts tend to combine economic theory with statistical testing (t and F-tests) and model selection criteria (e.g. R^2, AIC and BIC). However, economic theory often provides insufficient guidance, for example, as to how many lags should be included into a time-series model. A popular framework is the general-to-specific approach that has been put forward by David F. Hendry (see overview in Campos, Ericsson, and Hendry, 2005). The approach starts from a general model including a large number of variables and the model complexity is reduced step-wise guided by significance testing and goodness-of-fit measures. One of the major drawbacks of the general-to-specific approach is that the step-wise model building poses the risk of pre-testing bias. Furthermore, including too many variables in the first stages raises the risk of Type I errors.

The lack of a uniform framework for model selection also leaves room for specification errors and manipulation. Olken (2015) considers a thought experiment that illustrates this point:

> The researcher has 100 different variables he could examine, and the truth is that the experiment has no impact. By construction, the researcher should find an average of five of these variables statistically significantly different between the treatment group and the control group at the 5 percent level [...] Thus, if a researcher can discard or not report all the variables that do not agree with his desired outcome, the researcher is virtually guaranteed a few positive and statistically significant results, even if in fact the experiment has no effect.
>
> Olken, 2015, p. 61

Given that academic journals favour studies which report significant effects, researchers "may [...] search for specifications delivering just-significant results and ignore specifications giving just-insignificant results in order to increase their chances of being published" (Brodeur et al., 2016, p. 2). The empirical analysis in Brodeur et al. (2016) finds evidence supporting this claim. Similarly, Franco, Malhotra, and Simonovits (2014) argue that the publication bias occurs before submission and that authors do not write up results with insignificant statistical results. The results of this publication bias is that p-values slightly below the commonly used 5% significance level appear more frequent than p-values just above 5% (e.g. Winter and Dodou, 2015).[2]

Against this background, this thesis argues that applied econometrics should make extensive use of formal model selection techniques to reduce the risk of misspecification and to facilitate robustness of estimation results. The Lasso and related techniques provide promising methods for model selection. Specifically, Chapter 2

[2] On this topic, see `https://xkcd.com/882/` and `https://xkcd.com/1478/`.

and Chapter 4 demonstrate that the work of Belloni et al. (2012) and Belloni, Chernozhukov, and Hansen (2014a) is of great value when the researcher is faced with a larger number of instruments or a large number of control variables.

The oracle property is a desirable property in this context that has received much attention in the literature on high-dimensional models (Fan and Li, 2001; Zou, 2006). The oracle property is given if, first, asymptotic normality over the true support of the unknown parameter vector and, secondly, that the true support is recovered as the sample size increases (i.e., model selection consistency). Methods exhibiting this property could therefore greatly benefit research in empirical sciences and replace Ordinary Least Squares (OLS) as the benchmark method.

The Lasso estimator requires the irrespresentable condition for model selection consistency, which seems too restrictive in many contexts (see Chapter 1.2.3). Recent studies propose promising alternative methods. For instance, the Adaptive Lasso achieves the oracle property when the number of regressors is fixed (Zou, 2006). Another example is given by Wasserman and Roeder (2009) who propose a multi-stage procedure that combines sample-splitting, cross-validation, screening and hypothesis testing, and can achieve model selection consistency. Fan and Liao (2014) establish the oracle property for the Focused Generalised Methods of Moments (FGMM) estimator under the assumption of over-identification. It remains to be explored further how these recently developed methods can facilitate model selection in applied econometrics.

5.2 Endogeneity in high dimensions

Chapter 3 develops a Two-step Lasso (TSL) estimator that is inspired by Two-stage Least Squares (2SLS) and allows for endogeneity in high-dimensional models. It is shown that the spatial autoregressive panel model is a special case of a high-dimensional model with many endogenous regressors. Based on this insight, the developed TSL estimator is applied to estimate the spatial weights matrix.

This insight suggests that alternative estimators allowing for endogeneity in high dimensions can be exploited to estimate the spatial weights matrix in a similar way. To date, the literature on endogeneity in high-dimensional models is relatively small, but further advances are to be expected in the near future. The Self-Tuning Instrumental Variable (STIV) estimator developed in Gautier and Tsybakov (2014) is one exception. The estimator, which is an extension of the Dantzig estimator due to Candes and Tao (2007), allows for a large number of endogenous regressors assuming that exogenous instruments are available. The FGMM estimator discussed above is based on the Generalised Methods of Moments (GMM) and can also accommodate endogenous regressors. Furthermore, another estimation procedure for high-dimensional models with endogenous regressors could be based on the Square-root Lasso, which is likely to exhibit a smaller finite sample bias than the TSL as the optimal penalty of the Square-root Lasso is independent of the noise level (see Chapter 1.2.4).

Both STIV and FGMM are one-step procedures and, thus, may outperform the TSL and Two-step Post-Lasso (TSPL) considered in this thesis. To my knowledge, there is no attempt to estimate the spatial weights matrix using either estimator, although they seem to provide a promising framework for estimating the spatial weights matrix in a spatial autoregressive panel model. Future research could compare the theoretical properties of TSL, TSPL, STIV and FGMM and assess their performance in Monte Carlo simulations.

5.3 Estimation of the spatial weights matrix

The use of pre-specified spatial weights is one of the weaknesses of the conflict application in Chapter 2. While different specifications were considered, the risk of misspecification is substantial and selecting the true spatial weights matrix is near impossible. Indeed, the use of pre-specified weights and the lack of guidance as to how to select the spatial weights matrix is a major point of criticism of spatial econometrics. Bhattacharjee and Jensen-Butler (2013, p. 618) summarise the problem as follows: "the choice of weights is frequently arbitrary, there is substantial uncertainty regarding the choice, and empirical results vary considerably according to the choice of spatial weights". Another disadvantage is that pre-specified spatial weights do not provide any additional insights into the drivers of network effects, even when the weights matrix is correctly specified. The coefficient on the spatial lag, referred to as spatial autoregressive parameter, only allows to infer the strength of spill-over effects in a network.

For these reasons, Chapter 3 and Chapter 4 attempt to estimate spatial interaction effects in a spatial panel model with lagged dependent variable and lagged exogenous regressors, respectively. Monte Carlo results in Chapter 3 suggest that the TSL estimator, as well as the related TSPL estimator, can successfully recover the structure of the spatial weights matrix as the time dimension grows. The housing market application in Chapter 4 shows that the estimated weights are statistically associated with observable distance measures, indicating that the estimated interaction effects are consistent with theoretical expectations.

While this thesis demonstrates that high-dimensional methods are able to estimate the spatial weights matrix consistently, the issue of conducting statistical inference about the distribution of the spatial weight parameters is challenging and remains for future studies. The Post-double-selection Estimator (PDSE) employed in Chapter 4 is asymptotically normal for the low-dimensional local parameters, but the asymptotic distribution of the high-dimensional spatial parameters is likely to be non-standard. A recently proposed significance test for the Lasso may provide a basis for statistical inference in a high-dimensional setting and could potentially be utilised in the spatial context (Lockhart et al., 2014).

Two further extensions to the developed framework for estimating the spatial weights matrix are worth considering. First, the Lasso-based estimation methodology relies on the assumption that the time dimension (i.e., T) is large and approaches infinity. However, the time dimension is small relative to the cross-section dimension in many econometric applications (particularly, microeconometric panels). Future studies could explore estimation strategies using high-dimensional statistics for panel models where T is fixed or even consider the cross-section case where $T = 1$. However, the small T and cross-section setting is undoubtedly more challenging since less data is available. Secondly, one disadvantage of the TSL-based approach in Chapter 3 is the requirement that exogenous regressors are available. The pure spatial autoregressive model without exogenous regressors is left for future research.

Bibliography

Ahn, Seung C and Peter Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68(1), pp. 5–27.

Ahrens, Achim (2015). Civil Conflicts, Economic Shocks and Night-time Lights. *Peace Economics, Peace Science and Public Policy* 21(4), pp. 433–444.

Ahrens, Achim and Arnab Bhattacharjee (2015). Two-Step Lasso Estimation of the Spatial Weights Matrix. *Econometrics* 3(1), pp. 128–155.

Aidt, Toke S and Gabriel Leon (2015). The Democratic Window of Opportunity: Evidence from Riots in Sub-Saharan Africa. *Journal of Conflict Resolution*, Forthcoming.

Akaike, H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), pp. 716–723.

Anderson, T W and Cheng Hsiao (1981). Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association* 76(375), pp. 598–606.

— (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18(1), pp. 47–82.

Angrist, Joshua D and Alan B Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics* 106(4), pp. 979–1014.

Angrist, Joshua D and J S Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, New Jersey: Princeton University Press.

Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Springer Netherlands.

Anselin, Luc (2001). Spatial econometrics. In: *A companion to theoretical econometrics*. Ed. by Badi H Baltagi. Blackwell Publishing Ltd. Chap. 14, pp. 310–330.

— (2010). Thirty years of spatial econometrics. *Papers in Regional Science* 89(1), pp. 3–25.

Anselin, Luc and Anil K Bera (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In: *Handbook of Applied Economic Statistics*. Ed. by Aman Ullah and David Giles. Statistics: A Series of Textbooks and Monographs. Taylor & Francis. Chap. 7, pp. 237–289.

Anselin, Luc, Julie Le Gallo, and Hubert Jayet (2008). Spatial Panel Econometrics. In: *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Ed. by Lászlo Mátyás and Patrick Sevestre. 3rd ed. Heidelberg: Springer-Verlag. Chap. 19, pp. 625–660.

Arbia, Giuseppe and Bernard Fingleton (2008). New spatial econometric techniques and applications in regional science. *Papers in Regional Science* 87(3), pp. 311–317.

Arellano, Manuel and Stephen Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies* 58(2), pp. 277–297.

Arraiz, Irani et al. (2010). A Spatial Cliff-Ord-Type Model With Heteroskedastic Innovations: Small and Large Sample Results. *Journal of Regional Science* 50(2), pp. 592–614.

Ashworth, John and Simon C Parker (1997). Modelling Regional House Prices in the UK. *Scottish Journal of Political Economy* 44(3), pp. 225–246.

Bai, Jushan (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica* 77(4), pp. 1229–1279.

Bailey, Natalia, Sean Holly, and M Hashem Pesaran (2016). A Two Stage Approach to Spatiotemporal Analysis with Strong and Weak Cross-Sectional Dependence. *Journal of Applied Econometrics* 31(1), pp. 248–280.

Baller, Robert D, Luc Anselin, and Steven F; Messner (2001). Structural covariates of U.S. county homicide rates: Incorporating spatial effects. *Criminology* 39(3), pp. 561–590.

Baltagi, Badi H (2008). *Econometric Analysis of Panel Data*. John Wiley & Sons.

Baltagi, Badi H, Bernard Fingleton, and Alain Pirotte (2014). Estimating and Forecasting with a Dynamic Spatial Panel Data Model. *Oxford Bulletin of Economics and Statistics* 76(1), pp. 112–138.

Baltagi, Badi H and Jing Li (2014). Further Evidence on the Spatio-temporal Model of House Prices in the United States. *Journal of Applied Econometrics* 29(3), pp. 515–522.

Barrios, Salvador, Luisito Bertinelli, and Eric Strobl (2010). Trends in Rainfall and Economic Growth in Africa: A Neglected Cause of the African Growth Tragedy. *Review of Economics and Statistics* 92(2), pp. 350–366.

Bazzi, Samuel and Christopher Blattman (2014). Economic Shocks and Conflict: Evidence from Commodity Prices. *American Economic Journal: Macroeconomics* 6(4), pp. 1–38.

Beck, Nathaniel and Jonathan N Katz (2011). Modeling Dynamics in Time-Series–Cross-Section Political Economy Data. *Annual Review of Political Science* 14, pp. 331–352.

Becker, Gary S (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76(2), pp. 169–217.

Beenstock, Michael and Daniel Felsenstein (2007). Spatial Vector Autoregressions. *Spatial Economic Analysis* 2(2), pp. 167–196.

— (2012). Nonparametric Estimation of the Spatial Connectivity Matrix Using Spatial Panel Data. *Geographical Analysis* 44(4), pp. 386–397.

Belloni, Alexandre and Victor Chernozhukov (2011). High Dimensional Sparse Econometric Models: An Introduction. In: *Inverse Problems and High-Dimensional Estimation SE - 3*. Ed. by Pierre Alquier, Eric Gautier, and Gilles Stoltz. Lecture Notes in Statistics. Springer Berlin Heidelberg, pp. 121–156.

Belloni, Alexandre and Victor Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19(2), pp. 521–547.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014a). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives* 28(2), pp. 29–50.

— (2014b). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies* 81, pp. 608–650.

Belloni, Alexandre, Victor Chernozhukov, and Lie Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika* 98(4), pp. 791–806.

— (2014). Pivotal estimation via square-root Lasso in nonparametric regression. *The Annals of Statistics* 42(2), pp. 757–788.

Belloni, Alexandre et al. (2012). Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica* 80(6), pp. 2369–2429.

Belloni, Alexandre et al. (2016). Inference in High Dimensional Panel Models with an Application to Gun Control. *Journal of Business & Economic Statistics* 34(4), pp. 590–605.

Bergholt, Drago and Päivi Lujala (2012). Climate-related natural disasters, economic growth, and armed civil conflict. *Journal of Peace Research* 49(1), pp. 147–162.

Berman, Nicolas and Mathieu Couttenier (2015). External Shocks, Internal Shots: the Geography of Civil Conflicts. *Review of Economics and Statistics* 95(4), pp. 758–776.

Bhattacharjee, Arnab and Sean Holly (2011). Structural interactions in spatial panels. *Empirical Economics* 40(1), pp. 69–94.

— (2013). Understanding Interactions in Social Networks and Committees. *Spatial Economic Analysis* 8(1), pp. 23–53.

Bhattacharjee, Arnab and Chris Jensen-Butler (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics* 43(4), pp. 617–634.

Bickel, Peter J, Ya'acov Ritov, and Alexandre B Tsybakov (2009). Simultaneous Analysis of Lasso and Dantzig Selector. *The Annals of Statistics* 37(4), pp. 1705–1732.

Blattman, Christopher and Edward Miguel (2010). Civil War. *Journal of Economic Literature* 48(1), pp. 3–57.

Blundell, Richard W and Stephen Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87(1), pp. 115–143.

Bohlken, Anjali Thomas and Ernest John Sergenti (2010). Economic growth and ethnic violence: An empirical investigation of Hindu–Muslim riots in India. *Journal of Peace Research* 47(5), pp. 589–600.

Bound, John, David A Jaeger, and Regina M Baker (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association* 90(430), pp. 443–450.

Brady, Ryan R (2011). Measuring the diffusion of housing prices across space and over time. *Journal of Applied Econometrics* 26(2), pp. 213–231.

— (2014). The spatial diffusion of regional housing prices across U.S. states. *Regional Science and Urban Economics* 46, pp. 150–166.

Breitung, Jörg and M Hashem Pesaran (2008). Unit Roots and Cointegration in Panels. In: *The Econometrics of Panel Data*. Ed. by László Mátyás and Patrick Sevestre. 2nd ed. Vol. 3. Berlin, Heidelberg: Springer-Verlag. Chap. 9, pp. 279–322.

Bridgman, Benjamin (2008). What Does the Atlas Narodov Mira Measure? *Economics Bulletin* 10(6), pp. 1–8.

Brodeur, Abel et al. (2016). Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics* 8(1), pp. 1–32.

Brückner, Markus and Antonio Ciccone (2010). International commodity prices, growth and the outbreak of civil war in Sub-Saharan Africa. *Economic Journal* 120(544), pp. 519–534.

— (2011). Rain and the Democratic Window of Opportunity. *Econometrica* 79(3), pp. 923–947.

Bruk, Solomon I and V S Apenchenko (1964). *Atlas Narodov Mira.* Moscow: Glavnoe Upravlenie Geodezii i Kartografii.

Brunsdon, C, A S Fotheringham, and M E Charlton (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28(4), pp. 281–298.

Buhaug, Halvard (2010). Climate not to blame for African civil wars. *Proceedings of the National Academy of Sciences of the United States of America* 107(38), pp. 16477–16482.

Buhaug, Halvard and Kristian Skrede Gleditsch (2008). Contagion or confusion? Why conflicts cluster in space. *International Studies Quarterly* 52, pp. 215–233.

Buhaug, Halvard et al. (2011). It's the Local Economy, Stupid! Geographic Wealth Dispersion and Conflict Outbreak Location. *Journal of Conflict Resolution* 55(5), pp. 814–840.

Buhaug, Halvard et al. (2015). Climate variability, food production shocks, and violent conflict in Sub-Saharan Africa. *Environmental Research Letters*, Forthcoming.

Bühlmann, Peter and Sara Van de Geer (2011). *Statistics for High-Dimensional Data.* Berlin, Heidelberg: Springer-Verlag.

Burke, Marshall B, Solomon M Hsiang, and Edward Miguel (2015). Global non-linear effect of temperature on economic production. *Nature* 527(7577), pp. 235–239.

Burke, Marshall B et al. (2009). Warming increases the risk of civil war in Africa. *Proceedings of the National Academy of Sciences of the United States of America* 106, pp. 20670–20674.

Burke, Paul J and Andrew Leigh (2010). Do Output Contractions Trigger Democratic Change? *American Economic Journal: Macroeconomics* 2(4), pp. 124–157.

Burridge, Peter and Bernard Fingleton (2010). Bootstrap inference in spatial econometrics: the J-test. *Spatial Economic Analysis* 5(1), pp. 93–119.

Butts, Carter T (2008). network: a Package for Managing Relational Data in R. *Journal of Statistical Software* 24(2).

— (2015). *network: Classes for Relational Data.* The Statnet Project. URL: `http://cran.r-project.org/package=network`.

CIESIN and CIAT (2005). *Gridded Population of the World, Version 3 (GPWv3): Population Count Grid.* Palisades and NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Campbell, John Y and João F Cocco (2007). How do house prices affect consumption? Evidence from micro data. *Journal of Monetary Economics* 54(3), pp. 591–621.

Campos, Julia, Neil R Ericsson, and David F Hendry (2005). General-to-specific modeling: an overview and selected bibliography. *Unpublished manuscript.* International Finance Discussion Papers No. 839. URL: `http://www.federalreserve.gov/pubs/ifdp/2005/838/ifdp838.pdf`.

Candes, Emmanuel and Terence Tao (2007). The Dantzig Selector: Statistical Estimation When p Is Much Larger than n. *The Annals of Statistics* 35(6), pp. 2313–2351.

Caner, Mehmet (2009). Lasso-Type Gmm Estimator. *Econometric Theory* 25(1), pp. 270–290.

Case, Karl E and Robert J Shiller (2003). Is There a Bubble in the Housing Market? *Brookings Papers on Economic Activity* 34(2), pp. 299–362.

Chassang, Sylvain and Gerard i Miquel (2009). Economic Shocks and Civil War. *Quarterly Journal of Political Science* 4(3), pp. 211–228.

Chow, Gregory C and An-loh Lin (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics* 53(4), pp. 372–375.

Chudik, Alexander and M Hashem Pesaran (2013). *Large panel data models with cross-sectional dependence: a survey.* Globalization and Monetary Policy Institute Working Paper No. 153. Federal Reserve Bank of Dallas. URL: http://www.dallasfed.org/assets/documents/institute/wpapers/2013/0153.pdf.

— (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics* 188(2), pp. 393–420.

Chudik, Alexander, M Hashem Pesaran, and Elisa Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal* 14(1), pp. C45–C90.

Ciccone, Antonio (2011). Economic shocks and civil conflict: A comment. *American Economic Journal: Applied Economics* 3, pp. 215–227.

— (2013). Estimating the Effect of Transitory Economic Shocks on Civil Conflict. *Review of Economics and Institutions* 4(2), pp. 1–14.

Cliff, A D and Keith Ord (1973). *Spatial autocorrelation.* Monographs in spatial and environmental systems analysis. London: Pion.

— (1981). *Spatial Processes: Models & Applications.* London: Pion.

Coakley, Jerry, Ana-Maria Fuertes, and Ron Smith (2002). *A Principal Components Approach to Cross-Section Dependence in Panels.* 10th International Conference on Panel Data, Berlin, July 5-6, 2002. International Conferences on Panel Data. URL: https://ideas.repec.org/p/cpd/pd2002/b5-3.html.

Collier, Paul and Anke Hoeffler (1998). On economic causes of civil war. *Oxford Economic Papers* 50(4), pp. 563–573.

— (2004). Greed and grievance in civil war. *Oxford Economic Papers* 56(4), pp. 563–595.

Collier, Paul, Anke Hoeffler, and Dominic Rohner (2009). Beyond greed and grievance: Feasibility and civil war. *Oxford Economic Papers* 61(1), pp. 1–27.

Corrado, Luisa and Bernard Fingleton (2012). Where is the Economics in Spatial Econometrics? *Journal of Regional Science* 52(2), pp. 210–239.

Couttenier, Mathieu and Raphael Soubeyran (2014). Drought and Civil War In Sub-Saharan Africa. *The Economic Journal* 124(575), pp. 201–244.

Croissant, Yves and Giovanni Millo (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software* 27(2), pp. 1–43.

Dal Bó, Ernesto and Pedro Dal Bó (2011). Workers, warriors, and criminals: Social conflict in general equilibrium. *Journal of the European Economic Association* 9(4), pp. 646–677.

Davidson, Russell and James G Mackinnon (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica* 49(3), pp. 781–794.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken (2008). Climate change and economic growth: Evidence from the last half century. *Unpublished manuscript.* NBER Working Papers No. 14132. URL: http://www.nber.org/papers/w14132.

— (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics* 4, pp. 66–95.

Dickey, David A and Wayne A Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74(366), pp. 427–431.

Dijk, Bram van et al. (2011). Modelling regional house prices. *Applied Economics* 43(17), pp. 2097–2110.

Djankov, Simeon and Marta Reynal-Querol (2010). Poverty and Civil War: Revisiting the Evidence. *Review of Economics and Statistics* 92(4), pp. 1035–1041.

Doll, Christopher (2008). *CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications.* Tech. rep. Palisades, NY, USA: Center for International Earth Science Information Network (CIESIN).

Donohue, John J and Steven D Levitt (2001). The Impact of Legalized Abortion on Crime. *The Quarterly Journal of Economics* 116(2), pp. 379–420.

Drake, Leigh (1995). Testing for Convergence between UK Regional House Prices. *Regional Studies* 29(4), pp. 357–366.

Drukker, David M, Peter Egger, and Ingmar R Prucha (2013). On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors. *Econometric Reviews* 32(5-6), pp. 686–733.

Drukker, David M, Ingmar R Prucha, and Rafal Raciborski (2013a). A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. *Stata Journal* 13(2), pp. 287–301.

— (2013b). Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. *Stata Journal* 13(2), pp. 221–241.

Dube, Oeindrila and Juan F Vargas (2013). Commodity price shocks and civil conflict: Evidence from Colombia. *Review of Economic Studies* 80(4), pp. 1384–1421.

Eberhardt, Markus and Francis Teal (2013). No Mangoes in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis. *Oxford Bulletin of Economics and Statistics* 75(6), pp. 914–939.

Egger, Peter, Michael Pfaffermayr, and Hannes Winner (2005). An unbalanced spatial panel data approach to US state tax competition. *Economics Letters* 88(3), pp. 329–335.

Elhorst, J Paul (2005). Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels. *Geographical Analysis* 37(1), pp. 85–106.

— (2010a). Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis* 5(1), pp. 9–28.

— (2010b). Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics* 40(5), pp. 272–282.

— (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels.* Heidelberg: Springer-Verlag.

Elliott, Robert J R, Eric Strobl, and Puyang Sun (2015). The local impact of typhoons on economic activity in China: A view from outer space. *Journal of Urban Economics* 88, pp. 50–66.

Elvidge, Christopher D et al. (2009a). A fifteen year record of global natural gas flaring derived from satellite data. *Energies* 2, pp. 595–622.

Elvidge, Christopher D et al. (2009b). Overview of DMSP nightime lights and future possibilities. In: *2009 Joint Urban Remote Sensing Event*.

Ertur, Cem and Wilfried Koch (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics* 22(6), pp. 1033–1062.

Fama, Eugene F (2014). Two pillars of asset pricing. *The American Economic Review* 104(6), pp. 1467–1485.

Fan, Jianqing and Runze Li (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association* 96(456), pp. 1348–1360.

Fan, Jianqing and Yuan Liao (2014). Endogeneity in high dimensions. *The Annals of Statistics* 42(3), pp. 872–917.

Fearon, James D (2008). Economic Development, Insurgency, and Civil War. In: *Institutions and Economic Performance*. Ed. by Elhanan Helpman. Cambridge: Harvard University Press.

Fearon, James D and David D Laitin (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97(1), pp. 75–90.

Fingleton, Bernard (2008). A Generalized Method of Moments Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors. *Spatial Economic Analysis* 3(1), pp. 27–44.

Fingleton, Bernard and Julie Le Gallo (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science* 87(3), pp. 319–339.

Fjelde, Hanne and Nina von Uexkull (2012). Climate triggers: Rainfall anomalies, vulnerability and communal conflict in Sub-Saharan Africa. *Political Geography* 31(7), pp. 444–453.

Franco, Annie, Neil Malhotra, and Gabor Simonovits (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science* 345(6203), pp. 1502–1505.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33(1), pp. 1–22.

Gallin, Joshua (2006). The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets. *Real Estate Economics* 34(3), pp. 417–438.

Gautier, Eric and Alexandre B Tsybakov (2014). High-dimensional instrumental variables regression and confidence sets. *Unpublished manuscript.* URL: `http://arxiv.org/abs/1105.2454`.

Getis, Arthur and Jared Aldstadt (2004). Constructing the Spatial Weights Matrix Using a Local Statistic. *Geographical Analysis* 36, pp. 90–104.

Giussani, Bruno and George Hadjimatheou (1991). Modeling Regional House Prices in the United Kingdom. *Papers in Regional Science* 70(2), pp. 201–219.

Goodhart, Charles and Boris Hofmann (2008). House prices, money, credit, and the macroeconomy. *Oxford Review of Economic Policy* 24(1), pp. 180–205.

Grossman, Herschel I (1995). Insurrections. *Handbook of Defense Economics* 1, pp. 191–212.

Haining, R P (1978). The Moving Average Model for Spatial Interaction. *Transactions of the Institute of British Geographers* 3(2), pp. 202–225.

Harari, Mariaflavia and Eliana La Ferrara (2013). Conflict, Climate and Cells: A disaggregated analysis. *Unpublished manuscript.* IGIER Working Paper No. 461. URL: `http://economics.mit.edu/files/10058`.

Harris, Richard, John Moffat, and Victoria Kravtsova (2011). In Search of 'W'. *Spatial Economic Analysis* 6(3), pp. 249–270.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning.* 2nd ed. New York: Springer-Verlag.

Hastie, Trevor, Robert Tibshirani, and Martin J Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations.* Monographs on Statistics & Applied Probability. Boca Raton: CRC Press, Taylor & Francis.

Heal, Geoffrey M and Jisung Park (2013). Feeling the Heat: Temperature, Physiology & the Wealth of Nations. *Unpublished manuscript.* NBER Working Papers No. 19725. URL: `http://www.nber.org/papers/w19725`.

Hegre, H and Nicholas Sambanis (2006). Sensitivity Analysis of Empirical Results on Civil War Onset. *Journal of Conflict Resolution* 50(4), pp. 508–535.

Henderson, J Vernon, Adam Storeygard, and David N Weil (2012). Measuring economic growth from outer space. *American Economic Review* 102(2), pp. 994–1028.

Hendrix, Cullen S and Sarah M Glaser (2007). Trends and triggers: Climate, climate change and civil conflict in Sub-Saharan Africa. *Political Geography* 26(6), pp. 695–715.

Hendrix, Cullen S and Idean Salehyan (2012). Climate change, rainfall, and social conflict in Africa. *Journal of Peace Research* 49(1), pp. 35–50.

Hidalgo, F Daniel et al. (2010). Economic Determinants of Land Invasions. *Review of Economics and Statistics* 92(3), pp. 505–523.

Hijmans, Robert J and Jacob van Etten (2014). *raster: Geographic analysis and modeling with raster data.* URL: `http://cran.r-project.org/package=raster`.

Hodler, Roland and Paul A Raschky (2014). Economic shocks and civil conflict at the regional level. *Economics Letters* 124(3), pp. 530–533.

Hoerl, Arthur E and Robert W Kennard (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* 12(1), pp. 55–67.

Holly, Sean, M Hashem Pesaran, and Takashi Yamagata (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics* 158(1), pp. 160–173.

— (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics* 69(1), pp. 2–23.

Holm, Sture (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6(2), pp. 65–70.

Horrace, William C and Ronald L Oaxaca (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters* 90(3), pp. 321–327.

Hotelling, Harold (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24(6), p. 417.

Hsiang, Solomon M, Marshall B Burke, and Edward Miguel (2013). Quantifying the influence of climate on human conflict. *Science* 341, p. 1235367.

Huang, Jian, Shuangge Ma, and Cun-Hui Zhang (2008). Adaptive Lasso for Sparse High-Dimensional Regression Models. *Statistica Sinica* 18(4), pp. 1603–1618.

Hughes, T L (2013). *Speaking Up and Speaking Out*. Xlibris.

IPCC (2007). *Climate Change 2007: Synthesis Report: Fourth Assessment Report (AR4)*. Tech. rep. Intergovernmental Panel on Climate Change.

Iacoviello, Matteo and Stefano Neri (2010). Housing Market Spillovers: Evidence from an Estimated DSGE Model. *American Economic Journal: Macroeconomics* 2(2), pp. 125–164.

Im, Kyung So, M Hashem Pesaran, and Yongcheol Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115(1), pp. 53–74.

James, G et al. (2014). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer New York.

Jensen, Peter Sandholt and Kristian Skrede Gleditsch (2009). Rain, Growth and Civil War: The Importance of Location. *Defence and Peace Economics* 20(5), pp. 359–372.

Jing, Bing-Yi, Qi-Man Shao, and Qiying Wang (2003). Self-normalized Cramér-type large deviations for independent random variables. *The Annals of Probability* 31(4), pp. 2167–2215.

Johnston, J and J E DiNardo (1997). *Econometric Methods*. 4th ed. McGraw-Hill economics series. McGraw-Hill.

Jorda, Oscar (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review* 95(1), pp. 161–182.

Kapetanios, G, M Hashem Pesaran, and Takashi Yamagata (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics* 160(2), pp. 326–348.

Kapoor, Mudit, Harry H Kelejian, and Ingmar R Prucha (2007). Panel data models with spatially correlated error components. *Journal of Econometrics* 140(1), pp. 97–130.

Kelejian, Harry H (2008). A spatial J-test for model specification against a single or a set of non-nested alternatives. *Letters in Spatial and Resource Sciences* 1(1), pp. 3–11.

Kelejian, Harry H and Gianfranco Piras (2011). An extension of Kelejian's J-test for non-nested spatial models. *Regional Science and Urban Economics* 41(3), pp. 281–292.

Kelejian, Harry H and Ingmar R Prucha (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* 17(1), pp. 99–121.

— (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40(2), pp. 509–533.

— (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157(1), pp. 53–67.

Kim, Nam Kyu (2014). Revisiting Economic Shocks and Coups. *Journal of Conflict Resolution*, Forthcoming.

Kleibergen, Frank and Richard Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133(1), pp. 97–126.

Knight, Keith and Wenjiang Fu (2000). Asymptotics for Lasso-Type Estimators. *The Annals of Statistics* 28(5), pp. 1356–1378.

Koubi, Vally et al. (2012). Climate variability, economic growth, and civil conflict. *Journal of Peace Research* 49(1), pp. 113–127.

Kuethe, Todd H and Valerien O Pede (2011). Regional Housing Price Cycles: A Spatio-temporal Analysis Using US State-level Data. *Regional Studies* 45(5), pp. 563–574.

Lake, David A and Donald Rothchild (1998). Spreading fear: The genesis of transnational ethnic conflict. In: *The international spread of ethnic conflict: Fear, diffusion, and escalation*. Ed. by David A Lake and Donald Rothchild. Princeton, New Jersey. Chap. 1, pp. 3–32.

Lam, Clifford and Pedro CL Souza (2013). Regularization for Spatial Panel Time Series Using Adaptive Lasso. *Unpublished manuscript*. URL: `http://stats.lse.ac.uk/lam/RSPT.pdf`.

Lanzafame, Matteo (2014). Temperature, rainfall and economic growth in Africa. *Empirical Economics* 46(1), pp. 1–18.

Laphroaig (2016). *Laphroaig Single Malt Whisky - Triple Wood*. Port Ellen: Laphroaig distillery.

LeSage, James and R Kelly Pace (2009). *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. Boca Raton: Taylor & Francis.

Lee, Lung-fei (2002). Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models. *Econometric Theory* 18(2), pp. 252–277.

— (2004). Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* 72(6), pp. 1899–1925.

Lee, Lung-fei and Jihai Yu (2010a). A Spatial Dynamic Panel Data Model With Both Time And Individual Fixed Effects. *Econometric Theory* 26(2), pp. 564–597.

— (2010b). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154(2), pp. 165–185.

— (2010c). Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40(5), pp. 255–271.

— (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics* 180(2), pp. 174–197.

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics* 108(1), pp. 1–24.

Li, Zitong and Mikko J. Sillanpää (2012). Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. *Theoretical and Applied Genetics* 125(3), pp. 419–435.

Lin, Wei, Rui Feng, and Hongzhe Li (2014). Regularization Methods for High-Dimensional Instrumental Variables Regression With an Application to Genetical Genomics. *Journal of the American Statistical Association* (509), p. 0.

Lockhart, Richard et al. (2014). A Significance Test for the Lasso. *Annals of Statistics* 42(2), pp. 413–468.

Malpezzi, Stephen (1999). A Simple Error Correction Model of House Prices. *Journal of Housing Economics* 8(1), pp. 27–62.

Manresa, Elena (2013). Estimating the Structure of Social Interactions Using Panel Data. *Unpublished manuscript.* URL: `http://www.hec.unil.ch/documents/seminars/deep/1281.pdf`.

Manski, Charles F (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* 60(3), pp. 531–542.

McMillen, Daniel P (1996). One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach. *Journal of Urban Economics* 40(1), pp. 100–124.

Meen, Geoffrey (1996). Spatial aggregation, spatial dependence and predictability in the UK housing market. *Housing Studies* 11(3), pp. 345–372.

— (1999). Regional House Prices and the Ripple Effect: A New Interpretation. *Housing Studies* 14(6), pp. 733–753.

— (2002). The Time-Series Behavior of House Prices: A Transatlantic Divide? *Journal of Housing Economics* 11(1), pp. 1–23.

Meinshausen, Nicolai and Peter Bühlmann (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics* 34(3), pp. 1436–1462.

Michalopoulos, Stelios and Elias Papaioannou (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica* 81(1), pp. 113–152.

— (2015). Further evidence on the link between pre-colonial political centralization and comparative economic development in Africa. *Economics Letters* 126, pp. 57–62.

Miguel, Edward (2005). Poverty and witch killing. *Review of Economic Studies* 72, pp. 1153–1172.

Miguel, Edward and Shanker Satyanath (2010). Understanding Transitory Rainfall Shocks, Economic Growth and Civil Conflict. *Unpublished manuscript.* NBER Working Paper No. 16461. URL: `http://www.nber.org/papers/w16461`.

— (2011). Re-examining economic shocks and civil conflict. *American Economic Journal: Applied Economics* 3, pp. 228–232.

Miguel, Edward, Shanker Satyanath, and Ernest John Sergenti (2004). Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy* 112, pp. 725–753.

Mol, Christine De, Ernesto De Vito, and Lorenzo Rosasco (2009). Elastic-net regularization in learning theory. *Journal of Complexity* 25(2), pp. 201–230.

Murdoch, James C and Todd Sandler (2004). Civil Wars and Economic Growth: Spatial Dispersion. *American Journal of Political Science* 48(1), pp. 138–151.

Mutl, Jan (2006). Dynamic Panel Data Models with Spatially Correlated Disturbances. PhD thesis. University of Maryland. URL: `http://drum.lib.umd.edu/bitstream/handle/1903/3729/umi-umd-3553.pdf;sequence=1`.

Mutl, Jan and Michael Pfaffermayr (2011). The Hausman test in a Cliff and Ord panel model. *The Econometrics Journal* 14(1), pp. 48–76.

Nickell, Stephen (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* 49(6), pp. 1417–1426.

Nordhaus, William D and Xi Chen (2012). Improved Estimates of Using Luminosity as a Proxy for Economic Statistics: New Results and Estimates of Precision. *Unpublished manuscript.* Cowles Foundation Discussion Paper No. 1857. URL:

http://cowles.yale.edu/sites/default/files/files/pub/d18/d1857.pdf.

O'Loughlin, John et al. (2012). Climate variability and conflict risk in East Africa, 1990-2009. *Proceedings of the National Academy of Sciences of the United States of America* 109(45), pp. 18344–18349.

Olken, Benjamin A (2015). Promises and Perils of Pre-analysis Plans. *Journal of Economic Perspectives* 29(3), pp. 61–80.

Ord, Keith (1975). Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association* 70(349), pp. 120–126.

Ord, Keith and Arthur Getis (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* 27(4), pp. 286–306.

Palmer, Wayne C (1965). *Meteorological drought*. Research Paper No.45, U.S. Department of Commerce. URL: http://www.ncdc.noaa.gov/temp-and-precip/drought/docs/palmer.pdf.

Pesaran, M Hashem (2004). General Diagnostic Tests for Cross Section Dependence in Panels. *Unpublished manuscript*. Cambridge Working Papers in Economics No. 0435. URL: http://www.econ.cam.ac.uk/research/repec/cam/pdf/cwpe0435.pdf.

— (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4), pp. 967–1012.

— (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22(2), pp. 265–312.

— (2015). Testing Weak Cross-Sectional Dependence in Large Panels. *Econometric Reviews* 34(6-10), pp. 1089–1117.

Pesaran, M Hashem and Ron Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68(1), pp. 79–113.

Pesaran, M Hashem and Elisa Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, pp. 182–202.

Pinkse, Joris, Margaret E Slade, and Craig Brett (2002). Spatial Price Competition: A Semiparametric Approach. *Econometrica* 70(3), pp. 1111–1153.

Piras, Gianfranco and Nancy Lozano-Gracia (2012). Spatial J-test: some Monte Carlo evidence. *Statistics and Computing* 22(1), pp. 169–183.

Pollakowski, Henry O and Traci S Ray (1997). Housing price diffusion patterns at different aggregation levels: an examination of housing market efficiency. *Journal of Housing Research* 8(1), p. 107.

Raleigh, Clionadh and Dominic Kniveton (2012). Come rain or shine: An analysis of conflict and climate variability in East Africa. *Journal of Peace Research* 49(1), pp. 51–64.

Raskutti, Garvesh, Martin J Wainwright, and Bin Yu (2010). Restricted Eigenvalue Properties for Correlated Gaussian Designs. *Journal of Machine Learning Research* 11, pp. 2241–2259.

Robertson, Donald and James Symons (2007). Maximum likelihood factor analysis with rank-deficient sample covariance matrices. *Journal of Multivariate Analysis* 98(4), pp. 813–828.

Roodman, David (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics* 71(1), pp. 135–158.

Salehyan, Idean and Kristian Skrede Gleditsch (2006). Refugees and the Spread of Civil War. *International Organization* 60(2), pp. 335–366.

Sanderson, Eleanor and Frank Windmeijer (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics* 190(2), pp. 212–221.

Sarafidis, Vasilis and Tom Wansbeek (2012). Cross-Sectional Dependence in Panel Data Analysis. *Econometric Reviews* 31(5), pp. 483–531.

Sax, Christoph and Peter Steiner (2013). Temporal Disaggregation of Time Series. *The R Journal* 5(2), pp. 80–87.

Schaffer, Mark E (2012). *XTIVREG2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models.* URL: https://ideas.repec.org/c/boc/bocode/s456501.html.

Schwarz, Gideon (1978). Estimating the Dimension of a Model. *The Annals of Statistics* 6(2), pp. 461–464.

Souza, Pedro CL (2012). Estimating Networks: Lasso for Spatial Weights. *Unpublished manuscript*. URL: http://bibliotecadigital.fgv.br/ocs/index.php/sbe/EBE12/paper/download/3641/1437.

Stern, Nicholas (2007). *The Economics of Climate Change: The Stern Review*. Cambridge: Cambridge University Press.

Stiglitz, Joseph E (1990). Symposium on Bubbles. *The Journal of Economic Perspectives* 4(2), pp. 13–18.

Stock, James H and Motohiro Yogo (2005). Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. Ed. by Donald W. K. Andrews, James H. Stock, and Thomas J. Rothenberg, pp. 80–108.

Stone, Richard (1947). On the Interdependence of Blocks of Transactions. *Supplement to the Journal of the Royal Statistical Society* 9(1), pp. 1–45.

Su, Liangjun and Zhenlin Yang (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185(1), pp. 230–258.

Sundberg, Ralph and Erik Melander (2010). *UCDP GED Codebook version 1.9-2015*. Tech. rep. Department of Peace and Conflict Research, Uppsala University.

— (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50(4), pp. 523–532.

Tetlock, Philip E and Daniel Gardner (2015). *Superforecasting: The Art and Science of Prediction*. Random House.

Tibshirani, Robert (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1), pp. 267–288.

Van de Geer, Sara and Peter Bühlmann (2009). On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics* 3, pp. 1360–1392.

Varian, Hal R (2014). Big Data: New Tricks for Econometrics. *The Journal of Economic Perspectives* 28(2), pp. 3–27.

Wainwright, Martin J (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using L1-constrained quadratic programming. *IEEE Transactions on Information Theory* 55, pp. 2183–2202.

Wasserman, Larry and Kathryn Roeder (2009). High-dimensional variable selection. *Annals of Statistics* 37(5A), pp. 2178–2201.

Weezel, Stijn van (2015). Economic shocks & civil conflict onset in Sub-Saharan Africa, 1981-2010. *Defence and Peace Economics* 26(2), pp. 153–177.

Weidmann, Nils B, Jan Ketil Rød, and Lars-Erik Cederman (2010). Representing ethnic groups in space: A new dataset. *Journal of Peace Research* 47(4), pp. 491–499.

Wendt, Daniel, Luc JC van Loon, and Wouter D van Marken Lichtenbelt (2007). Thermoregulation during Exercise in the Heat. *Sports Medicine* 37(8), pp. 669–682.

Whittle, P (1954). On Stationary Processes in the Plane. *Biometrika* 41(3/4), pp. 434–449.

Willmott, Cort J and Kenji Matsuura (2013). *Terrestrial Air Temperature and Precipitation: Monthly and Annual Climatologies (Version 3.01)*. Center for Climatic Research, Department of Geography, University of Delaware.

Winter, Joost C F de and Dimitra Dodou (2015). A surge of p-values between 0.041 and 0.049 in recent decades (but negative results are increasing rapidly too). *PeerJ* 3, e733.

Wooldridge, Jeffrey M (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Yu, Jihai, Robert de Jong, and Lung-fei Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146(1), pp. 118–134.

Zhao, Peng and Bin Yu (2006). On Model Selection Consistency of Lasso. *Journal of Machine Learning Research* 7, pp. 2541–2563.

Zou, Hui (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101(476), pp. 1418–1429.

Zou, Hui and Trevor Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 67(2), pp. 301–320.

Index