ESSAYS ON INVESTMENTS AND ENVIRONMENT:
A SPATIAL ECONOMETRICS PERSPECTIVE

PhD Thesis

Submitted to the Faculty of Economics at the University of Neuchatel

by

José-Antonio Monteiro

Under the supervision of

Jean-Marie Grether, Professor of Economics, University of Neuchâtel

Members of the Committee:

Robert J.R. Elliott, Professor of Economics, University of Birmingham

Peter Egger, Professor of Economics, Swiss Federal Institute of Technology

Milad Zarin, Professor of Economics, University of Neuchâtel

Defended in Neuchâtel, 16 December 2010

UNIVERSITÉ DE NEUCHÂTEL

IMPRIMATUR POUR LA THÈSE

Essays on Investments and Environment : A Spatial Econometrics Perspective

José-Antonio MONTEIRO

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

La Faculté des sciences économiques,
sur le rapport des membres du jury

Prof. Jean-Marie Grether (directeur de thèse, Université de Neuchâtel)
Prof. Milad Zarin-Nejadan (président du jury, Université de Neuchâtel)
Prof. Peter Egger (Ecole Polytechnique Fédérale de Zürich)
Prof. Robert J.R. Elliot (Université de Birmingham, Grande-Bretagne)

Autorise l'impression de la présente thèse.

Neuchâtel, le 24 février 2011

Le doyen

J-M Grether

Jean-Marie Grether

■ Téléphone : +41 32 718 15 00 ■ Fax : +41 32 718 15 01 ■ E-mail : secretariat.seco@unine.ch ■ www.unine.ch

To my parents, for all their sacrifices that made this thesis a reality.

And to all the tree huggers in the world.

PS: *praise the bridge that carried you over.*

EXECUTIVE SUMMARY

This PhD dissertation investigates the links between foreign direct investment (FDI), pollution and environmental policies in an interdependent world. To tackle the issue of spatial dependence, I propose to apply new spatial estimators.

The thesis consists of four papers. The first chapter, entitled *Spatial Dynamic Panel and System GMM: a Monte-Carlo Investigation*, investigates the finite sample properties of estimators for spatial dynamic panel models in the presence of several endogenous variables. I show that in order to account for the endogeneity of several covariates and get unbiased estimates, you should apply the system generalized moments method (GMM) estimator instead of traditional spatial dynamic estimators. From a practical point of view, system GMM is computationally less demanding and easier to adapt to unbalanced panel data.

In the second paper, *Pollution Havens: a Spatial Vector Autoregression Approach* (VAR), I investigate the mutual relationship between FDI, pollution and trade openness for 141 countries during the nineties. After applying second generation panel unit root tests, I estimate a trivariate spatial VAR model where a shock occurring in a given country can potentially affect the economic conditions in the neighboring countries. Based on generalized empirical likelihood estimations, the results highlight a reverse causality between FDI and pollution emission, suggesting the existence of a potential endogenous pollution haven effect to low income countries in particular. In addition, spatial spillovers play a key role in the linkage between FDI, trade openness and SO_2 emission and highlight the strategic nature of these variables. Multilateral and regional cooperation between developed, emerging and developing countries should be extended to capitalize on these spillovers, address potential issues and seize new opportunities.

In line with the previous study, in the third paper, entitled *Complex FDI and Environmental Regulation: the Role of Spatial Dependence*, I investigate if differences in environmental regulations influenced OECD's FDI allocation during 1981-2000 taking into account "third-country" effects in a multi-country setting. The findings, based on system GMM estimates of a spatial dynamic gravity model, confirm the existence of a negative relationship between FDI and environmental stringency, once I correct for endogeneity and spatial dependence. In addition, the evidence of positive "third-country" effects for FDI suggests the prevalence of highly complex vertically integrated FDI from OECD countries to developing economies. Multinationals tend to allocate each part of the production process in different countries to ensure cost minimization. Emerging and developing countries, which might be tempted to use environmental regulation as an instrument to attract FDI, will only attract the most polluting part of the production process. This is not the best way to ensure sustainable economic growth, at least on the long run.

In the last chapter called *Unequal Diffusion of Eco-labels: a Spatial Econometric Approach*, I analyze the decision to introduce an eco-labelling scheme through a heteroskedastic Bayesian spatial probit. This framework allows the government's decision to introduce an eco-label program to be influenced by the decision of the neighboring countries. I propose to estimate this model by implementing a new algorithm with higher mixing and converge. Based on a sample, including 141 developing and developed economies, I show that the probability for a country to introduce an eco-labelling scheme depends on the eco-label programs adopted by countries which are spatially close or sharing a strong trade intensity relationship with each other. These results explain why developing countries, which are standards takers, have naturally been put at a disadvantage in terms of eco-labelling adoption. Therefore, eco-label programs should be as transparent as possible and rely on standard harmonization and mutual recognition to remove any potential technical trade barriers and encourage the participation of developing countries.

Keywords: Spatial Econometrics; Generalized Moments Methods; Empirical Likelihood; Bayesian Econometrics; Dynamic Panel Model; Monte-Carlo Simulations.

International Trade; Environmental Policy; Complex Foreign Direct Investment; Pollution Haven; Eco-label.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 4: *Unequal Spatial Diffusion of Eco-label Programs*

LIST OF FIGURES

Introduction

This thesis is an attempt to better understand the role of spatial dependence in the determination of the linkages between foreign direct investments (FDI), pollution emissions and environmental policies at the world wide level.

Motivation

The so-called globalization process has been at the center of several heated debates between environmental and trade communities. The impact of globalization on the environment both within and across countries is a controversial debate among economists and policymakers. Climate Change has further put the attention to energy efficiency and environmental damages. Thus, it is important to understand the mechanisms through which the economy is affecting and affected by environmental degradation.

Emerging and developing countries are likely to face the highest share of the damage costs of environmental impacts in terms of potential agricultural productivity losses, limited water access, poverty and forced migration. While developing economies should engage in green growth strategies through the use of energy efficient technologies and renewable alternative energies, they cannot be expected to make these changes right away without impeding productivity and growth. Therefore, an understanding of the spatial and temporal dynamic linkages between foreign direct investment, trade, pollution emissions and environmental policies is necessary to be able to lay down the foundations for a sustainable economic path in an interdependent world.

Conceptually, the presence of spatial interdependence necessitates a specific modelling and econometric treatment, which can be accommodated through the inclusion of an endogenous spatial autoregressive term, called spatial lag. Yet, most empirical studies fail to explicitly account for the fact that some policy and investment decisions depend on the behavior of other governments and firms, while pollution emissions (air/water) are characterized by spillovers to geographically close countries. If spatial interdependence does matter, but is not accounted for, empirical findings might be biased and invalid. Therefore it is particularly important to overcome the lack of robustness by mitigating the variables omission bias.

Contributions

Beyond the environmental/trade perspective, each empirical essay proposes a new estimator to be applied in a spatial framework. Despite an increasing diffusion of spatial estimators in the past 10 years, a major limitation in spatial econometrics is the lack of available estimators in standard statistical and econometric software packages (e.g. Stata, Eviews, ...). In particular, I show the flexibility of the system GMM estimator and generalized empirical likelihood in order to address the endogeneity of the time as well as spatial dependence and other covariates. In a cross-section setting, I extend the "joint updating" Monte Carlo Markov Chain algorithm to a spatial dependent probit model. This new method displays higher mixing and relies on smaller draws to obtain a stationary chain making it relatively more time efficient. I compiled and implemented most of the new estimators in MATLAB. Although time consuming, the construction of all these routines were a true learning process. They remain available upon request.

Thesis Structure

The structure of this thesis is composed of three parts. The first one is purely econometric and investigates the finite sample properties of spatial dynamic estimators and system GMM in a spatial dynamic panel framework. The second part includes two essays that analyze the pollution haven effect through a spatial Vector Autoregression specification and a bilateral spatial dynamic panel model. Finally, in the third part, I investigate the adoption of eco-labelling schemes in an international trade setting through a Bayesian spatial autoregressive probit model.

The starting point of this thesis is the investigation of the finite sample properties of several estimators for spatial dynamic panel models in the presence of several endogenous variables. So far, none of the available estimators in spatial econometrics allows to consider spatial dynamic models with endogenous variables, beside the time and spatial autoregressive variables. I propose to apply system GMM, since it can correct for the endogeneity of the lagged dependent variable, the spatial lag as well as other potentially endogenous variables using internal and/or external instruments. The results demonstrate that system GMM can be as efficient as standard spatial dynamic estimators (Elhorst, 2005; Yu et al., 2008). In addition, the need to account for additional sources of endogeneity leads to favor system GMM. On a practical ground, unlike spatial dynamic estimators which are computationally cumbersome, system GMM is easy to handle and readily available in most econometric softwares.

The second chapter, "Pollution Havens: a Spatial Vector Autoregression Approach", starts by recognizing that empirical studies designed to test the pollution haven effect have so far shown little and conflicting evidence. One of the crucial problems inherent in existing studies is that, to date, no specific causality analysis of the mutual relationship between FDI, environmental policy and capital/trade openness has been conducted. The most popular approaches consist of either assuming environmental regulation as a function of FDI/trade or vice-versa. Only recently, some authors showed the theoretical and empirical importance of endogenizing environmental stringency with respect to FDI/Trade. The second essay disentangles which relationship(s) is (are) more prevalent by estimating a trivariate spatial panel vector autoregression where an economic shock occurring in a given country can potentially affect the economic conditions in the neighbouring countries. Explicitly accounting for spatial dependence and spillovers is particularly important for two interrelated reasons. First, environmental policies, FDI allocation as well as trade openness have been shown to be spatially correlated. Second, taking into account spatial dependence and spillovers can potentially mitigate the variable omissions bias. This is the first time that the temporal and spatial dynamics through which these three variables interact are carefully analyzed at the world-wide level. From a methodological perspective, this study presents several innovative features. First, the generalized empirical likelihood (GEL) estimator based on the system GMM's moments conditions is applied to a spatial VAR model. Second, because ignoring the possibility of non-stationary parameters can lead to unreliable inference, second generation panel unit root tests, that explicitly account for heterogeneity and cross-section dependence, are implemented. Last but not least, the role of spatial spillovers is highlighted through spatial complements to panel Granger causality and impulse-response functions. Based on 124 countries during the nineties (1993-2000), the results highlight a reverse causality between FDI and environmental policy (proxied by SO_2/GDP emission), confirming the existence of both a pollution haven and a pollution halo effect, for low income countries in particular. In addition, spatial spillovers play a key role in the linkages between FDI, trade openness and SO_2 emissions and highlight the strategic nature of these variables. Overall, these findings suggest that the analysis of the linkages between FDI and the environment should explicitly account for the potential endogeneity and spatial nature of the variables in order to reduce variable omissions.

Based on the results provided by the first two chapters, the third essay, entitled "Complex FDI and Environmental Regulation", reconsiders the analysis of the pollution haven effect by accounting for the spatial and endogenous nature of environmental regulation with respect to the allocation of FDI. Until recently, most multinational enterprises' (MNE) motivations illustrated in the literature relied on a two-country framework, where FDI between home country *i* and host country *j* was only affected by both countries' characteristics. However, new types of FDI have emerged in the last twenty years. Multinationals can allocate FDI in a host country but can also engage in trade or FDI in a third country. These more complex integrated FDI are embedded in a multilateral decision-making process, which means that FDI decisions across various host countries are not spatially independent. Other elements may also lead to interdependent FDI decisions across host countries, including imperfect capital markets and agglomeration externalities (spillovers) which limit the necessary funds a multinational company has to commit abroad. The distinction between types of FDI is also important in the analysis of the pollution haven effect, because these four forms of FDI respond differently to the host and neighboring countries' environmental stringency. In particular, highly integrated vertical FDI is expected to be more sensitive to environmental stringency. On the other hand, horizontal FDI should not be very sensitive to environmental regulation. Thus, the reason the evidence of pollution haven effect is weak (almost inexistent) at the cross-country level could be related to the fact that the most prevalent type of FDI is horizontal, not that the pollution haven effect does not exist. This paper employs a spatial dynamic panel model to estimate the effects of environmental regulations on the allocation of FDI flows among countries taking into account spatial dependence in FDI decision. I examine bilateral FDI flows using an extended OECD investment database which covers a great number of host countries and a long sample period (1981-2000). The findings are largely plausible across specifications and confirm the existence of a pollution haven effect, once I correct for endogeneity and spatial dependence. It is further corroborated by the evidence of prevalence of complex vertical integrated FDI from developed to developing countries.

The last chapter takes a departure from the pollution haven literature to analyze why some governments have adopted multi-sectors (manufactured products) type I eco-label programs and others did not. It also looks at the presence of strategic interactions among partners in their decision to adopt an eco-labelling scheme. Due to increasing competition in domestic and export markets, a country's decision to introduce an eco-label program might be influenced by the behaviour of its neighbouring countries. A government might consider introducing an eco-labelling scheme in order to restrict market access to foreign products or to boost its exports earnings. Whatever the case, its decision will depend on how many and which countries have or are expected to adopt an eco-labelling program. Such interdependence may help explain the very unequal diffusion of eco-label programs among countries. While most industrialized countries have adopted an eco-labelling scheme, African and Latin American countries have yet to decide to implement this type of market-based initiative. This asymmetry has led to trade tensions, as eco-labels are criticized for potentially imposing the environmental concerns of (high income) importing countries on the production methods of (low income) trading partners. This is particularly problematic for developing countries depending heavily on exports to sustain their growth. So far, in spite of these important concerns, most of the literature on eco-labels and its international trade linkages has taken a conceptual or descriptive approach due to a lack of data. As prior studies have not controlled for the spatial dependence in the eco-label decision, their results may also be biased or inconsistent. To tackle the issue of cross-sectional interdependence, I estimate a Bayesian spatial autoregressive probit model by developing a new Markov-Chain-Monte-Carlo algorithm. The empirical evidence confirms the role of the economy's stage of development and innovation capacity in the government's decision to introduce an eco-label. Moreover, results highlight the strategic nature of the eco-labelling decision. In particular, the probability for a country to introduce an eco-labelling scheme depends on the eco-label programs adopted by countries which are spatially close or sharing a strong trade intensity relationship with each other. These results explain why developing countries, which are standards takers, have been naturally put at a disadvantage in terms of eco-labelling adoption.

Chapter 1

Spatial Dynamic Panel Model and System GMM:
A Monte Carlo Investigation[1]

1 Introduction

Although dynamic panel models have drawn a lot of attention in the last decade, the econometric analysis of spatial and dynamic panel models is scarce. So far, none of the available estimators allows to consider a dynamic spatial lag panel model with one or more endogenous variables as explanatory variables, besides the time and spatial lag. Despite this econometric gap, there is an increasing number of empirical studies interested in analyzing the time and spatial patterns of the data in a unified framework.

The spatial dynamic model finds its justification in two main arguments. First, from a theoretical point of view, the spatial lag variable can be viewed as the empirical representation of the reaction function of the theoretical model (e.g. Nash game), while the autoregressive variable captures the dynamic mechanism of the underlying process[2]. Second, from an econometric viewpoint, the inclusion of a spatial and time lag can potentially reduce variables omission, which are spatially and/or temporally correlated. Despite the inclusion of both variables, additional endogeneity in the right hand side might still persist. This endogeneity is usually associated with additional variables omission, measurement error or the existence of a simultaneous relationship between the dependent and some explanatory variables.

Empirically, there are numerous examples where the presence of a dynamic process, spatial dependence and endogeneity might occur. For instance, the empirical test of the New Economic Geography model faces not only spatial and dynamic dependence, but also endogeneity. In the so-called wage equation, the nominal wage rate of the monopolistically competitive sector usually displays high persistency over time and can be determined strategically with respect to the remaining regions or industries. Moreover, the main explanatory variable of the equation, the "market potential" variable, is likely to be endogenous, since the level of income and the price index of each region are partially determined by the wage rate. Recently, empirical work

[1] Written in collaboration with Madina Kukenova, PhD Candidate, University of Lausanne.
[2] The spatial autoregressive term is also referred as endogenous interaction effects in social economics or as interdependence process in political science.

on FDI based on the Capital-Knowledge model is set in a spatial dynamic framework to account for the complementary and/or substitutive nature of FDI (e.g. horizontal, vertical and complex FDI). Additional FDI determinants might also be endogenous, especially in small countries whose FDI are a large source of income.

The inclusion of the lagged dependent variable and additional endogenous variables in a spatial lag model invalidate the use of traditional spatial maximum likelihood estimators. In fact, applying standard spatial likelihood estimators in this framework yields inconsistent and biased estimates, because they only account for one or two source(s) of endogeneity, namely the spatial and the time lag. To overcome the absence of a proper spatial estimator, several empirical studies (Kukenova and Monteiro, 2009; Madariaga and Poncet, 2007) apply the system generalized method of moments (GMM) estimator, developed by Arellano and Bover(1995) and Blundell and Bond (1998)[3]. The main reason behind the use of system GMM to estimate a spatial dynamic panel model is its ability to correct for the endogeneity introduced by the spatial lag variable and other potentially endogenous regressors. In addition, system GMM is robust to some econometrics issues such as measurement error and weak instruments. It can also control for arbitrary heteroskedasticity and autocorrelation in the disturbance error. Moreover its implementation is computationally tractable as it avoids the inversion of high dimension spatial weights matrix *W* and the computation of its eigenvalues.

Going beyond this intuitive motivation, this paper addresses several issues through an extensive Monte Carlo study. First, I show that system GMM estimates consistently the spatial lag parameter. Second, I investigate how the presence of additional endogenous variables influence the performance of system GMM. Third, I compare the relative performance of recently available spatial dynamic estimators with respect to system GMM in the presence of an additional endogenous variable. In particular, I compare the performance of spatial MLE (SMLE), Spatial Dynamic MLE (SDMLE) (Elhorst, 2005) and Spatial Dynamic QMLE (SDQMLE) (Yu et al, 2008). To our knowledge, this paper is the first to investigate the consequences of an additional endogenous explanatory variable in a spatial dynamic framework and compare the finite sample properties of the current available spatial dynamic estimators in a thorough analysis. Overall, system GMM displays good finite sample properties. Despite the fact that system GMM displays higher dispersion in the parameter estimation, it is less biased. As the

[3] System GMM is also known as SYS-GMM or extended GMM (Binder, 2005).

sample size increases, system GMM outperforms any spatial estimators considered in this study. These findings suggest that it might be appropriate to apply system GMM when the spatial dynamic panel model includes several endogenous variables and the sample size (N or T) is relatively large. The remaining of the paper is structured as follows. Section 2 presents the dynamic spatial lag model and reviews the available methods to estimate it, including system GMM. The Monte Carlo investigation is described and performed in section 3. Section 4 checks the robustness of the main results. Section 5 brings the paper to a conclusion.

2 Spatial Dynamic Panel Model

The increasing popularity of spatial models in empirical studies is intimately linked to the recent progress in spatial econometrics theory. The basic spatial model was suggested by Cliff and Ord (1981), but it did not receive important theoretical extensions until the middle of the 1990s. Anselin (2001, 2006), Elhorst (2003) and Lee and Yu (2009) provide thorough surveys of the different spatial models and suggest econometric strategies to estimate them.

2.1 Dynamic Spatial Lag Model

A general spatial dynamic panel model, also known as a spatial dynamic autoregressive model with spatial autoregressive error (SARAR(1,1)), can be described as follows:

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + \rho W_{1t} Y_t + Z_t \beta + X_t \gamma + \varepsilon_t \\ \varepsilon_t &= \eta + \phi W_{2t} \varepsilon_t + v_t,\ t = 1,\ldots,T \end{aligned} \tag{1}$$

where Y_t is a $N \times 1$ vector, W_{1t} and W_{2t} are $N \times N$ row-standardized spatial weight matrices which are non-stochastic and exogenous to the model, η is the vector of individual effects, Z_t is a $N \times p$ matrix of p exogenous explanatory variables ($p \geq 0$) and X_t is a $N \times q$ matrix of q endogenous and predetermined explanatory variables with respect to Y_t $(q \geq 0)$. Finally, v_t is assumed to be distributed as $(\mathbf{0}, \Omega)$. By adding some restrictions to the parameters, two popular spatial model specifications can be derived from this general spatial model, namely the dynamic spatial lag model ($\phi = 0$) and the dynamic spatial error model $(\rho = 0)$[4].

[4] The analysis of the spatial error panel model and the spatial lag with spatial error model is beyond the scope of this paper. For further details, see Elhorst (2005), Kapoor et al. (2007), as well as Lee and Yu (2009). Recently, Jacobs et al. (2009) estimate a spatial dynamic model with a spatial error term using system GMM. They show that the impact of ignoring the spatial error term is marginal. This confirms that the presence of spatial dependence in the error term does not alter the consistence of the remaining parameters.

The spatial lag model accounts directly for relationships between dependent variables that are believed to be related in some spatial way. Somewhat analogous to a lagged dependent variable in time series analysis, the estimated spatial lag coefficient characterizes the contemporaneous correlation between one cross-section and other geographically-proximate cross-sections. The following equation gives the basic spatial dynamic panel specification:

$$Y_t = \alpha Y_{t-1} + \rho W_t Y_t + Z_t \beta + X_t \gamma + \eta + v_t \tag{2}$$

Model (2) is also known as the "time-space simultaneous" model according to Anselin (1988, 2001) typology[5]. The spatial autoregressive coefficient (ρ) associated with $W_t Y_t$ represents the effect of the weighted average of the neighborhood, i.e. $\left[W_t Y_t\right]_i = \Sigma_{j=1 \ldots N_t} w_t\left(d_{ij}\right) \cdot Y_{jt}$, $w_t\left(d_{ij}\right)$ being the weight between location *i and j*. The spatial lag term allows to determine if the dependent variable Y_t is (positively/negatively) affected by the Y_t from other close locations weighted by a given criterion (usually distance or contiguity). In other words, the spatial lag coefficient captures the impact of Y_t from neighborhood locations. Let $\omega_{\min}$ and $\omega_{\max}$ be the smallest and highest characteristic root of the spatial matrix W, then this spatial effect is assumed to lie between $1/\omega_{\min}$ and $1/\omega_{\max}$. Most of the spatial econometrics literature constrains the spatial lag to lie between -1 and +1. However, this might be restrictive, because if the spatial matrix is row-normalized, then the highest characteristic root is equal to unity (i.e. $\omega_{\max} = 1$), but the smallest eigenvalue can be bigger than -1, which would lead the lower bound to be smaller than -1.

Given that expression (2) is a combination of a time and spatial autoregressive models, I need to ensure that the resulting process is stationary. This is particularly important for the spatial estimators, since Binder et al. (2005) showed that system GMM does not break down in the presence of a unit root. Note that the stationarity restrictions in this type of model are stronger than the individual restrictions imposed on the coefficients of a spatial or dynamic model. The process is covariance stationary if $\left|\left(I_N - \rho W_t\right)^{-1} \alpha\right| < 1$, or, equivalently, if[6]

$$\begin{aligned} |\alpha| &< 1 - \rho \omega_{\max} \quad \text{if } \rho \geq 0 \\ |\alpha| &< 1 - \rho \omega_{\min} \quad \text{if } \rho < 0 \end{aligned}$$

[5] Besides the "time-space simulatenous" model, Anselin distinguishes three other spatial lag panel models: the "pure space recursive" model which only includes a lagged spatial lag coefficient; the "time-space recursive" specification which considers a lagged dependent variable as well as a lagged spatial lag (Korniotis, 2007); and the "time-space dynamic" model, which includes a time lag, a spatial lag and a lagged spatial lag (Yu et al., 2008).

[6] According to Kelejian and Prucha (1999), one necessary stationary condition in space is that rows and columns sums of the spatial weight matrix are uniformly bounded in absolute values (as $n \to \infty$).

From an econometric viewpoint, equation (2) faces simultaneity and endogeneity problems, which in turn means that OLS estimation will be biased and inconsistent (Anselin, 1988). To see this point more formally, note that the reduced form of equation (2) is given by the following expression:

$$Y_t = (I_N - \rho W_t)^{-1}(\alpha Y_{t-1} + Z_t \beta + X_t \gamma + \eta + v_t)$$

Each element of Y_t is a linear combination of all of the error terms. Moreover, as pointed out by Anselin (2003), assuming $|\rho| < 1$ and each element of W_t is smaller than one imply that $(I_N - \rho W_t)^{-1}$ can be reformulated as a Leontief expansion $(I_N - \rho W_t)^{-1} = I + \rho W_t + \rho^2 W_t^2 + ...$ Accordingly, the spatial lag model features two types of global spillovers effects: a multiplier effect for the predictor variables as well as a diffusion effect for the error process. Since the spatial lag term $W_t Y_t$ is correlated with the disturbances, even if v_t are independently and identically distributed, it must be treated as endogenous and a proper estimation method must account for this endogeneity.

Despite the fact that dynamic panel models have been the object of recent important developments (Phillips and Moon, 2000), econometric analysis of spatial dynamic panel models is almost inexistent. In fact, the spatial dynamic panel model without additional endogenous variables was first formulated by Anselin in 1988. Yet there was no available estimator until the mid 2000's. Currently, there is only a limited number of available estimators that deal with spatial and time dependence in a panel setting. Table 1 summarizes the different estimators proposed in the literature.

In the absence of spatial dependence, there are three main types of estimators available to estimate a dynamic panel model. The first type of estimators consists of estimating an unconditional likelihood function (Hsiao et al., 2002). The second type of procedure corrects the bias associated with the least square dummy variables (LSDV) estimator (Bun and Carree, 2005). The last type, which is the most popular, relies on GMM estimators, like difference GMM (Arellano and Bond, 1992), system GMM (Arellano and Bover, 1995), Blundell and Bond, 1998) or continuously updated GMM (Hansen et al., 1996).

Table 1: Spatial Dynamic Estimators Survey

Model	Estimation Methods	Endogenous
$Y_t = \alpha Y_{t-1} + \beta Z_t + \varepsilon_t$	GMM (Arellano et al. 1991, 1995; Blundell et al., 1998; Hansen et al., 1996) MLE/Minimum Distance (Hsiao, Pesaran & Tahmiscioglu, 2002) CLSDV (Kiviet, 1995; Hahn and Kuersteiner, 2002; Bun and Carree, 2005)	Y_{t-1}
$Y_t = \alpha Y_{t-1} + \beta Z_t + \gamma X_t + \varepsilon_t$	GMM (Arellano et al., 1991, 1995; Blundell et al., 1998; Hansen et al., 1996)	$Y_{t-1};\ X_t$
$Y_t = \alpha Y_{t-1} + \rho W Y_{t-1} + \beta Z_t + \gamma X_t + \varepsilon_t$	LSDV-IV (Korniotis, 2010)	$Y_{t-1};\ X_t$
$Y_t = \rho W Y_t + \beta Z_t + \varepsilon_t$	Spatial-MLE / Spatial 2SLS (Anselin, 1988, 2001; Elhorst, 2003) Minimum Distance (Azomahou, 2008)	WY_t
$Y_t = \rho W Y_t + \beta Z_t + \gamma X_t + \varepsilon_t$	Spatial 2SLS (Dall'erba and Le Gallo, 2007)	$WY_t;\ X_t$
$Y_t = \alpha Y_{t-1} + \rho W Y_t + \beta Z_t + \varepsilon_t$	Spatial Dynamic MLE (Elhorst, 2003, 2005, 2008) Spatial Dynamic QMLE (Yu, de Jong and Lee, 2007, 2008; Lee and Yu 2007) C2SLSDV (Beenstock and Felsenstein, 2007) Spatial MLE-GMM / Spatial MLE-Spatial Dynamic MLE (Elhorst, 2008)	$WY_t;\ Y_{t-1}$
$Y_t = \alpha Y_{t-1} + \rho W Y_t + \beta Z_t + \gamma X_t + \varepsilon_t$	System-GMM (Arellano and Bover, 1995; Blundell and Bond, 1998)	$WY_t;\ Y_{t-1};\ X_t$

Assuming all explanatory variables are exogenous beside the spatial autoregressive term, the spatial lag panel model without any time dynamic is usually estimated using spatial maximum likelihood (Elhorst, 2003) or spatial two-stage least squares methods (S2SLS) (Anselin, 1988, 2001). The ML approach consists of estimating the spatial coefficient by maximizing the non-linear reduced form of the spatial lag model. The spatial 2SLS uses the exogenous variables and their spatially weighted averages (Z_t , $W_t \cdot Z_t$) as instruments[7]. In a Monte Carlo study, Franzese and Hays (2007) show that S2SLS is less biased than SMLE but suffers from poor efficiency in small samples. When the number of cross-sections is larger than the period sample, Anselin (1988) suggests to estimate the model using MLE, 2SLS or 3SLS in a spatial seemingly unrelated regression (SUR) framework. More recently, Azomahou (2008) proposes to estimate a spatial panel autoregressive model with random effects applying a two-step minimum distance estimator. In the presence of endogenous variables but no lagged dependent variable, Dall'erba and Le Gallo (2007) suggest applying spatial 2SLS with lower orders of the spatially weighted sum of the exogenous variables as instrument for the spatial autoregressive term and the endogenous variables.

In a dynamic context, the estimation of spatial lag panel models is usually based on a ML function. Elhorst (2003, 2005) proposes to estimate the unconditional loglikelihood function of the reduced form of the model in first-difference. While the absence of explanatory variables besides the time and spatial lags leads to an exact likelihood function, this is no longer the case when additional regressors are included. Moreover, when the sample size T is relatively small the initial observations contribute greatly to the overall likelihood. That is why the pre-sample values of the explanatory variables and likelihood function are approximated using the Bhargava and Sargan approximation or the Nerlove and Balestra approximation. More recently, Yu et al. (2008) provide a theoretical analysis on the asymptotic properties of the quasi-maximum likelihood (Spatial Dynamic QML), which relies on the maximization of the concentrated likelihood function of the demeaned model. They show that the limit distribution is not centered around zero and propose a bias-corrected estimator[8]. Beside the fact that Yu et al. (2008) do not assume normality, the main difference with Elhorst's ML estimator lies in the asymptotic structure. Elhorst considers fixed T and large N $(N \to \infty)$, while Yu et al. assume

[7] In a cross-section setting, Kelejian and Prucha (1998) suggest also additional instruments ($W_t^2 Z_t$, $W_t^3 W_t$, …). Lee (2003) shows that the estimator proposed by Kelejian and Prucha is not an asymptotically optimal estimator and suggests a three-steps procedure with an alternative instrument for the spatial autoregressive coefficient in the last step: $W_t(Z_t - \tilde{\rho} W_t)^{-1} Z_t \tilde{\beta}$, where $\tilde{\rho}$ and $\tilde{\beta}$ are estimates obtained using the S2SLS proposed by Kelejian and Prucha, 1998.

[8] In two other related working papers, Lee and Yu (2007) and Yu et al.(2007) investigate the presence of non-stationarity and time fixed effects, respectively, in a spatial dynamic panel framework.

large N and T $(N \to \infty; T \to \infty)$. Consequently, the way the individual effects are taken out differs. Elhorst considers first-difference variables, while Yu et al. demean the variables. Assuming large T avoids the problem associated with initial values and the use of approximation procedures. Finally, Yu et al's approach allows to recover the estimated individual effects, which is not the case with the estimator proposed by Elhorst. In a recent paper, Elhorst (2008) analyzes the finite sample performance of three estimators for a spatial dynamic panel model with one exogenous variable: Spatial MLE (referred as CLSDV by Elhorst), Spatial Dynamic MLE and difference GMM. He finds that Spatial Dynamic MLE has the better overall performance in terms of bias reduction and root mean squared errors (RMSE), although the Spatial MLE presents the smallest bias for the spatial autoregressive coefficient. Based on these results, Elhorst proposes two mixed estimators, where the spatial lag dependent variable is based on the spatial ML estimator and the remaining parameters are estimated using either difference GMM or Spatial Dynamic ML conditional on the spatial ML's estimate of the spatial autoregressive coefficient. The mixed Spatial MLE/Spatial Dynamic MLE estimator shows superior performance in terms of bias reduction and RMSE in comparison with mixed Spatial MLE/difference GMM. However, the latter can be justified on a practical ground if the number of cross-sections in the panel is large, since the time needed to compute Spatial MLE/Spatial Dynamic MLE can be substantial. In a spatial vector autoregression (spVAR) setting, Beenstock and Felsenstein (2007) suggest a spatial 2SLS procedure where the bias of the lagged dependent variable is corrected using Hsiao (1986)'s asymptotic bias expression.

If one is willing to consider some explanatory variables as potentially endogenous in a dynamic spatial panel setting, then no spatial estimator is currently available. In many empirical applications, endogeneity can arise from measurement errors, variables omission or the presence of simultaneous relationship(s) between the dependent and the covariate(s). The main drawback of applying SMLE, SDMLE or SDQMLE is that, while the spatial autoregressive coefficient is considered endogenous, no instrumental treatment is applied to other potential endogenous variables. This leads to biased estimates, which would invalidate empirical results. To address this issue, I propose system-GMM.

2.2 System GMM

Empirical papers dealing with a spatial dynamic panel model with several endogenous variables usually apply system GMM (see for example Kukenova and Monteiro, 2008; Madriaga and Poncet, 2007). Back in 1978, Haining proposed to instrument a first order spatial autoregressive model using lagged dependent variables. Since it does not use efficiently all the available information, this method is not efficient in a cross-section setting (Anselin, 1988). This is no longer the case in a panel framework. In line with Haining's suggestion, GMM instruments the spatial lag using lagged values of the dependent, spatial autoregressive and exogenous variables[9]. For simplicity, equation (2) is reformulated for a given cross-section i ($i = 1,..,N$) at time t ($t = 1,..,T$):

$$Y_{it} = \alpha Y_{it-1} + \rho [W_t Y_t]_i + Z_{it}\beta + X_{it}\gamma + \eta_i + v_{it} \tag{3}$$

According to Arellano and Bond (1991), the first step in difference GMM consists of eliminating the individual effects (η_i) correlated with the covariates and the lagged dependent variable, by rewriting equation (3) in first difference for individual i at time t:

$$\Delta Y_{it} = \alpha \Delta Y_{it-1} + \rho \Delta [W_t Y_t]_i + \Delta Z_{it}\beta + \Delta X_{it}\gamma + \Delta v_{it} \tag{4}$$

Even if the fixed effects (within) estimator cancels the individual fixed effects, the lagged endogenous variable (ΔY_{it-1}) is still correlated with the idiosyncratic error terms (v_{it}). Nickell (1981) as well as Anderson and Hsiao (1981) show that the within estimator is only consistent for large T. Given that this condition is usually not satisfied in practice, the estimation using first-difference remains also biased and inconsistent. To overcome this issue, Arellano and Bond (1991) propose the following moment conditions associated with equation (4):

$$E(Y_{i,t-\tau} \Delta v_{it}) = 0; \text{ for } t = 3,...,T \text{ and } 2 \le \tau \le t-1 \tag{5}$$

If the strict exogeneity assumption of the covariates (Z_{it}) has not been verified, the estimation based only on these moment conditions (5) is insufficient. The explanatory variables constitute valid instruments to improve the estimator's efficiency only when the strict exogeneity assumption is satisfied:

$$E(Z_{i\tau} \Delta v_{it}) = 0; \text{ for } t = 3,...,T \text{ and } 1 \le \tau \le T \tag{6}$$

However, the GMM estimator based on the moment conditions (5) and (6) can still be inconsistent when $\tau < 2$ and in presence of reverse causality, i.e. $E(Z_{it} v_{it}) \neq 0$. In order to overcome this problem, one can assume that the covariates are weakly exogenous for $\tau < t$, which means that the moment conditions (6) can be rewritten as:

[9] Badinger et al. (2004) recommend to apply system GMM, once the data has been spatially filtered. This approach can be considered only when spatial dependence is viewed as a nuisance parameter.

$$E\left(Z_{i,t-\tau}\Delta v_{it}\right)=0;\ \text{for}\ t=3,\ldots,T\ \text{ and } 1\leq\tau\leq t-1 \qquad (7)$$

For the different endogenous variables, including the spatial lag, the valid moment conditions are

$$E\left(X_{i,t-\tau}\Delta v_{it}\right)=0;\ \text{for}\ t=3\ldots T\ \text{ and } 2\leq\tau\leq t-1 \qquad (8)$$

$$E\left(\left[W_{t-\tau}Y_{t-\tau}\right]_i\Delta v_{it}\right)=0;\ \text{for}\ t=3\ldots T\ \text{ and } 2\leq\tau\leq t-1 \qquad (9)$$

In small samples, the GMM estimator based on these moment conditions can still yield biased coefficients. Blundell and Bond (1998) show that the imprecision of this estimator increases when the individual effects are important and the variables are persistent over time making the lagged levels of the endogenous variables weak instruments. To overcome those problems, the authors propose system GMM, which estimate simultaneously equation (3) and equation (4). The additional moment conditions for system GMM are based on the model specified in level:

$$E\left(\Delta Y_{i,t-1}v_{it}\right)=0;\ \text{for}\ t=3,\ldots,T \qquad (10)$$

$$E\left(\Delta Z_{it}v_{it}\right)=0;\ \text{for}\ t=2,\ldots,T \qquad (11)$$

$$E\left(\Delta X_{it-1}v_{it}\right)=0;\ \text{for}\ t=3,\ldots,T \qquad (12)$$

$$E\left(\Delta\left[W_{t-1}Y_{t-1}\right]_i v_{it}\right)=0;\ \text{for}\ t=3,\ldots,T \qquad (13)$$

The consistency of the system GMM estimator relies on the validity of these moment conditions, which depends on the assumption of absence of serial correlation of the level residuals and the exogeneity of the explanatory variables. Therefore, it is necessary to apply specification tests to ensure that these assumptions hold. Arellano and Bond suggest two specification tests in order to verify the consistency of the GMM estimator. First, the overall validity of the moment conditions is checked by the Sargan/Hansen test. Second, the Arellano-Bond test examines the serial correlation property of the level residuals.

More generally, one should keep in mind that the estimation of the spatial autoregressive coefficient via system GMM, although "potentially" consistent, is usually not the most efficient one. Efficiency relies on the "proper" choice of instruments, which is not an easy task to implement. Another issue lies in the fact that the instrument count grows as the sample size T rises. A large number of instruments can overfit endogenous variables (i.e. fail to correct for endogeneity) and leads to inaccurate estimation of the optimal weight matrix, downward biased two-step standard errors and wrong inference in the Hansen test of instruments validity. Okui (2009) demonstrates that the bias of system GMM does not result from the total number of instruments, but from the number of instruments for each equation. As pointed out by Roodman (2009), it is advisable to restrict the number of instruments by defining a maximum number of lags and/or by collapsing the instruments[10]. *Collapsing* means combining instruments through addition into subsets. For instance, if the instruments are collapsed, the moment conditions (5) become:

$$E\left(Y_{i,t-\tau}\Delta v_{it}\right)=0;\ \text{for } 2\leq\tau\leq t-1 \tag{14}$$

These modified moment conditions still impose the orthogonality of $Y_{i,t-\tau}$ and Δv_{it}, but rather to hold for each t and τ, it is only valid for each τ. Roodman (2009) shows that collapsed instruments lead to less biased estimates, although the associated standard errors tend to increase. GMM results in this paper are based on collapsed instruments[11]. This allows us to use the same number of instruments independently of the sample size.

3 A Monte-Carlo Study

In this section, I investigate the finite sample properties of several estimators including Spatial MLE, Spatial Dynamic MLE and Spatial Dynamic QMLE, LSDV, difference GMM and system GMM in the presence of an endogenous variable in a spatial dynamic panel data context using Monte-Carlo simulations[12]. Following Blundell et al. (2000), the data generating process (DGP) is defined as follows:

$$Y_{it}=\alpha Y_{i,t-1}+\rho\left[WY_t\right]_i+\beta Z_{it}+\gamma X_{it}+\eta_i+v_{it} \tag{15}$$

$$Z_{it}=\delta Z_{i,t-1}+u_{it} \tag{16}$$

$$X_{it}=\lambda X_{i,t-1}+\psi\eta_i+\theta v_{it}+e_{it} \tag{17}$$

with $\eta_i\sim N\left(0,\sigma_\eta^2\right)$; $v_{it}\sim N\left(0,\sigma_v^2\right)$; $u_{it}\sim N\left(0,\sigma_u^2\right)$; $e_{it}\sim N\left(0,\sigma_e^2\right)$.

[10] This approach has been adopted in several empirical papers, including Beck and Levine (2004).

[11] See Appendix 6.1 for further details on extended GMM and Appendix 6.2 for spatial ML estimators.

[12] All simulations are performed using Matlab R2008b. The estimation of a spatial dynamic panel model with system-GMM can also be done in Stata (with the *xtabond*2 command), Gauss (with the *DPD*98 routine) or Eviews (with *gmm*() and *@dyn* commands).

Note that the variable X_{it} is predetermined and not endogenous *per se*. Yet, GMM estimators correct for the presence of predetermined and endogenous variables through the same type of moment conditions. In order to avoid results being influenced by initial observations, the covariates Y_{i0}, Z_{i0} and X_{i0} are set to 0 for all i and each variable is generated $(100+T)$ times according to their respective DGP. The first 100 observations are then discarded. Note that the dependent variable is generated according to the reduced form of equation (15):

$$Y_{it} = \left(1-\rho[W]_i\right)^{-1}\left[\alpha Y_{i,t-1} + \beta Z_{it} + \gamma X_{it} + \eta_i + v_{it}\right] \tag{18}$$

Following Kapoor et al. (2007) and Kelejian and Prucha (1999), I consider four different types of spatial weight matrix. In each case, the matrix is row-standardized so that elements in each row sums to one. The matrices considered are characterized by different degree of sparseness. More precisely, the first three "theoretical" spatial matrices rely on a perfect "idealized" circular world, referred as Rook type of order 1, 3 and 5, so that each location is related to the one/three/five locations immediately before and after it and each nonzero elements are equal to $0.5\ /\ 0.\bar{3}\ /\ 0.1$, respectively[13]. In addition, as a robustness check, I consider real distance data between capitals among 224 countries[14]. In order to avoid giving some positive weight to very remote countries, I consider the negative exponential weighting scheme. This is done by dividing the distance between locations *J* and *K* by the minimum distance within the region r (where location j lies within region r): $w(d_{j,k}) = \exp(-d_{j,k}/MIN_{r,j})$ if $j \neq k$.

The Monte Carlo experiments rely on the following designs:

$$T \in \{5,10,20,30,40\}; \quad N \in \{20;\ 30;\ 50;\ 70\}; \quad \alpha \in \{0.2;\ 0.4;\ 0.5;\ 0.7\}; \quad \rho \in \{0.1;\ 0.3;\ 0.5;\ 0.7\};$$
$$\beta = 1;\ \delta = 0.65;\ \gamma = 0.5;\ \lambda = 0.45;\ \psi = 0.25;\ \theta = 0.6;\ \sigma_u^2 = 0.05;\ \sigma_v^2 = 0.05;\ \sigma_e^2 = 0.05;\ \sigma_\eta^2 = 0.08$$

In order to ensure stationarity, I consider the designs that satisfy the restrictions $|\alpha| < 1-\rho\omega_{\max}$ if $\rho \geq 0$. The values of the spatial lag are restricted to be positive, as most empirical studies highlight a positive spatial autoregressive parameter. This stationarity restriction reduces the total number of designs to 200. For each of these designs, I performed 1000 trials with the same initial conditions and spatial weight matrix.

[13] For example, the rook-type matrix of order 1, also known as "one ahead, one behind" is such that its i^{th} row, $1 \leq i \leq N$, has nonzero elements in positions $i-1$ and $i+1$. Since the matrix is defined in a circular world, the nonzero elements in rows 1 and *N* are in positions $(1,2),(1,N)$ and $(N-1,N),(N,1)$, respectively. The other two matrices are defined in a similar way. Note that these matrices have the property of remaining symmetric even after the row-standardization, which reduces significantly the computation time of Spatial Dynamic MLE and Spatial Dynamic QMLE.

[14] The data is taken from CEPII database.

As a measure of consistency, I consider the root mean square error which is defined as the square root of the sum of the variance and the squared bias of the estimator. As suggested by Kelejian and Prucha (1999), I also consider the approximated RMSE, which converges to the standard RMSE under a normal distribution. Approximated RMSE is given by $\sqrt{b^2+\left(IQ^2/1.35\right)^2}$, where b is the difference between the true value of the coefficient and the median of the estimated coefficients, and IQ is the difference between the 75% and 25% quantile. This definition has the advantage of being more robust to outliers that may be generated by the Monte-Carlo simulations.

Based on the Rook type of order 1 spatial weight matrix, the Monte Carlo investigation highlights several important facts[15]. First, the results in terms of bias and efficiency depend on the values assigned to the spatial and time lag parameters. Second, the computation time is always larger with any spatial ML estimators than with system GMM. Since the results based on RMSE and approximated RMSE are qualitatively similar, I only present the results associated with standard RMSE. The Monte Carlo analysis is presented in two parts. First, I check if GMM are appropriate estimators in a spatial dynamic panel framework. Second, I compare the relative performance of system GMM with respect to the spatial estimators.

3.1 System GMM vs. Difference GMM

In this section, I study the finite sample properties of difference and system GMM. For illustrative purpose, I compare the performance of GMM estimators with respect to LSDV estimator. In order to have the same number of instruments independently of the sample size, I consider two-step GMM based on a collapsed instruments structure[16]. To avoid imprecise estimate of the optimal weight matrix (when $T > N$), I also restrict the lags to two and three. In other words, each endogenous variables (i.e. Y_{t-1}, WY_t, X_t) is instrumented with their 2nd and 3rd lags values and the exogenous variable X_t.

In order to assess the global consistency and efficiency of GMM estimators: the bias and RMSE results are averaged over the whole range of parameters. As Figure 1 and 2 indicate, system and difference GMM outperform the fixed effect estimator in terms of bias and efficiency. In fact,

[15] The full results are given in Appendix 6.3, 6.4 and 6.5.

[16] Different types of instruments structure have been considered (no lag limit, no *collapse* option, ...) as well as different initial weight of the moment conditions matrix used in the first-step estimation. All results are available upon request.

the LSDV estimator has a negative bias which does not decrease with the size of the panel, invalidating the use of LSDV in this type of model. Consequently, spatial dynamic panel models should definitively not be estimated with LSDV. The differences in terms of performance between difference and system GMM are relatively marginal. Hayakawa (2007) shows that difference and level GMM estimators can be more biased than system GMM, because the bias of system GMM is the sum of two elements. The first one is the weighted sum of the bias associated with difference and level GMM, while the second one originates from using the level and difference estimators jointly. Since the bias of difference and level GMM evolve in opposite directions, the first component of the bias of system GMM will tend to be small due to a partial cancelling out effect. In addition, the weight of the first element plays also a role in the difference of the magnitude of the biases. The latter could explain why both GMM estimators perform almost equivalently. However, as depicted in Figure 2, system GMM displays smaller RMSE in small samples, that is why, I decide to focus on system GMM.

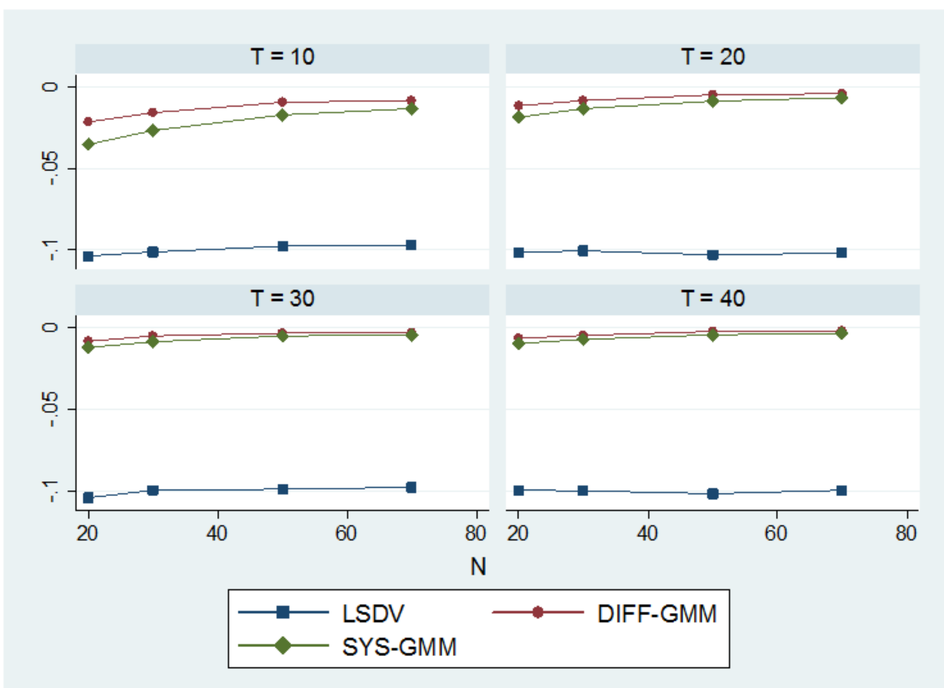


Figure 1: LSDV *vs*. GMM Bias

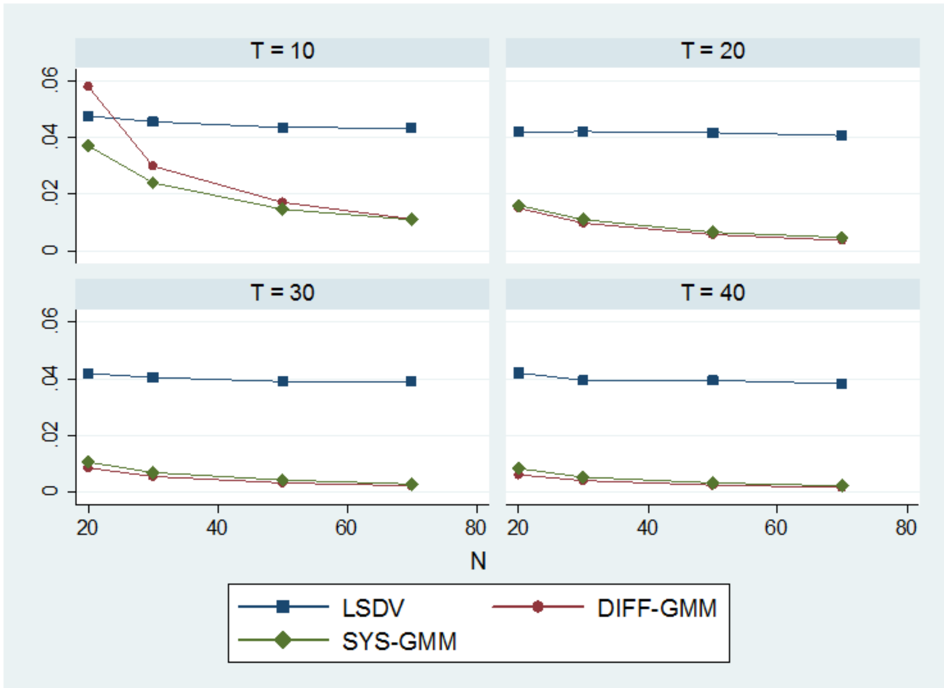


Figure 2: LSDV *vs*. GMM RMSE

3.2 System GMM vs. Spatial Estimators

In this section, I compare the performance of system GMM with respect to spatial ML estimators by considering the average bias and RMSE for the different values of the autoregressive and spatial autoregressive coefficient[17]. For each parameter, Figure 3 plots the averaged bias of the respective estimators over a range of the cross-sectional dimension, N. The upper panels of Figure 3 display the average bias associated with the autoregressive and spatial autoregressive variables respectively, while the lower panels present the average bias for the endogenous and exogenous variables. Although the bias associated with the spatial estimators seems to be independent of the cross-section dimension, it does decrease as the time dimension increases. This corroborates Yu et al.'s findings, who show that spatial dynamic QMLE remains slightly biased as the number of cross-section increases (see Table 2 and 3 in Yu et al., 2008). Unlike spatial maximum likelihood estimators, system GMM's bias decreases as N and/or T increase.

[17] Note that the spatial estimators rely on a numerical optimization algorithm which can lead to a non-convergent solution. When convergence is not obtained, the trial is dropped from the analysis. See Appendix 6.2 for further details about the implementation of the spatial estimators.

According to the plot on the top left of Figure 3, all estimators tend to underestimate the autoregressive parameter (e.g. positive bias) with system GMM exhibiting the smallest bias. The opposite happens for the spatial autoregressive parameter. Note that the spatial estimators tend to dominate system GMM in relatively small sample. This is not surprising since spatial estimators explicitly account for the spatial structure of the data by estimating the reduced form of the model. However, as the time dimension increases, system GMM tends to outperform the spatial estimators. Even though all estimators overestimate the exogenous variable, system GMM display smaller bias. Most importantly, the plot on the bottom left of Figure 3 shows how important it is to correct for endogeneity[18]. In fact, if not corrected, the bias associated with the endogenous variable can represent more than 40% of the true value of the parameter. Moreover, the magnitude of the bias of the endogenous covariate does not seem to depend on the sample dimension (*N* and *T*). This finding suggests that estimating a spatial dynamic panel model with endogenous variables using traditional spatial estimators would not be advisable. Besides system GMM, spatial dynamic QMLE displays the lowest bias for all coefficients.

Figure 4 depicts the performance of the estimators in terms of RMSE. The results are slightly different from the ones obtained for the bias. Despite the fact that spatial ML estimators yield larger bias, in general, they tend to render less disperse estimates than system GMM. This is confirmed by Figure 5, which plots the histogram distribution for each estimated parameter based on 1000 trials for $N = 70$ and $T = 20$. However, as the panel size increases, the performance of GMM converges to the level of RMSE of the spatial ML estimators.

While the spatial estimators reach their steady-state level of RMSE in relatively small samples (making the slope of RMSE and bias almost flat), the RMSE of system GMM decreases as the sample size increases (the slope of RMSE and bias being more steeper). Interestingly, for moderate size samples, the estimation of the time lag and exogenous variables with system GMM yields a lower RMSE than any spatial ML estimators. This result might seems to be paradoxical with the accepted notion that ML is more efficient than GMM, but it is not. Actually, this finding is an extension of Das et al.'s (2003) conclusion to the panel framework. The main reason for this seemingly contradictory result is that the spatial ML approach requires the estimation of more parameters than does system GMM. All spatial ML estimators involve the

[18] I run simulations using different values for the correlation parameter (θ) of the endogenous variable and the error term of the dependent variable (v_{it}). Results remain qualitatively similar. The main difference is that depending on the sign of the parameter of the endogenous variable will be overestimated ($\theta > 0$) or underestimated ($\theta < 0$).

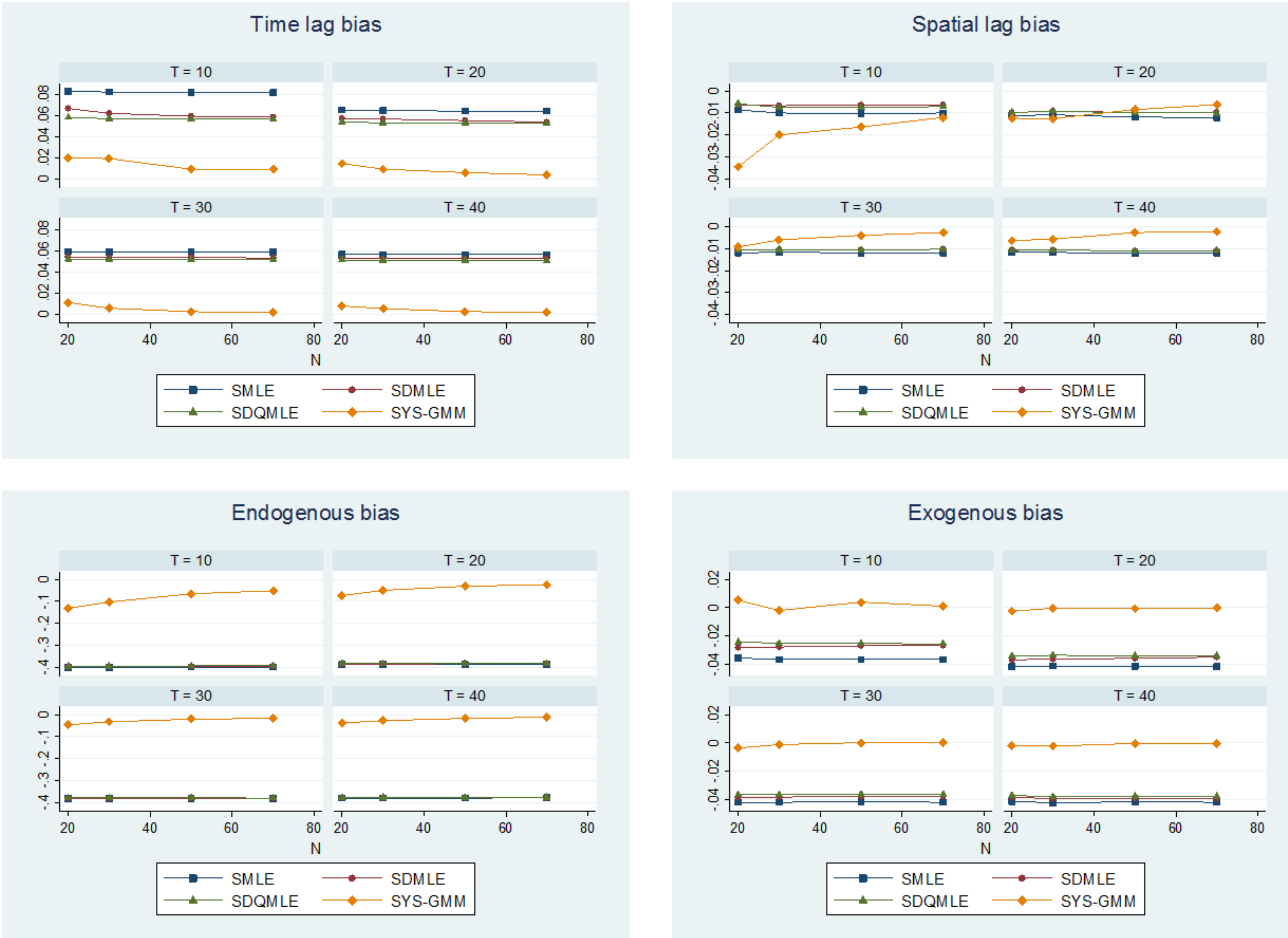


Figure 3: System GMM *vs.* Spatial Estimators Bias

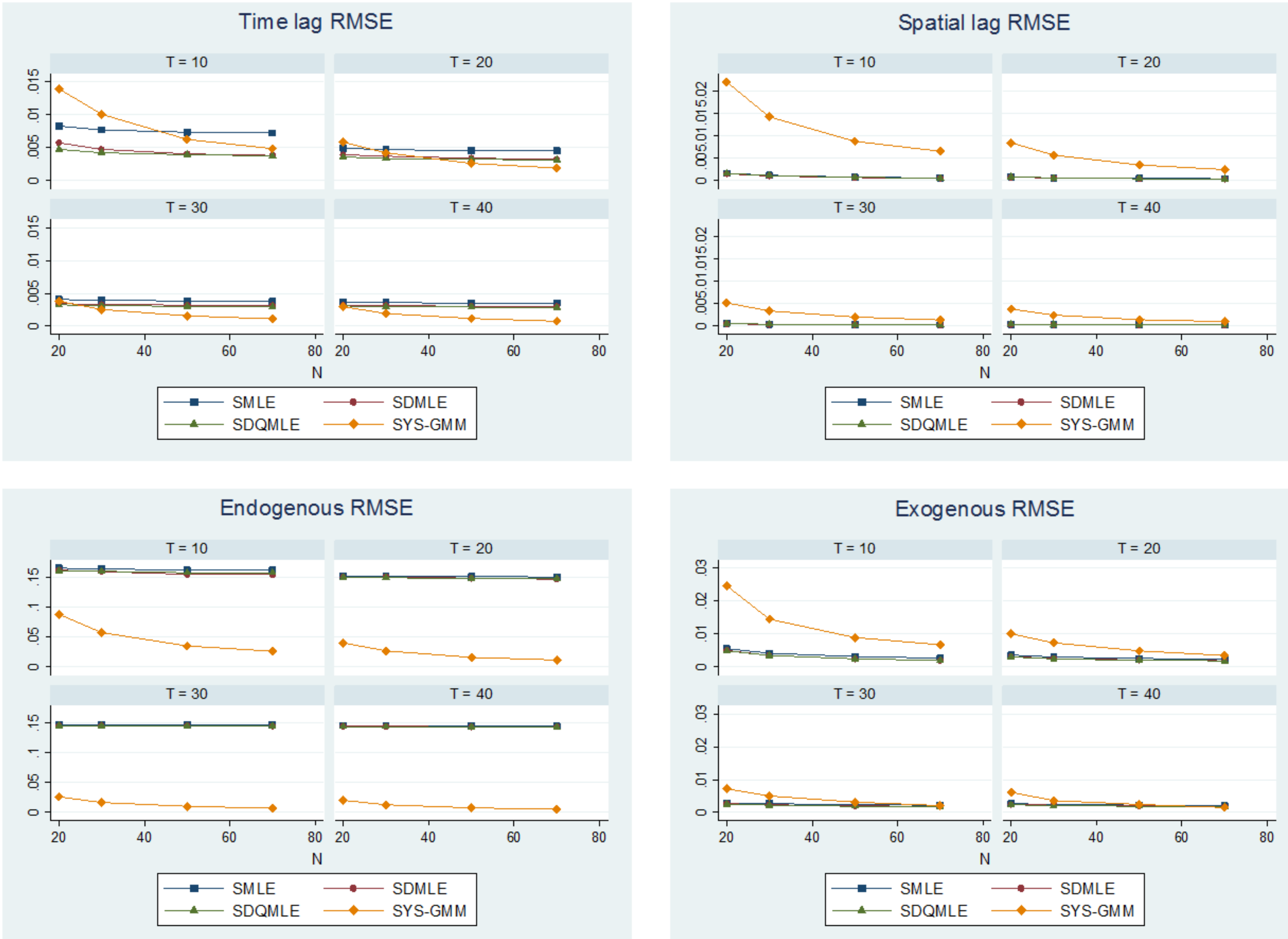


Figure 4: System GMM *vs.* Spatial Estimators RMSE

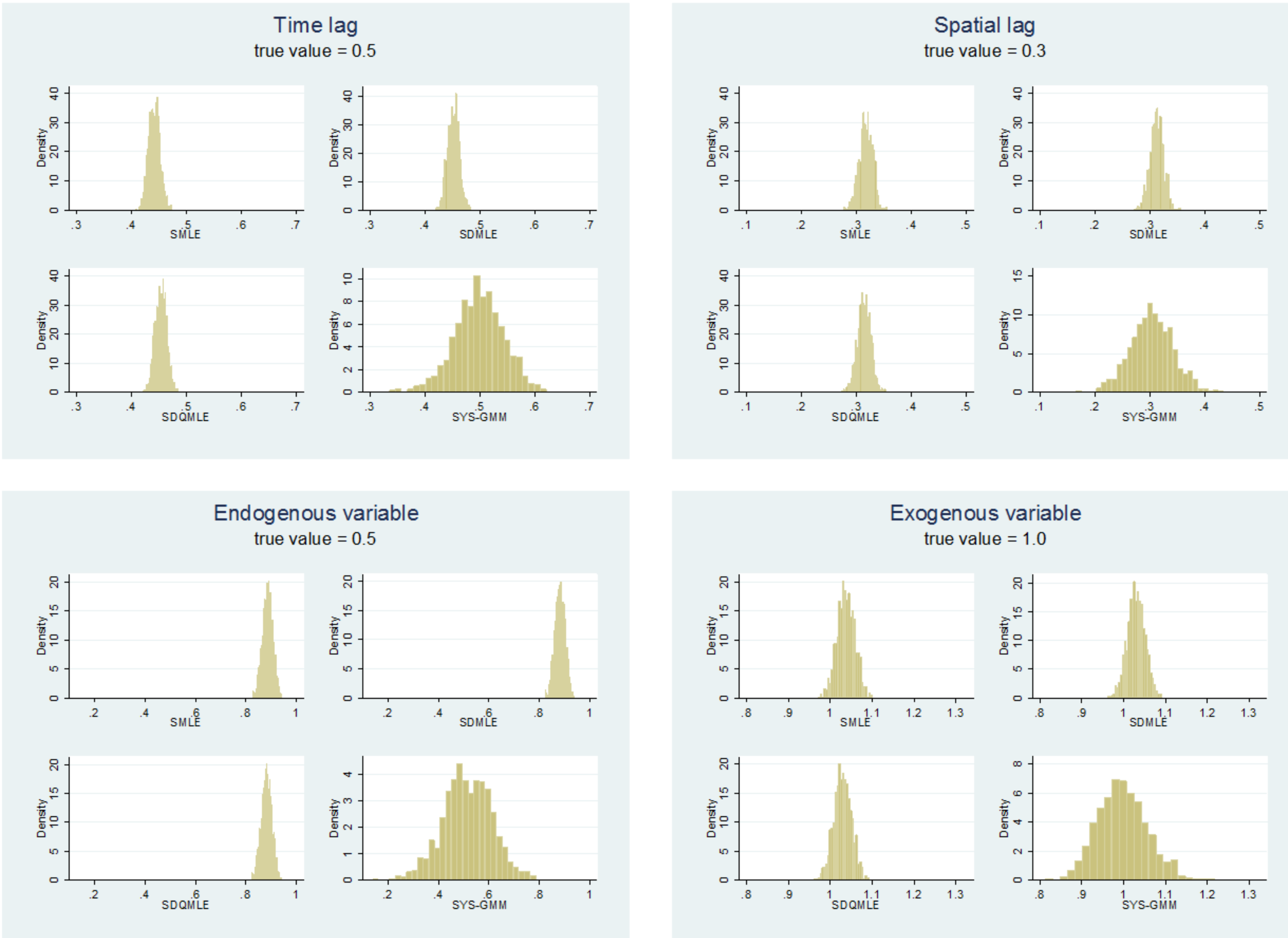


Figure 5: Parameters' Histograms (N = 70 and T = 20)

estimation of the additional parameter σ^2. Moreover, Elhorst's approach implies the estimation of the parameters for initial condition, while Yu et al's method is explicitly designed for large N and T. Therefore, the classical arguments relating to relative efficiency do not apply here.

As I already mentioned it in the bias analysis, the estimation of the endogenous covariate with system GMM yields the lowest RMSE. The slope of the RMSE line in Figure 4 is almost null for the QMLE and MLE, which suggests that increasing the sample dimension cannot decrease the RMSE associated with the endogenous variable. In other words, the use of spatial ML estimators is not recommended in the presence of endogenous or predetermined variables. Among the spatial estimators, SDQMLE is again the best one in terms of the RMSE criterion. Note that the estimation of the spatial lag parameter by simple spatial MLE yields a similar RMSE than any other spatial estimators. That is why Elhorst (2008) suggests estimating separately the spatial lag with SMLE and the remaining parameters with either SDMLE or difference GMM.

4 Robustness Check

Before investigating the sensitivity of our main results, I summarize the results in terms of RMSE response functions. This allows us to highlight potential issues that need to be addressed in the robustness checks. Specifically, I investigate four departures from the baseline model. First, I check if the results are sensitive to the choice of the spatial matrix W. Additionally, I analyze what happens when the spatial weight scheme is miss-specified. Second, I drop the assumption of Gaussian process for the error terms and individual effects. Third, I check the impact of heteroskedasticity in the error terms on the performance of the estimators. Last, I investigate the consequence of dealing with spatially dependent endogenous and exogenous variables[19].

4.1 RMSE response function

As previously commented, the relationship between the performance of the estimators and the model parameters is not easily determined. That is why, I report the results in terms of response functions. Using the RMSE for each estimator and parameters of the entire set of designs, the following equation is estimated by OLS:

$$\log\left(\sqrt{N_i \cdot T_i} \cdot RMSE_i\right) = a_1 + a_2 \frac{1}{W_i} + a_3\alpha_i + a_4\rho_i + a_5(\alpha_i\rho_i) + a_6\frac{W_i}{N_i} + a_7\frac{W_i}{T_i} + a_8\frac{1}{N_i} + a_9\frac{1}{T_i} + \varepsilon_i$$

[19] To conserve space, I do not display the results tables, but they remain available upon request.

where $RMSE_i$ is the RMSE of a given parameter obtained using a given estimator in the *i*-th design. W_i corresponds to the value attributed to the construction of the spatial weight matrix and is a measure of the degree of sparseness of matrix W (e.g. $W_i \in (1,3,5)$). Note that the dependent variable is expressed in logarithm in order to rule out negative predicted RMSE. Therefore, the interpretation of the estimated coefficients is not straightforward (i.e. a negative coefficient implies a small (close to zero) effect).

The RMSE response function estimation results are displayed in Table 2. As suggested by R^2, the fit of the response functions to the data is relatively good. The comments made in the previous section are confirmed by the estimation results. As expected, the main factor that contributes to the efficiency of system GMM is the panel size (N and T). This is not the case for the spatial estimators, because the rate of efficiency is only marginally affected by the panel size. This was already confirmed by Figures 3 and 4, where the slope of the spatial estimators' RMSE were relatively flat. The effects of the autoregressive and spatial autoregressive parameters on the RMSE are clearly non-linear. While the individual effects of the parameters are negative and significant, the interaction term enters the RMSE regression positively and significantly in almost all cases. In other words, it is the combination of the time and spatial lag coefficients which affects the efficiency of the estimators rather than each parameter taken individually. Another important finding relates to the specification of the spatial weight matrix W. The performances of the spatial estimators are definitively more sensitive to the spatial weight matrix than system GMM estimator. The spatial lag parameter is the only parameter where performance of GMM seems to be affected by the spatial matrix W. This finding suggests that spatial estimators are more sensitive to the specification of the spatial weight matrix than GMM. I address this question in the following sub-section.

4.2 Role of the Spatial Weight Matrix

In order to check the sensitivity of our results to the choice of spatial matrix, I re-estimate the model with data generated according to a theoretical Rook type of order 3 and 5 spatial weight matrix as well as a negative exponential matrix based on real data distance for 224 countries. As in the baseline case, I perform 1000 simulations for each design with respect to each type of spatial matrix.

Table 2: RMSE Response Functions

	Time Lag α :			Spatial Lag ρ :			Endogenous γ :			Exogenous β :		
	SYS-GMM	SDMLE	SDQMLE	SYS-GMM	SDMLE	SDQMLE	SYS-GMM	SDMLE	SDQMLE	SYS-GMM	SDMLE	SDQMLE
$1/W_i$	0.0251	-0.308***	-0.295***	-0.909***	-0.819***	-0.770***	-0.0352	-0.0138	-0.0146	-0.0396	-0.284***	-0.295***
	(-0.057)	(-0.064)	(-0.061)	(-0.081)	(-0.093)	(-0.092)	(-0.049)	(-0.018)	(-0.018)	(-0.063)	(-0.072)	(-0.072)
α_i	-0.229**	-3.520***	-3.821***	-2.887***	-1.858***	-1.800***	-1.002***	-0.195***	-0.187***	0.0625	-1.822***	-1.907***
	(-0.105)	(-0.122)	(-0.115)	(-0.143)	(-0.15)	(-0.151)	(-0.087)	(-0.032)	(-0.032)	(-0.109)	(-0.12)	(-0.121)
ρ_i	-0.736***	-3.072***	-3.131***	-2.615***	-2.156***	-2.012***	-0.500***	-0.157***	-0.159***	-0.441***	-1.911***	-1.899***
	(-0.116)	(-0.159)	(-0.153)	(-0.187)	(-0.205)	(-0.207)	(-0.1)	(-0.038)	(-0.039)	(-0.14)	(-0.147)	(-0.148)
$\alpha_i \rho_i$	1.437***	4.837***	5.041***	0.0916	5.945***	5.876***	0.348	0.146	0.152	0.786*	2.471***	2.380***
	(-0.384)	(-0.436)	(-0.406)	(-0.556)	(-0.646)	(-0.66)	(-0.317)	(-0.121)	(-0.123)	(-0.433)	(-0.461)	(-0.461)
W_i / N_i	0.706**	0.0427	-0.0181	1.454***	1.991***	2.055***	0.393	0.0724	0.0782	0.582	-0.155	-0.182
	(-0.357)	(-0.384)	(-0.374)	(-0.483)	(-0.484)	(-0.495)	(-0.299)	(-0.107)	(-0.11)	(-0.414)	(-0.369)	(-0.363)
W_i / T_i	-0.297*	0.0969	0.147	0.517**	1.366***	1.310***	-0.0852	0.0107	0.0049	-0.322*	-0.211	-0.206
	(-0.166)	(-0.188)	(-0.177)	(-0.24)	(-0.294)	(-0.297)	(-0.137)	(-0.048)	(-0.049)	(-0.186)	(-0.201)	(-0.195)
$1/N_i$	13.40***	-11.79***	-13.13***	13.28***	4.830***	3.885**	16.85***	-16.87***	-17.16***	15.56***	-1.461	-1.483
	(-1.317)	(-1.482)	(-1.41)	(-1.731)	(-1.657)	(-1.691)	(-1.079)	(-0.38)	(-0.391)	(-1.402)	(-1.46)	(-1.444)
$1/T_i$	13.22***	-3.730***	-4.408***	12.70***	1.976**	3.291***	11.04***	-7.673***	-7.513***	9.222***	-3.755***	-3.329***
	(-0.613)	(-0.723)	(-0.669)	(-0.865)	(-0.92)	(-0.939)	(-0.52)	(-0.166)	(-0.173)	(-0.647)	(-0.723)	(-0.711)
Constant	-3.409***	0.202***	0.302***	-0.637***	-3.258***	-3.345***	-1.191***	2.476***	2.474***	-2.762***	-1.206***	-1.228***
	(-0.057)	(-0.072)	(-0.068)	(-0.084)	(-0.09)	(-0.091)	(-0.049)	(-0.02)	(-0.019)	(-0.065)	(-0.072)	(-0.072)
R^2	0.824	0.824	0.856	0.893	0.797	0.791	0.85	0.962	0.961	0.698	0.612	0.618

Notes: The standard deviation is reported in parenthesis. *** p < 0.01, ** p < 0.05, * p < 0.1.

.

The results seem to be qualitatively similar to what I obtained in the benchmark scenario. System GMM outperforms the spatial estimators according to the unbiasness criterion. The only exception is the spatial lag parameter which is better estimated by spatial ML estimators in relatively small samples. As before, all estimators tend to underestimate the autoregressive parameter and overestimate the spatial autoregressive parameter. The coefficients of the endogenous and exogenous variable tend to be overestimated by spatial ML estimators. In addition, the bias of the endogenous and exogenous variables does not seem to decrease with an increase in the sample size (N and/or T). As highlighted by the response function analysis, the bias of the spatial estimators increases with the number of non-zeros elements in the spatial weight matrix. This is not the case for system GMM, whose performance is not affected by the degree of sparseness of the matrix W. Nevertheless, the spatial estimators' sensitivity to the type of spatial weight matrix tends to disappear as the panel dimension increases.

Most of the spatial econometrics literature presumes that the spatial weight matrix is known and well specified. However, in practice, one has to define the spatial weight matrix. This is usually done based on some underlying theory. But since there is no theoretical guidance on the choice of the matrix W, the latter can be misspecified. In order to study the consequences of the misspecification of the spatial weight matrix, I consider the data based on a Rook type of order 3 matrix, but estimating the model applying instead Rook type of order 1 or 5 spatial weight matrices. In the first case, I assume the spatial dependence to be more local than it actually is, while in the second scenario I presume the opposite.

The results depend on the type of misspecification. In the case of over-spatial dependence, the results are similar to what I obtained in the benchmark case, although the parameters' bias tends to be higher. In particular, assuming more global spatial dependence introduces additional noise, which affects the performance of system GMM. As found earlier, the autoregressive variable tends to be underestimated while the spatial autoregressive variable, endogenous and exogenous variables tend to be overestimated. Moreover, the estimation of the autoregressive and endogenous variables by the spatial estimators are dominated by system GMM according to the unbiasness criterion.

In the case of under-spatial dependence, the results for all estimators are severely affected. The bias for the spatial autoregressive parameter changes from negative to positive. In fact, limiting spatial dependence to the close neighborhood tends to underestimate the spatial effect. The main explanation for this finding lies probably in the fact that spatial methods estimate the reduced form of the model, which means that the entire set of parameters, beside the spatial lag, are affected by the miss-specification (see equation (18)). As noted previously, simple SMLE seems to estimate the spatial lag as accurate as SDMLE and SDQMLE (Elhorst, 2008).

From an applied econometric point of view, these results suggest that spatial dynamic panel models should not be estimated using only one type of spatial weight matrix. In fact, different types of spatial weight matrices (contiguity, geographical distance, economic distance) should be applied to check the robustness and consistency of the results.

4.3 Non-Gaussian Distribution

Most spatial estimators rely on the assumption of normality of the individual effects and error term. In empirical application, the data does not necessarily follow a normal distribution. That is why I investigate the consequences of dropping the Gaussian assumption through three modifications: student distribution and chi-square distribution for the error term as well as a non-normal distribution for the individual fixed effects.

First, when the error are generated according to a Student distribution with five degrees of freedom ($v_{it} \sim t(5)$), characterized by heavier tails than the normal distribution, the results remain qualitatively similar. System GMM tends to outperform spatial ML estimators although the rate of convergence seems to be slower. Among the spatial estimators, SDQMLE continues to be more robust. This corroborates Lee's (2004) finding who demonstrates that QMLE can be consistent when the disturbances are independently and identically distributed without normality.

Second, the performance of system GMM deteriorates when the shocks are generated according to a chi-square distribution with 1 degree of freedom $(v_{it} \sim \chi^2(1))$. The bias reduction associated with an increase in the panel dimension tends to be smaller than in the benchmark case. Spatial ML estimators tend to outperform system GMM according to the unbiasness criterion for the exogenous and spatial autoregressive parameters. However, the estimation of the lagged dependent and endogenous variables with system GMM continues to dominate the spatial ML estimators.

Last, following Binder et al. (2005), I also investigate the way the individual effects are generated, since the performance of system GMM depends on the ratio of the individual effect variance with respect to the variance of the error term (Hayakawa, 2006). Note that in the benchmark setting, the ratio σ_η^2/σ_v^2 is different from 1 (=1.6), which already favors the spatial estimators against system GMM. Going one step further, the individual specific effects are no longer normally distributed but generated as follows:

$$\eta_i = \sqrt{\tau}\left(\frac{q_i - 1}{\sqrt{2}}\right)m_i, \qquad q_i \overset{iid}{\sim} \chi^2(1), \qquad m_i \overset{iid}{\sim} N(0, 0.05)$$

The parameter τ measures the degree of cross-section to the time-series variations. Two values are considered $\tau = 1$ and $\tau = 5$.

As expected, spatial estimators are not affected by the way the individual specific effects are generated. This result is in line with Binder et al.'s (2005) findings. The performance of system GMM deteriorates slightly, but continues to display better results for the endogenous variable than any other spatial estimators.

4.4 Heteroskedastic Errors

Theoretically, maximum likelihood allows for heteroskedasticity in the error term. However, the practical application of such estimator is difficult and complex, which dissuades researchers to directly account for heterogeneity. This is also the case for spatial dynamic maximum likelihood estimators, which do not accommodate easily for heteroskedasticity. That is why I investigate the impact of three different types of heteroskedasticity on the performance of the estimators. First, cross-sectional heteroskedasticity in the variance of the disturbance term is defined as follows: $\sigma_{vit}^2 = \sigma_{vi}^2 \sim \chi^2(1)$. Second, time-series heteroskedasticity in the variance is specified as a function of time: $\sigma_{vit}^2 = \sigma_{vt}^2(t)$. Third, both cross-section and time-series heteroskedasticity, which are common in empirical applications, is considered by combining the first two processes. In these three specifications I make sure that the average variance is equal to one (Bun and Carree, 2006):

$$\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \approx \frac{1}{T}\sum_{i=1}^{T}\sigma_t^2 \approx \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}\sigma_t^2 \approx 1$$

Overall, the performance of all estimators deteriorates in the presence of heteroskedasticity. The absolute bias as well as the RMSE of the parameters increase in comparison with the baseline scenario. In addition, the convergence slows down for all estimators. This should not come as a surprise, since heteroskedasticity increases volatility in the data. According to the bias criterion, system GMM continues to dominate any spatial estimators for most variables. But spatial estimators still show superior performance when estimating the spatial lag parameter. However, the pattern of the bias associated with system GMM indicates that for higher time and cross-section dimensions system GMM will ultimately catch with the spatial estimators. The results associated with RMSE seem to be qualitatively similar to the main results. Once again, system GMM dominates the spatial estimators for the estimation of the lagged dependent and endogenous variables. For the exogenous and spatial lag, the rate of convergence is slower but shows a similar pattern as the benchmark results.

4.5 Additional Spatial Dependence

Finally, I modify the data generating process to account for spatial dependence in the exogenous and endogenous variables. Each one follows a spatial moving average process (with $\rho_Z = \rho_X = 0.5$):

$$Z_{it} = \delta Z_{i,t-1} + \rho_Z [Wu_t]_i + u_{it}$$
$$X_{it} = \lambda X_{i,t-1} + \psi \eta_i + \theta v_{it} + \rho_X [We_t]_i + e_{it}$$

The performance of the spatial estimators and GMM remains relatively robust to the presence of spatial dependence. Unlike the main results, the spatial ML estimators tend to underestimate the spatial lag parameter, while overestimating the effect of endogenous and exogenous variables. The spatial estimators seem to allocate higher spatial dependence to the endogenous and exogenous variables at the expense of the spatial autoregressive variable. System GMM is still less biased and more efficient than the spatial estimators for all but the spatial lag variable. The estimation of the spatial lag with system GMM leads to a negative bias, which means that system GMM continues to overestimate the effect of the spatial autoregressive term. In relatively small samples, the spatial estimators continue to dominate system GMM for the spatial lag variable. However, as sample size increases, system GMM reaches the same level of RMSE than the spatial estimators. Note that the bias of the exogenous variable for system GMM tends to be higher compared to the baseline case. Despite the fact that RMSE tends to be higher for all the estimators in small samples, the relative performance of system GMM remains qualitatively unchanged.

The performance of system GMM improves, when the spatially weighted sum of the exogenous variable ($W \cdot Z_t$) is included as an additional instrument as suggested by Kelejian and Prucha (1998). Note that this is only true when there is spatial dependence in the exogenous and endogenous variables. In fact, in the benchmark Monte Carlo setting, where $\rho_Z = \rho_X = 0$, the inclusion of $W \cdot Z_t$ as instruments does not improve significantly the performance of system GMM. From a practical viewpoint, this finding suggests to include the spatially weighted average of the exogenous variables in the set of instruments ($W \cdot Z_t$, $W^2 \cdot Z_t$,...), once the presence of spatial dependence is verified (with a Moran's *I* or Geary's *C* test for instance).

5 Conclusion

In this paper I investigate the finite samples behavior of system GMM in a spatial dynamic panel framework in the presence of an additional endogenous variable. Our Monte Carlo analysis demonstrates that while the simultaneity bias of the spatial lag remains relatively low, the bias of the endogenous variable is large if it is not corrected.

The need to account for additional sources of endogeneity leads to favour system GMM. In fact, system GMM emerges clearly dominant by an unbiasedness criterion for most variables. In addition, system GMM can even be as efficient as spatial maximum likelihood estimators in moderate and large samples. Moreover, from a purely practical point of view, system GMM is less cumbersome than any spatial estimators. It avoids the inversion of large spatial weight matrices. It is also easier to implement and its computation time is lower than any spatial dynamic estimators. Furthermore, the efficiency of system GMM could be improved through iterated GMM or continuously updated GMM.

However, the use of spatial dynamic estimators (SDMLE or SDQMLE) would still be advisable under some conditions. When the sample size (*N* and *T*) is relatively small, potential additional endogeneity can be ignored or the interpretation and measurement of the spatial autoregressive coefficient has important policy relevance (e.g. tax reaction function study), the spatial dynamic panel model should be estimated with SDMLE or SDQMLE. Recently, a lot of attention has been drawn to the impact of heterogenous and cross-section error on the bias in dynamic panel estimation with fixed effects. A possible extension of this study would be to consider weak cross-section dependence by allowing the error term to be spatially dependent. This could be done by extending the spatial HAC framework proposed by Kelejian and Prucha (2007) or Driscoll and Kray's (1998) approach to system GMM.

6 Appendices

6.1 GMM Estimators

This appendix section presents the procedure associated with the different GMM estimators considered in this study. Let Y, Y_{-1}, WY, U be $N \cdot T$ column vectors, Z is a $N \cdot T \times p$ exogenous matrix and X is a $N \cdot T \times q$ endogenous matrix. Note that the data is first sorted by time T and then by cross-section N. Thus, $Y = (Y_1; Y_2; ...; Y_T)'$, where $Y_t = (Y_{1t}; Y_{2t}; ...; Y_{Nt})'$. The same structure is applied to the remaining vectors and matrices.

As mentioned previously, the time lag $(Y_{i,t-1})$, spatial lag $([WY_t]_i)$ and endogenous variable (X_{it}) are treated as endogenous covariates, while the exogenous variable (Z_{it}) is considered as strictly exogenous. Each endogenous variable is instrumented by the strictly exogenous variable and the second and third lags of each endogenous variable. In order to restrict the number of instruments, the instruments matrix is constructed by collapsing them[20].

6.1.1 Difference GMM

The difference GMM estimator, proposed by Arellano and Bond (1991), consists of estimating the model expressed in first-difference. More specifically, the estimation steps are:

1) Construct the "collapsed" instruments matrix for each cross-section *i*:

$$Z_i^D = \begin{bmatrix} Y_{i,0} & 0 & [WY_0]_i & 0 & \Delta Z_{i,1} & X_{i,0} & 0 \\ Y_{i,1} & Y_{i,0} & [WY_1]_i & [WY_0]_i & \Delta Z_{i,2} & X_{i,1} & X_{i,0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{i,T-2} & Y_{i,T-3} & [WY_{T-2}]_i & [WY_{T-3}]_i & \Delta Z_{i,T} & X_{i,T-2} & X_{i,T-3} \end{bmatrix}$$

2) Construct the weighting matrix:

$$A^{D1} = \left(\sum_i Z_i^{D'} \cdot H_i^{D1} \cdot Z_i^D \right)^{-1}, \text{ where } H_i^{D1} = \begin{bmatrix} 1 & -0.5 & 0 & \cdots & 0 \\ -0.5 & 1 & -0.5 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & -0.5 \\ 0 & 0 & 0 & -0.5 & 1 \end{bmatrix}$$

[20] Numerous instruments can lead to two types of small-sample issues. The first problem leads to overfitting endogenous variables, i.e. failure to remove endogeneity. The second problem concerns imprecise estimation of the optimal weighting matrix in the two-step procedure. This affects the computation of two-step standard errors and the validity of Hansen's weak instruments (see Roodman, 2009).

3) Carry out the one-step estimation given by

$$\begin{bmatrix}\hat{\alpha}_1\\ \hat{\rho}_1\\ \hat{\beta}_1\\ \hat{\gamma}_1\end{bmatrix}=\left[Q_{XZ}\cdot A^{D1}\cdot Q'_{XZ}\right]^{-1}\cdot Q_{XZ}\cdot A^{D1}\cdot Q_{ZY}\ ,\ \text{where } Q_{XZ}=\sum_i X_i^{*'}\cdot Z_i \text{ and } Q_{ZX}=\sum_i Z_i^{'}\cdot Y_i^{*}$$

$$X_i^{*}=\begin{bmatrix}\Delta Y_{i,1} & \left[W\Delta Y_2\right]_i & \Delta Z_{i,2} & \Delta X_{i,2}\\ \vdots & \vdots & \vdots & \vdots\\ \Delta Y_{i,T-1} & \left[W\Delta Y_T\right]_i & \Delta Z_{i,T} & \Delta X_{i,T}\end{bmatrix} \text{ and } Y_i^{*}=\begin{bmatrix}\Delta Y_{i,2}\\ \vdots\\ \Delta Y_{i,T}\end{bmatrix}$$

– The associated variance are computed as follows:

$$\hat{V}_1=\hat{\sigma}_1^2\cdot\left[Q_{XZ}\cdot A^{D1}\cdot Q'_{XZ}\right]^{-1}\text{, where } \hat{\sigma}_1^2=\frac{1}{N-4}\sum_i\left(Y_i^{*}-X_i^{*}\left[\hat{\alpha}_1,\hat{\rho}_1,\hat{\beta}_1,\hat{\gamma}_1\right]\right)'\left(Y_i^{*}-X_i^{*}\left[\hat{\alpha}_1,\hat{\rho}_1,\hat{\beta}_1,\hat{\gamma}_1\right]\right)$$

– The robust one-step variance is given by:

$$\hat{V}_{1,Robust}=\left[Q_{XZ}\cdot A^{D1}\cdot Q'_{XZ}\right]^{-1}\cdot A^{D1}\left(A^{D2}\right)^{-1}A^{D1}Q'_{XZ}\cdot\left[Q_{XZ}\cdot A^{D1}\cdot Q'_{XZ}\right]^{-1},$$

$$\text{where } A^{D2}=\left(\sum_i Z_i^{D'}\cdot H_i^{D2}\cdot Z_i^{D}\right)^{-1} \text{ and } H_i^{D2}=\left(Y_i^{*}-X_i^{*}\left[\hat{\alpha},\hat{\rho},\hat{\beta},\hat{\gamma}\right]\right)\left(Y_i^{*}-X_i^{*}\left[\hat{\alpha},\hat{\rho},\hat{\beta},\hat{\gamma}\right]\right)'$$

– The two-step estimates are given by:

$$\begin{bmatrix}\hat{\alpha}_2\\ \hat{\rho}_2\\ \hat{\beta}_2\\ \hat{\gamma}_2\end{bmatrix}=\left[Q_{XZ}\cdot A^{D2}\cdot Q'_{XZ}\right]^{-1}\cdot Q_{XZ}\cdot A^{D2}\cdot Q_{ZY}$$

– The associated two-step variance is computed as:

$$\hat{V}_2=\left[Q_{XZ}\cdot A^{D2}\cdot Q'_{XZ}\right]^{-1}$$

6.1.2 System GMM

The system GMM estimator, proposed by Arellano and Bover (1995) and Blundell and Bond (1998), consists of combining the moment conditions from the model in first-difference with the moment conditions from the model in levels. These are the estimation steps:

1) Construct the "collapsed" instruments matrix for each cross-section i:

$$Z_i = \begin{bmatrix} Y_{i,0} & 0 & 0 & [WY_0]_i & 0 & 0 & \Delta Z_{i,1} & X_{i,0} & 0 & 0 \\ Y_{i,1} & Y_{i,0} & 0 & [WY_1]_i & [WY_0]_i & 0 & \Delta Z_{i,2} & X_{i,1} & X_{i,0} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{i,T-2} & Y_{i,T-3} & 0 & [WY_{T-2}]_i & [WY_{T-3}]_i & 0 & \Delta Z_{i,T} & X_{i,T-2} & X_{i,T-3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & Z_{i1} & 0 & 0 & 0 \\ 0 & 0 & \Delta Y_{i,1} & 0 & 0 & [W\Delta Y_2]_i & Z_{i2} & 0 & 0 & \Delta X_{i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \Delta Y_{i,T-1} & 0 & 0 & [W\Delta Y_{T-1}]_i & Z_{iT} & 0 & 0 & \Delta X_{i,T-1} \end{bmatrix}$$

2) Construct the weighting matrix:

$$A^1 = \left(\sum_i Z_i' \cdot H_i \cdot Z_i \right)^{-1}, \text{ where } H_i = \begin{bmatrix} H_i^D & 0 \\ 0 & I_i \end{bmatrix}$$

3) Carry out the one-step estimation given by:

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\rho}_1 \\ \hat{\beta}_1 \\ \hat{\gamma}_1 \end{bmatrix} = \left[Q_{XZ} \cdot A^1 \cdot Q_{XZ}' \right]^{-1} \cdot Q_{XZ} \cdot A^1 \cdot Q_{ZY}, \text{ where } Q_{XZ} = \sum_i X_i^{*'} \cdot Z_i \text{ and } Q_{ZX} = \sum_i Z_i' \cdot Y_i^*$$

$$X_i^* = \begin{bmatrix} \Delta Y_{i,1} & [W\Delta Y_2]_i & \Delta Z_{i,2} & \Delta X_{i,2} \\ \vdots & \vdots & \vdots & \vdots \\ \Delta Y_{i,T-1} & [W\Delta Y_T]_i & \Delta Z_{i,T} & \Delta X_{i,T} \\ Y_{i,0} & [WY_1]_i & Z_{i,1} & X_{i,1} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{i,T-1} & [WY_T]_i & Z_{i,T} & X_{i,T} \end{bmatrix} \text{ and } Y_i^* = \begin{bmatrix} \Delta Y_{i,2} \\ \vdots \\ \Delta Y_{i,T} \\ Y_{i,1} \\ \vdots \\ Y_{i,T} \end{bmatrix}$$

4) The associated variance are computed as follows:

$$\hat{V}_1 = \hat{\sigma}_1^2 \cdot \left[Q_{XZ} \cdot A^1 \cdot Q_{XZ}' \right]^{-1}, \text{ where } \hat{\sigma}_1^2 = \frac{1}{N-4} \sum_i \left(Y_i^* - X_i^* \left[\hat{\alpha}_1, \hat{\rho}_1, \hat{\beta}_1, \hat{\gamma}_1 \right] \right)' \left(Y_i^* - X_i^* \left[\hat{\alpha}_1, \hat{\rho}_1, \hat{\beta}_1, \hat{\gamma}_1 \right] \right)$$

5) The robust one-step variance is given by:

$$\hat{V}_{1,Robust} = \left[Q_{XZ} \cdot A^1 \cdot Q_{XZ}' \right]^{-1} \cdot A^1 \cdot \left(A^2 \right)^{-1} \cdot A^1 \cdot Q_{XZ}' \cdot \left[Q_{XZ} \cdot A^1 \cdot Q_{XZ}' \right]^{-1},$$

where $A^2 = \left(\sum_i Z_i' \cdot H_i^2 \cdot Z_i \right)^{-1}$ and $H_i^2 = \left(Y_i^* - X_i^* \left[\hat{\alpha}, \hat{\rho}, \hat{\beta}, \hat{\gamma} \right] \right) \left(Y_i^* - X_i^* \left[\hat{\alpha}, \hat{\rho}, \hat{\beta}, \hat{\gamma} \right] \right)'$

6) The two-step estimates are given by:

$$\begin{bmatrix} \hat{\alpha}_2 \\ \hat{\rho}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_2 \end{bmatrix} = \left[Q_{XZ} \cdot A^2 \cdot Q'_{XZ}\right]^{-1} \cdot Q_{XZ} \cdot A^2 \cdot Q_{ZY}$$

7) The associated two-step variance is computed as:

$$\hat{V}_2 = \left[Q_{XZ} \cdot A^2 \cdot Q'_{XZ}\right]^{-1}$$

6.2 Spatial Estimators

This appendix section presents the procedure associated with the different spatial estimators. For further details, the reader is referred to Anselin (1988), Elhorst (2003, 2005, 2008) and Yu et al. (2008). Let Y, Y_{-1}, WY, U be $N \cdot T$ column vectors, Z is a $N \cdot T \times p$ exogenous matrix and X is a $N \cdot T \times q$ endogenous matrix. Note that the data is first sorted by time T and then by cross-section N. Thus, $Y = \left(Y_1; Y_2; ...; Y_T\right)'$, where $Y_t = \left(Y_{1t}; Y_{2t}; ...; Y_{Nt}\right)'$. The same structure is applied to the remaining vectors and matrices. As initial values for the parameters, the estimates obtained by system GMM can be used.

6.2.1 Spatial MLE

The classical spatial maximum likelihood estimator relies on the concentrated likelihood in the spatial lag parameter, which is conditional upon the others' coefficient values. Operationally, "standard" spatial maximum estimation can be achieved in five steps:

1) Demean all variables, denoted by ~.

2) Carry out the following OLS regressions:

$$\tilde{Y} = \left[\tilde{Y}_{-1}; \tilde{Z}; \tilde{X}\right] b_0 + U_0$$
$$W\tilde{Y} = \left[\tilde{Y}_{-1}; \tilde{Z}; \tilde{X}\right] b_L + U_L$$

3) Compute the associated residuals $\hat{U}_0$ and $\hat{U}_L$.

4) Given $\hat{U}_0$ and $\hat{U}_L$, find ρ that maximizes the following concentrated likelihood:

$$\ln L(\rho) = -\frac{NT}{2}\ln 2\pi - \frac{NT}{2}\ln\sigma^2 + T\ln\left|I_N - \rho W\right| - \frac{NT}{2}\ln\left[\left(\hat{U}_0 - \rho\hat{U}_L\right)'\left(\hat{U}_0 - \rho\hat{U}_L\right)\right].$$

5) Given the estimate $\hat{\rho}$, the remaining coefficient estimates are computed as follows:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = b_0 - \hat{\rho} b_L \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{NT}\left(\hat{U}_0 - \hat{\rho}\hat{U}_L\right)'\left(\hat{U}_0 - \hat{\rho}\hat{U}_L\right)$$

As mentioned in Elhorst (2008), this spatial MLE is inconsistent, because of the presence of the lag dependent variable.

6.2.2 Spatial Dynamic MLE

The unconditional MLE, proposed by Elhorst (2005, 2008), involves a two-steps iterative procedure once the data has been first-differenced. Note that the initial observations are approximated using Bhargava and Sargan approach (1983). Estimation should proceed according to the following steps:

1) Take the first-difference of all variables.

2) Define some initial values for the parameters α, ρ and θ, where $\theta = \sigma_\xi^2 / \sigma^2$ and σ_ξ^2 is the variance associated with the approximation of the initial observations.

3) The two-steps iterative procedure begins here with the computation of the coefficients π_i associated with the initial observations' approximation as well as the parameters of the exogenous and endogenous covariates, and the variance σ^2 :

$$\begin{bmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \\ \vdots \\ \hat{\pi}_T \\ \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \left(\underline{\Delta X}' H_{V\theta}^{-1} \underline{\Delta X}\right)^{-1} \underline{\Delta X}' H_{V\theta}^{-1} \underline{\Delta Y} \quad \text{and} \quad \hat{\sigma}^2 = \frac{\Delta\hat{U}' H_{V\theta}^{-1} \Delta\hat{U}}{NT}.$$

where $\underline{\Delta X} = \begin{bmatrix} I_N & \Delta X_1 & \cdots & \Delta X_T & 0 \\ 0 & 0 & \cdots & 0 & \Delta X_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \Delta X_T \end{bmatrix}$; $\underline{\Delta Y} = \begin{bmatrix} (I_N - \rho W)\Delta Y_1 \\ (I_N - \rho W)\Delta Y_2 - \alpha \Delta Y_1 \\ \vdots \\ (I_N - \rho W)\Delta Y_T - \alpha \Delta Y_{T-1} \end{bmatrix}$;

$$H_{V\theta}=\begin{bmatrix} V_\theta & -I_N & 0 & \cdots & 0 & 0 \\ -I_N & 2\cdot I_N & -I_N & \ddots & 0 & 0 \\ 0 & -I_N & 2\cdot I_N & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2\cdot I_N & -I_N \\ 0 & 0 & 0 & \cdots & -I_N & 2\cdot I_N \end{bmatrix}; \quad \Delta\hat{U}=\underline{\Delta Y}-\underline{\Delta X}\left[\hat{\pi}_1;\cdots;\hat{\pi}_T;\hat{\beta}';\hat{\gamma}'\right]$$

$$\begin{aligned} V_\theta &= \theta I_N + I_N + (\alpha S - I_N)\left(I_N-\alpha^2 SS'\right)^{-1}(\alpha S - I_N)' \\ &\quad -(\alpha S - I_N)(\alpha S)^{m-1}\left(I_N-\alpha^2 SS'\right)^{-1}(\alpha S)^{m-1}(\alpha S - I_N)' - \left(\alpha^2 SS'\right)^{m-1}; \\ S &= (I_N-\rho W)^{-1}; \end{aligned}$$

The parameter *m*, which represents the number of periods since the process started, should be defined in advance. It must be such that the eigenvalues of the matrix αS lie inside the unit circle, because otherwise the matrix $(\alpha S)^{m-1}$ would become infinite and yield a corner solution. Instead of defining *m* in advance, Elhorst (2008) proposes to include a third step procedure to estimate it. Besides increasing the computation time, this additional step improves marginally the results.

4) Given the set of parameters obtained in step 3, maximize the unconditional likelihood function as follows:
   $$\ln L(\alpha,\rho,\theta)=-\frac{NT}{2}\ln 2\pi-\frac{NT}{2}\ln\sigma^2+T\ln\left|I_N-\rho W\right|-\frac{1}{2}\ln\left|H_{V\theta}-\frac{1}{2\sigma^2}\right|\Delta\hat{U}H_{V\theta}^{-1}\Delta\hat{U}$$
   w.r.t. $|\alpha|<1-\rho\omega_{\max}$ and $|\alpha|<1-\rho\omega_{\min}$

5) Repeat step 3, with the estimates obtained in step 4 and so on, until convergence is met.

Note that to reduce the computation time the Jacobian term, $\ln|I_N-\rho W|$, in the loglikelihood function is approximated by $\sum_{i=1}^{N}\ln(1-\rho\omega_i)$, where ω_i is the eigenvalue of matrix W. The inverse of matrix $H_{V\theta}$ is also estimated using summation operations instead of matrix calculus.

6.2.3 Spatial Dynamic QMLE

The QMLE, presented by Yu et al. (2008), requires first the maximization of the concentrated likelihood and then a bias correction. Note that the original model proposed by the authors includes a lagged spatial lag and corresponds to a "time-space dynamic" model (based on Anselin taxonomy (1988, 2001)). The estimation process involves the following steps:

1) Demean all variables, denoted by ~.

2) Maximize the following concentrated likelihood function in order to estimate $\hat{\alpha},\ \hat{\rho},\ \hat{\beta},\ \hat{\gamma}$ and $\hat{\sigma}^2$:

$$\ln L(\alpha,\rho,\beta,\gamma,\sigma^2) = -\frac{NT}{2}\ln 2\pi - \frac{NT}{2}\ln\sigma^2 + T\ln|I_N - \rho W| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{U}_t'\tilde{U}_t$$

w.r.t. $\sum_{t=1}^{T}\tilde{Y}_{-1}'\tilde{U}_t = 0$, $\sum_{t=1}^{T}\left(W\tilde{Y}_{-1}'\right)'\tilde{U}_t = tr\left(W(I_N - \rho W)^{-1}\right)$, $\sum_{t=1}^{T}\tilde{Z}'\tilde{U}_t = 0$, $\sum_{t=1}^{T}\tilde{X}'\tilde{U}_t = 0$ and $\sum_{t=1}^{T}\tilde{U}_t'\tilde{U}_t = N\sigma^2$, where $\tilde{U}_t = (I - \rho W)\tilde{Y}_t - \left[\tilde{Y}_t;\tilde{Z};\tilde{X}\right]\left[\alpha;\beta';\gamma'\right]'$.

3) The bias-corrected estimator is then given by:

$$\begin{bmatrix}\hat{\alpha}^c\\ \hat{\rho}^c\\ \hat{\beta}^c\\ \hat{\gamma}^c\\ \hat{\sigma}^{2c}\end{bmatrix} = \begin{bmatrix}\hat{\alpha}\\ \hat{\rho}\\ \hat{\beta}\\ \hat{\gamma}\\ \hat{\sigma}^2\end{bmatrix} - \frac{1}{T}\left(-\hat{\Sigma}^{-1}b\right)$$

where $\hat{\Sigma}^{-1}$ can be approximated by the empirical Hessian matrix of the concentrated log likelihood function (an analytical expression for the matrix Σ can also be found in Yu et al.) and the column matrix b is given by:

$$b = \begin{bmatrix} \frac{1}{N}tr\left(\left(I_N - \alpha(I_N - \hat{\rho}W)^{-1}\right)(I_N - \hat{\rho}W)^{-1}\right) \\ \frac{\hat{\alpha}}{N}tr\left(W(I_N - \hat{\rho}W)^{-1}\left(I_N - \alpha(I_N - \hat{\rho}W)^{-1}\right)(I_N - \hat{\rho}W)^{-1}\right) + \frac{1}{N}tr\left(W(I_N - \hat{\rho}W)^{-1}\right) \\ 0 \\ 0 \\ \frac{1}{2\hat{\sigma}^2}\end{bmatrix}$$

4) Finally, the individual effects are recovered as follows:

$$\hat{\eta} = \frac{1}{T}\sum_{t=1}^{T}(I_N - \hat{\rho}^c W)Y_t - [Y_{-1};Z;X]\left[\hat{\alpha}^c;\hat{\beta}^{c\prime};\hat{\gamma}^{c\prime}\right]'$$

6.3 Monte Carlo Results: Bias

Time lag variable α : Bias

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	0.1076	0.0860	0.0807	0.0935	0.0588	0.0272
10	20	0.5	0.3	0.0784	0.0633	0.0518	0.0618	0.0472	0.0244
10	20	0.7	0.1	0.0722	0.0562	0.0402	0.0455	0.0762	-0.0561
20	20	0.2	0.1	0.0913	0.0804	0.0784	0.0413	0.0166	0.0180
20	20	0.5	0.3	0.0585	0.0514	0.0467	0.0614	0.0226	0.0115
20	20	0.7	0.1	0.0478	0.0407	0.0352	0.0211	0.0200	-0.0052
30	20	0.2	0.1	0.0857	0.0783	0.0770	-0.0591	0.0122	0.0141
30	20	0.5	0.3	0.0531	0.0483	0.0455	0.0608	0.0141	0.0099
30	20	0.7	0.1	0.0404	0.0362	0.0328	-0.0134	0.0102	-0.0008
40	20	0.2	0.1	0.0819	0.0763	0.0756	0.0803	0.0082	0.0096
40	20	0.5	0.3	0.0500	0.0463	0.0443	-0.0666	0.0081	0.0046
40	20	0.7	0.1	0.0374	0.0343	0.0327	-0.0064	0.0082	0.0004
10	30	0.2	0.1	0.1081	0.0851	0.0821	0.1066	0.0298	0.0284
10	30	0.5	0.3	0.0780	0.0571	0.0506	-0.0066	0.0437	0.0013
10	30	0.7	0.1	0.0716	0.0459	0.0388	0.0651	0.0485	-0.0042
20	30	0.2	0.1	0.0917	0.0807	0.0785	0.0921	0.0122	0.0143
20	30	0.5	0.3	0.0588	0.0514	0.0464	0.0689	0.0148	0.0038
20	30	0.7	0.1	0.0471	0.0404	0.0344	0.0068	0.0119	-0.0082
30	30	0.2	0.1	0.0850	0.0775	0.0763	-0.0514	0.0100	0.0071
30	30	0.5	0.3	0.0525	0.0477	0.0447	0.0345	0.0090	0.0028
30	30	0.7	0.1	0.0398	0.0355	0.0327	-0.0119	0.0081	-0.0059
40	30	0.2	0.1	0.0828	0.0773	0.0762	0.0651	0.0057	0.0059
40	30	0.5	0.3	0.0486	0.0452	0.0432	0.0262	0.0047	0.0044
40	30	0.7	0.1	0.0366	0.0335	0.0320	0.0119	0.0047	-0.0042
10	50	0.2	0.1	0.1070	0.0825	0.0808	0.1064	0.0203	0.0177
10	50	0.5	0.3	0.0795	0.0547	0.0524	0.0601	0.0326	0.0004
10	50	0.7	0.1	0.0706	0.0437	0.0385	0.0589	0.0228	-0.0163
20	50	0.2	0.1	0.0897	0.0786	0.0768	0.0849	0.0091	0.0095
20	50	0.5	0.3	0.0580	0.0492	0.0461	0.0693	0.0087	0.0036
20	50	0.7	0.1	0.0473	0.0377	0.0344	0.0457	0.0090	-0.0106
30	50	0.2	0.1	0.0853	0.0780	0.0767	0.0488	0.0048	0.0051
30	50	0.5	0.3	0.0530	0.0483	0.0453	0.0361	0.0060	0.0007
30	50	0.7	0.1	0.0401	0.0358	0.0327	0.0194	0.0052	-0.0042
40	50	0.2	0.1	0.0818	0.0763	0.0755	0.0588	0.0032	0.0051
40	50	0.5	0.3	0.0499	0.0464	0.0443	0.0596	0.0040	0.0006
40	50	0.7	0.1	0.0369	0.0337	0.0323	-0.0440	0.0044	-0.0051
10	70	0.2	0.1	0.1064	0.0812	0.0802	0.0975	0.0149	0.0136
10	70	0.5	0.3	0.0787	0.0538	0.0515	0.0907	0.0200	0.0048
10	70	0.7	0.1	0.0692	0.0429	0.0375	0.0363	0.0193	-0.0129
20	70	0.2	0.1	0.0907	0.0789	0.0776	0.0910	0.0063	0.0077
20	70	0.5	0.3	0.0583	0.0473	0.0464	0.0628	0.0076	0.0003
20	70	0.7	0.1	0.0468	0.0350	0.0338	0.0352	0.0077	-0.0094
30	70	0.2	0.1	0.0847	0.0773	0.0761	0.0714	0.0034	0.0038
30	70	0.5	0.3	0.0516	0.0461	0.0440	0.0247	0.0036	0.0000
30	70	0.7	0.1	0.0392	0.0336	0.0314	0.0212	0.0041	-0.0025
40	70	0.2	0.1	0.0825	0.0770	0.0760	0.0612	0.0013	0.0020
40	70	0.5	0.3	0.0492	0.0456	0.0434	0.0597	0.0026	0.0019
40	70	0.7	0.1	0.0366	0.0335	0.0309	0.0322	0.0016	-0.0034

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0,0.08)$; $v_{it} \sim N(0,0.05)$; $u_{it} \sim N(0,0.05)$; $e_{it} \sim N(0,0.05)$.

Spatial lag variable ρ : Bias

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	-0.0013	-0.0006	0.0005	-0.0825	-0.0148	-0.0326
10	20	0.5	0.3	-0.0151	-0.0116	-0.0104	-0.0883	-0.0216	-0.0140
10	20	0.7	0.1	-0.0081	-0.0085	-0.0073	-0.0812	-0.0253	-0.0005
20	20	0.2	0.1	-0.0054	-0.0053	-0.0048	-0.0460	-0.0127	0.0020
20	20	0.5	0.3	-0.0152	-0.0130	-0.0124	-0.0482	-0.0125	-0.0164
20	20	0.7	0.1	-0.0073	-0.0066	-0.0065	-0.0665	-0.0036	-0.0009
30	20	0.2	0.1	-0.0045	-0.0042	-0.0041	-0.1479	0.0000	-0.0084
30	20	0.5	0.3	-0.0174	-0.0158	-0.0153	-0.0248	-0.0066	-0.0092
30	20	0.7	0.1	-0.0072	-0.0066	-0.0070	0.0752	-0.0035	0.0063
40	20	0.2	0.1	-0.0046	-0.0041	-0.0043	-0.0222	-0.0014	-0.0046
40	20	0.5	0.3	-0.0161	-0.0151	-0.0147	-0.0150	0.0003	-0.0024
40	20	0.7	0.1	-0.0070	-0.0062	-0.0070	-0.0163	-0.0015	-0.0021
10	30	0.2	0.1	-0.0027	-0.0015	-0.0015	-0.0176	-0.0111	-0.0078
10	30	0.5	0.3	-0.0150	-0.0113	-0.0120	-0.0462	-0.0139	-0.0039
10	30	0.7	0.1	-0.0074	-0.0053	-0.0077	-0.0067	-0.0008	-0.0016
20	30	0.2	0.1	-0.0039	-0.0033	-0.0031	-0.0286	-0.0075	-0.0037
20	30	0.5	0.3	-0.0170	-0.0146	-0.0140	-0.0518	-0.0013	-0.0051
20	30	0.7	0.1	-0.0077	-0.0067	-0.0069	0.0040	-0.0014	-0.0056
30	30	0.2	0.1	-0.0044	-0.0039	-0.0041	0.0948	-0.0056	-0.0016
30	30	0.5	0.3	-0.0173	-0.0158	-0.0153	0.0060	0.0017	-0.0029
30	30	0.7	0.1	-0.0068	-0.0059	-0.0060	0.0032	-0.0024	-0.0026
40	30	0.2	0.1	-0.0037	-0.0034	-0.0033	0.0070	0.0032	-0.0026
40	30	0.5	0.3	-0.0160	-0.0151	-0.0147	-0.0202	-0.0046	-0.0056
40	30	0.7	0.1	-0.0081	-0.0076	-0.0078	0.0032	-0.0014	-0.0010
10	50	0.2	0.1	-0.0034	-0.0010	-0.0025	-0.0267	-0.0077	-0.0051
10	50	0.5	0.3	-0.0144	-0.0089	-0.0101	-0.1200	-0.0052	-0.0219
10	50	0.7	0.1	-0.0063	-0.0040	-0.0061	-0.0286	-0.0005	-0.0069
20	50	0.2	0.1	-0.0032	-0.0028	-0.0026	-0.0195	-0.0008	0.0011
20	50	0.5	0.3	-0.0172	-0.0144	-0.0143	-0.0499	0.0001	-0.0035
20	50	0.7	0.1	-0.0079	-0.0060	-0.0068	-0.0219	-0.0005	-0.0027
30	50	0.2	0.1	-0.0039	-0.0034	-0.0035	-0.0522	0.0007	-0.0006
30	50	0.5	0.3	-0.0174	-0.0158	-0.0155	-0.0006	-0.0014	-0.0035
30	50	0.7	0.1	-0.0074	-0.0066	-0.0066	0.0060	-0.0037	-0.0023
40	50	0.2	0.1	-0.0036	-0.0034	-0.0033	-0.0448	0.0037	0.0043
40	50	0.5	0.3	-0.0173	-0.0161	-0.0158	-0.0518	-0.0018	-0.0037
40	50	0.7	0.1	-0.0075	-0.0068	-0.0072	-0.0200	-0.0027	-0.0034
10	70	0.2	0.1	-0.0041	-0.0024	-0.0031	-0.0287	-0.0043	-0.0032
10	70	0.5	0.3	-0.0159	-0.0104	-0.0111	-0.0550	-0.0075	-0.0154
10	70	0.7	0.1	-0.0064	-0.0038	-0.0056	0.0101	0.0027	-0.0039
20	70	0.2	0.1	-0.0044	-0.0038	-0.0038	-0.0230	-0.0033	-0.0029
20	70	0.5	0.3	-0.0176	-0.0134	-0.0144	-0.0443	-0.0006	-0.0037
20	70	0.7	0.1	-0.0077	-0.0058	-0.0068	-0.0081	-0.0017	-0.0034
30	70	0.2	0.1	-0.0045	-0.0040	-0.0041	-0.0114	0.0011	-0.0010
30	70	0.5	0.3	-0.0165	-0.0145	-0.0144	-0.0140	-0.0008	-0.0025
30	70	0.7	0.1	-0.0074	-0.0063	-0.0067	0.0131	0.0004	0.0003
40	70	0.2	0.1	-0.0043	-0.0040	-0.0039	-0.0302	-0.0019	-0.0026
40	70	0.5	0.3	-0.0171	-0.0163	-0.0158	-0.0557	-0.0003	-0.0022
40	70	0.7	0.1	-0.0078	-0.0072	-0.0073	-0.0091	0.0002	-0.0004

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0,0.08)$; $v_{it} \sim N(0,0.05)$; $u_{it} \sim N(0,0.05)$; $e_{it} \sim N(0,0.05)$.

Endogenous variable γ : Bias

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	-0.4125	-0.4065	-0.4050	-0.4107	-0.1361	-0.1614
10	20	0.5	0.3	-0.3994	-0.3958	-0.3932	-0.4003	-0.1305	-0.1308
10	20	0.7	0.1	-0.3942	-0.3931	-0.3905	-0.4014	-0.1193	-0.0894
20	20	0.2	0.1	-0.3987	-0.3950	-0.3941	-0.4117	-0.0589	-0.0822
20	20	0.5	0.3	-0.3845	-0.3826	-0.3808	-0.3863	-0.0617	-0.0695
20	20	0.7	0.1	-0.3817	-0.3793	-0.3776	-0.3870	-0.0404	-0.0389
30	20	0.2	0.1	-0.3932	-0.3905	-0.3898	-0.4266	-0.0366	-0.0472
30	20	0.5	0.3	-0.3786	-0.3766	-0.3758	-0.3838	-0.0384	-0.0388
30	20	0.7	0.1	-0.3779	-0.3761	-0.3741	-0.3989	-0.0451	-0.0341
40	20	0.2	0.1	-0.3907	-0.3885	-0.3882	-0.3906	-0.0334	-0.0455
40	20	0.5	0.3	-0.3790	-0.3775	-0.3766	-0.4267	-0.0213	-0.0332
40	20	0.7	0.1	-0.3733	-0.3714	-0.3701	-0.3862	-0.0208	-0.0252
10	30	0.2	0.1	-0.4192	-0.4138	-0.4131	-0.4192	-0.0764	-0.1107
10	30	0.5	0.3	-0.3975	-0.3943	-0.3924	-0.4378	-0.0808	-0.0781
10	30	0.7	0.1	-0.3947	-0.3902	-0.3909	-0.3982	-0.0877	-0.0805
20	30	0.2	0.1	-0.3980	-0.3941	-0.3936	-0.3970	-0.0341	-0.0552
20	30	0.5	0.3	-0.3861	-0.3837	-0.3820	-0.3838	-0.0448	-0.0475
20	30	0.7	0.1	-0.3806	-0.3787	-0.3771	-0.3946	-0.0390	-0.0323
30	30	0.2	0.1	-0.3925	-0.3901	-0.3893	-0.4420	-0.0180	-0.0327
30	30	0.5	0.3	-0.3802	-0.3785	-0.3776	-0.3999	-0.0270	-0.0256
30	30	0.7	0.1	-0.3780	-0.3764	-0.3749	-0.3952	-0.0248	-0.0189
40	30	0.2	0.1	-0.3918	-0.3896	-0.3892	-0.3983	-0.0221	-0.0325
40	30	0.5	0.3	-0.3771	-0.3757	-0.3748	-0.3924	-0.0193	-0.0272
40	30	0.7	0.1	-0.3759	-0.3748	-0.3730	-0.3859	-0.0143	-0.0168
10	50	0.2	0.1	-0.4095	-0.4031	-0.4026	-0.4076	-0.0540	-0.0762
10	50	0.5	0.3	-0.3980	-0.3939	-0.3927	-0.3930	-0.0598	-0.0483
10	50	0.7	0.1	-0.3945	-0.3906	-0.3898	-0.3978	-0.0490	-0.0444
20	50	0.2	0.1	-0.3990	-0.3951	-0.3945	-0.3999	-0.0247	-0.0440
20	50	0.5	0.3	-0.3856	-0.3828	-0.3816	-0.3838	-0.0277	-0.0333
20	50	0.7	0.1	-0.3811	-0.3784	-0.3773	-0.3814	-0.0210	-0.0119
30	50	0.2	0.1	-0.3946	-0.3918	-0.3914	-0.4024	-0.0133	-0.0228
30	50	0.5	0.3	-0.3805	-0.3789	-0.3777	-0.3984	-0.0148	-0.0142
30	50	0.7	0.1	-0.3762	-0.3744	-0.3729	-0.3833	-0.0127	-0.0123
40	50	0.2	0.1	-0.3918	-0.3895	-0.3891	-0.3964	-0.0155	-0.0257
40	50	0.5	0.3	-0.3778	-0.3764	-0.3757	-0.3773	-0.0137	-0.0134
40	50	0.7	0.1	-0.3753	-0.3741	-0.3726	-0.3992	-0.0093	-0.0056
10	70	0.2	0.1	-0.4105	-0.4037	-0.4032	-0.4130	-0.0448	-0.0558
10	70	0.5	0.3	-0.4009	-0.3955	-0.3952	-0.3969	-0.0471	-0.0408
10	70	0.7	0.1	-0.3918	-0.3877	-0.3871	-0.4046	-0.0400	-0.0271
20	70	0.2	0.1	-0.3982	-0.3941	-0.3937	-0.3975	-0.0174	-0.0300
20	70	0.5	0.3	-0.3862	-0.3829	-0.3823	-0.3886	-0.0224	-0.0155
20	70	0.7	0.1	-0.3809	-0.3773	-0.3769	-0.3850	-0.0223	-0.0184
30	70	0.2	0.1	-0.3940	-0.3911	-0.3907	-0.3985	-0.0152	-0.0228
30	70	0.5	0.3	-0.3804	-0.3784	-0.3775	-0.3998	-0.0122	-0.0141
30	70	0.7	0.1	-0.3773	-0.3752	-0.3744	-0.3844	-0.0056	-0.0049
40	70	0.2	0.1	-0.3903	-0.3882	-0.3878	-0.3960	-0.0023	-0.0110
40	70	0.5	0.3	-0.3768	-0.3754	-0.3747	-0.3751	-0.0067	-0.0083
40	70	0.7	0.1	-0.3745	-0.3734	-0.3722	-0.3766	-0.0086	-0.0062

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0,0.08)$; $v_{it} \sim N(0,0.05)$; $u_{it} \sim N(0,0.05)$; $e_{it} \sim N(0,0.05)$.

Exogenous variable β : Bias

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	-0.0487	-0.0387	-0.0358	-0.0365	0.0098	-0.0062
10	20	0.5	0.3	-0.0329	-0.0258	-0.0219	-0.0103	0.0088	-0.0013
10	20	0.7	0.1	-0.0234	-0.0189	-0.0129	-0.0118	0.0318	0.0517
20	20	0.2	0.1	-0.0581	-0.0512	-0.0499	-0.0218	-0.0029	-0.0145
20	20	0.5	0.3	-0.0390	-0.0344	-0.0307	-0.0308	0.0017	-0.0053
20	20	0.7	0.1	-0.0316	-0.0260	-0.0222	-0.0073	0.0096	0.0163
30	20	0.2	0.1	-0.0572	-0.0522	-0.0514	0.0466	0.0001	-0.0098
30	20	0.5	0.3	-0.0403	-0.0362	-0.0340	-0.0434	0.0007	-0.0068
30	20	0.7	0.1	-0.0342	-0.0298	-0.0278	0.0025	0.0008	0.0091
40	20	0.2	0.1	-0.0553	-0.0513	-0.0507	-0.0527	-0.0017	-0.0032
40	20	0.5	0.3	-0.0381	-0.0355	-0.0338	0.0631	0.0017	-0.0017
40	20	0.7	0.1	-0.0330	-0.0298	-0.0278	0.0089	-0.0028	0.0052
10	30	0.2	0.1	-0.0572	-0.0457	-0.0437	-0.0560	-0.0046	-0.0185
10	30	0.5	0.3	-0.0311	-0.0225	-0.0200	0.0162	0.0072	0.0097
10	30	0.7	0.1	-0.0263	-0.0164	-0.0149	-0.0252	0.0142	0.0132
20	30	0.2	0.1	-0.0553	-0.0489	-0.0477	-0.0546	-0.0022	-0.0069
20	30	0.5	0.3	-0.0344	-0.0298	-0.0264	-0.0326	0.0014	0.0053
20	30	0.7	0.1	-0.0326	-0.0275	-0.0231	-0.0040	0.0062	0.0129
30	30	0.2	0.1	-0.0554	-0.0504	-0.0496	0.0309	-0.0006	-0.0074
30	30	0.5	0.3	-0.0373	-0.0338	-0.0319	-0.0284	0.0009	0.0034
30	30	0.7	0.1	-0.0340	-0.0297	-0.0274	0.0084	0.0010	0.0090
40	30	0.2	0.1	-0.0574	-0.0536	-0.0529	-0.0463	-0.0036	-0.0077
40	30	0.5	0.3	-0.0393	-0.0364	-0.0347	-0.0177	-0.0009	-0.0030
40	30	0.7	0.1	-0.0349	-0.0320	-0.0291	-0.0118	0.0002	0.0044
10	50	0.2	0.1	-0.0512	-0.0389	-0.0382	-0.0500	0.0017	-0.0049
10	50	0.5	0.3	-0.0353	-0.0255	-0.0238	-0.0043	0.0050	0.0131
10	50	0.7	0.1	-0.0262	-0.0174	-0.0156	-0.0204	0.0040	0.0148
20	50	0.2	0.1	-0.0557	-0.0486	-0.0476	-0.0523	-0.0018	-0.0075
20	50	0.5	0.3	-0.0384	-0.0330	-0.0303	-0.0381	-0.0004	-0.0027
20	50	0.7	0.1	-0.0323	-0.0256	-0.0232	-0.0309	0.0034	0.0130
30	50	0.2	0.1	-0.0556	-0.0507	-0.0498	-0.0287	-0.0003	-0.0038
30	50	0.5	0.3	-0.0390	-0.0355	-0.0332	-0.0293	-0.0004	0.0029
30	50	0.7	0.1	-0.0331	-0.0293	-0.0265	-0.0173	-0.0009	0.0082
40	50	0.2	0.1	-0.0558	-0.0520	-0.0513	-0.0372	0.0002	-0.0029
40	50	0.5	0.3	-0.0395	-0.0368	-0.0350	-0.0366	-0.0018	0.0019
40	50	0.7	0.1	-0.0334	-0.0307	-0.0280	0.0445	-0.0012	0.0069
10	70	0.2	0.1	-0.0512	-0.0391	-0.0386	-0.0457	0.0018	-0.0024
10	70	0.5	0.3	-0.0328	-0.0228	-0.0210	-0.0303	0.0027	0.0059
10	70	0.7	0.1	-0.0240	-0.0148	-0.0123	-0.0119	0.0056	0.0172
20	70	0.2	0.1	-0.0552	-0.0478	-0.0472	-0.0544	0.0001	-0.0059
20	70	0.5	0.3	-0.0374	-0.0301	-0.0294	-0.0335	-0.0015	0.0049
20	70	0.7	0.1	-0.0330	-0.0251	-0.0241	-0.0259	0.0022	0.0100
30	70	0.2	0.1	-0.0561	-0.0514	-0.0507	-0.0467	-0.0007	-0.0018
30	70	0.5	0.3	-0.0381	-0.0341	-0.0325	-0.0164	0.0006	0.0008
30	70	0.7	0.1	-0.0332	-0.0284	-0.0266	-0.0198	-0.0002	0.0065
40	70	0.2	0.1	-0.0567	-0.0528	-0.0522	-0.0400	-0.0013	-0.0048
40	70	0.5	0.3	-0.0395	-0.0368	-0.0349	-0.0363	-0.0025	0.0001
40	70	0.7	0.1	-0.0334	-0.0305	-0.0281	-0.0293	0.0003	0.0055

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0, 0.08)$; $v_{it} \sim N(0, 0.05)$; $u_{it} \sim N(0, 0.05)$; $e_{it} \sim N(0, 0.05)$.

6.4 Monte Carlo Results: RMSE

Time lag variable α : RMSE

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	0.0135	0.0092	0.0085	0.0108	0.0235	0.0177
10	20	0.5	0.3	0.0074	0.0051	0.0038	0.0050	0.0207	0.0125
10	20	0.7	0.1	0.0062	0.0040	0.0026	0.0033	0.1781	0.0173
20	20	0.2	0.1	0.0089	0.0070	0.0067	0.0025	0.0048	0.0063
20	20	0.5	0.3	0.0039	0.0031	0.0026	0.0043	0.0059	0.0063
20	20	0.7	0.1	0.0026	0.0020	0.0016	0.0009	0.0085	0.0051
30	20	0.2	0.1	0.0078	0.0065	0.0063	0.0040	0.0026	0.0039
30	20	0.5	0.3	0.0031	0.0026	0.0023	0.0040	0.0027	0.0036
30	20	0.7	0.1	0.0018	0.0015	0.0013	0.0004	0.0032	0.0033
40	20	0.2	0.1	0.0071	0.0062	0.0061	0.0068	0.0020	0.0030
40	20	0.5	0.3	0.0027	0.0023	0.0022	0.0046	0.0020	0.0030
40	20	0.7	0.1	0.0015	0.0013	0.0012	0.0002	0.0016	0.0022
10	30	0.2	0.1	0.0126	0.0082	0.0077	0.0122	0.0094	0.0096
10	30	0.5	0.3	0.0069	0.0039	0.0033	0.0009	0.0183	0.0126
10	30	0.7	0.1	0.0057	0.0026	0.0021	0.0049	0.0282	0.0087
20	30	0.2	0.1	0.0088	0.0069	0.0066	0.0088	0.0033	0.0047
20	30	0.5	0.3	0.0037	0.0029	0.0024	0.0051	0.0032	0.0042
20	30	0.7	0.1	0.0024	0.0018	0.0014	0.0003	0.0056	0.0043
30	30	0.2	0.1	0.0075	0.0063	0.0061	0.0030	0.0017	0.0024
30	30	0.5	0.3	0.0030	0.0025	0.0022	0.0014	0.0018	0.0027
30	30	0.7	0.1	0.0017	0.0014	0.0012	0.0003	0.0018	0.0023
40	30	0.2	0.1	0.0070	0.0062	0.0060	0.0046	0.0012	0.0018
40	30	0.5	0.3	0.0025	0.0022	0.0020	0.0008	0.0011	0.0017
40	30	0.7	0.1	0.0015	0.0012	0.0012	0.0002	0.0014	0.0019
10	50	0.2	0.1	0.0121	0.0074	0.0071	0.0120	0.0046	0.0054
10	50	0.5	0.3	0.0066	0.0032	0.0031	0.0041	0.0119	0.0074
10	50	0.7	0.1	0.0053	0.0023	0.0019	0.0039	0.0119	0.0058
20	50	0.2	0.1	0.0083	0.0064	0.0062	0.0075	0.0018	0.0025
20	50	0.5	0.3	0.0035	0.0026	0.0023	0.0050	0.0020	0.0025
20	50	0.7	0.1	0.0023	0.0015	0.0013	0.0022	0.0028	0.0027
30	50	0.2	0.1	0.0074	0.0062	0.0060	0.0026	0.0011	0.0017
30	50	0.5	0.3	0.0029	0.0024	0.0021	0.0014	0.0011	0.0016
30	50	0.7	0.1	0.0017	0.0014	0.0011	0.0005	0.0011	0.0014
40	50	0.2	0.1	0.0069	0.0060	0.0059	0.0036	0.0007	0.0012
40	50	0.5	0.3	0.0026	0.0022	0.0020	0.0036	0.0007	0.0010
40	50	0.7	0.1	0.0014	0.0012	0.0011	0.0020	0.0007	0.0011
10	70	0.2	0.1	0.0116	0.0068	0.0067	0.0100	0.0031	0.0038
10	70	0.5	0.3	0.0064	0.0031	0.0029	0.0084	0.0060	0.0052
10	70	0.7	0.1	0.0051	0.0021	0.0017	0.0017	0.0094	0.0057
20	70	0.2	0.1	0.0084	0.0064	0.0062	0.0084	0.0011	0.0018
20	70	0.5	0.3	0.0035	0.0023	0.0023	0.0041	0.0014	0.0020
20	70	0.7	0.1	0.0023	0.0013	0.0012	0.0014	0.0019	0.0024
30	70	0.2	0.1	0.0073	0.0061	0.0059	0.0052	0.0006	0.0009
30	70	0.5	0.3	0.0027	0.0022	0.0020	0.0007	0.0007	0.0011
30	70	0.7	0.1	0.0016	0.0012	0.0011	0.0005	0.0007	0.0010
40	70	0.2	0.1	0.0069	0.0060	0.0059	0.0038	0.0005	0.0009
40	70	0.5	0.3	0.0025	0.0022	0.0020	0.0037	0.0005	0.0008
40	70	0.7	0.1	0.0014	0.0012	0.0010	0.0011	0.0005	0.0008

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0, 0.08)$; $v_{it} \sim N(0, 0.05)$; $u_{it} \sim N(0, 0.05)$; $e_{it} \sim N(0, 0.05)$.

Spatial lag variable ρ : RMSE

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	0.0018	0.0018	0.0018	0.0078	0.0979	0.0324
10	20	0.5	0.3	0.0015	0.0014	0.0015	0.0093	0.0333	0.0153
10	20	0.7	0.1	0.0015	0.0015	0.0016	0.0078	0.2125	0.0129
20	20	0.2	0.1	0.0011	0.0011	0.0011	0.0041	0.0320	0.0219
20	20	0.5	0.3	0.0008	0.0007	0.0007	0.0031	0.0100	0.0058
20	20	0.7	0.1	0.0006	0.0006	0.0006	0.0052	0.0331	0.0043
30	20	0.2	0.1	0.0005	0.0005	0.0005	0.0231	0.0111	0.0105
30	20	0.5	0.3	0.0005	0.0005	0.0005	0.0008	0.0046	0.0039
30	20	0.7	0.1	0.0003	0.0003	0.0003	0.0061	0.0054	0.0038
40	20	0.2	0.1	0.0005	0.0005	0.0005	0.0012	0.0075	0.0080
40	20	0.5	0.3	0.0005	0.0004	0.0005	0.0010	0.0020	0.0020
40	20	0.7	0.1	0.0002	0.0002	0.0002	0.0004	0.0022	0.0014
10	30	0.2	0.1	0.0015	0.0015	0.0015	0.0026	0.0393	0.0304
10	30	0.5	0.3	0.0011	0.0010	0.0010	0.0045	0.0252	0.0099
10	30	0.7	0.1	0.0010	0.0010	0.0011	0.0012	0.0307	0.0090
20	30	0.2	0.1	0.0006	0.0006	0.0006	0.0016	0.0122	0.0101
20	30	0.5	0.3	0.0006	0.0005	0.0005	0.0030	0.0082	0.0046
20	30	0.7	0.1	0.0003	0.0003	0.0003	0.0007	0.0108	0.0027
30	30	0.2	0.1	0.0004	0.0004	0.0004	0.0103	0.0061	0.0050
30	30	0.5	0.3	0.0006	0.0005	0.0005	0.0004	0.0029	0.0018
30	30	0.7	0.1	0.0002	0.0002	0.0002	0.0004	0.0027	0.0016
40	30	0.2	0.1	0.0003	0.0003	0.0003	0.0005	0.0059	0.0049
40	30	0.5	0.3	0.0004	0.0004	0.0004	0.0006	0.0016	0.0012
40	30	0.7	0.1	0.0002	0.0002	0.0002	0.0002	0.0033	0.0016
10	50	0.2	0.1	0.0009	0.0007	0.0009	0.0021	0.0140	0.0101
10	50	0.5	0.3	0.0007	0.0005	0.0006	0.0143	0.0771	0.0083
10	50	0.7	0.1	0.0006	0.0005	0.0006	0.0014	0.0174	0.0046
20	50	0.2	0.1	0.0004	0.0003	0.0003	0.0008	0.0094	0.0057
20	50	0.5	0.3	0.0005	0.0004	0.0004	0.0025	0.0032	0.0025
20	50	0.7	0.1	0.0002	0.0002	0.0002	0.0007	0.0058	0.0016
30	50	0.2	0.1	0.0003	0.0003	0.0003	0.0035	0.0042	0.0044
30	50	0.5	0.3	0.0004	0.0003	0.0003	0.0002	0.0019	0.0016
30	50	0.7	0.1	0.0001	0.0001	0.0001	0.0002	0.0016	0.0011
40	50	0.2	0.1	0.0002	0.0002	0.0002	0.0024	0.0030	0.0029
40	50	0.5	0.3	0.0004	0.0004	0.0004	0.0029	0.0012	0.0010
40	50	0.7	0.1	0.0002	0.0001	0.0002	0.0006	0.0012	0.0006
10	70	0.2	0.1	0.0005	0.0004	0.0004	0.0016	0.0143	0.0142
10	70	0.5	0.3	0.0006	0.0004	0.0005	0.0032	0.0199	0.0048
10	70	0.7	0.1	0.0004	0.0003	0.0004	0.0008	0.0121	0.0025
20	70	0.2	0.1	0.0002	0.0002	0.0002	0.0008	0.0050	0.0050
20	70	0.5	0.3	0.0005	0.0003	0.0004	0.0024	0.0024	0.0015
20	70	0.7	0.1	0.0002	0.0002	0.0002	0.0002	0.0032	0.0018
30	70	0.2	0.1	0.0002	0.0002	0.0002	0.0004	0.0026	0.0026
30	70	0.5	0.3	0.0004	0.0003	0.0003	0.0004	0.0012	0.0010
30	70	0.7	0.1	0.0001	0.0001	0.0001	0.0003	0.0009	0.0006
40	70	0.2	0.1	0.0001	0.0001	0.0001	0.0010	0.0020	0.0020
40	70	0.5	0.3	0.0004	0.0003	0.0003	0.0033	0.0007	0.0005
40	70	0.7	0.1	0.0001	0.0001	0.0001	0.0002	0.0007	0.0005

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0, 0.08)$; $v_{it} \sim N(0, 0.05)$; $u_{it} \sim N(0, 0.05)$; $e_{it} \sim N(0, 0.05)$.

Endogenous variable γ : RMSE

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	0.1733	0.1686	0.1676	0.1723	0.1073	0.1195
10	20	0.5	0.3	0.1634	0.1611	0.1593	0.1649	0.0892	0.0810
10	20	0.7	0.1	0.1603	0.1581	0.1572	0.1650	0.1134	0.0668
20	20	0.2	0.1	0.1615	0.1584	0.1579	0.1735	0.0424	0.0476
20	20	0.5	0.3	0.1499	0.1481	0.1470	0.1512	0.0329	0.0364
20	20	0.7	0.1	0.1477	0.1457	0.1448	0.1518	0.0288	0.0292
30	20	0.2	0.1	0.1549	0.1527	0.1524	0.1843	0.0248	0.0305
30	20	0.5	0.3	0.1446	0.1432	0.1425	0.1484	0.0227	0.0244
30	20	0.7	0.1	0.1431	0.1403	0.1408	0.1603	0.0173	0.0188
40	20	0.2	0.1	0.1529	0.1512	0.1509	0.1528	0.0184	0.0231
40	20	0.5	0.3	0.1441	0.1430	0.1425	0.1825	0.0142	0.0177
40	20	0.7	0.1	0.1408	0.1383	0.1387	0.1508	0.0122	0.0139
10	30	0.2	0.1	0.1754	0.1702	0.1696	0.1748	0.0634	0.0655
10	30	0.5	0.3	0.1608	0.1561	0.1569	0.1947	0.0521	0.0497
10	30	0.7	0.1	0.1596	0.1496	0.1564	0.1618	0.0468	0.0420
20	30	0.2	0.1	0.1603	0.1572	0.1567	0.1593	0.0292	0.0345
20	30	0.5	0.3	0.1500	0.1483	0.1471	0.1487	0.0188	0.0219
20	30	0.7	0.1	0.1469	0.1451	0.1439	0.1577	0.0201	0.0211
30	30	0.2	0.1	0.1549	0.1528	0.1524	0.1962	0.0152	0.0192
30	30	0.5	0.3	0.1455	0.1440	0.1434	0.1611	0.0110	0.0137
30	30	0.7	0.1	0.1433	0.1420	0.1408	0.1569	0.0115	0.0123
40	30	0.2	0.1	0.1540	0.1524	0.1521	0.1592	0.0125	0.0149
40	30	0.5	0.3	0.1433	0.1423	0.1417	0.1551	0.0094	0.0115
40	30	0.7	0.1	0.1418	0.1404	0.1396	0.1488	0.0080	0.0089
10	50	0.2	0.1	0.1691	0.1633	0.1632	0.1678	0.0354	0.0372
10	50	0.5	0.3	0.1597	0.1499	0.1555	0.1568	0.0365	0.0358
10	50	0.7	0.1	0.1560	0.1525	0.1528	0.1587	0.0254	0.0239
20	50	0.2	0.1	0.1601	0.1569	0.1565	0.1610	0.0146	0.0183
20	50	0.5	0.3	0.1491	0.1469	0.1462	0.1478	0.0129	0.0146
20	50	0.7	0.1	0.1456	0.1387	0.1426	0.1457	0.0098	0.0105
30	50	0.2	0.1	0.1552	0.1530	0.1527	0.1616	0.0097	0.0130
30	50	0.5	0.3	0.1449	0.1437	0.1428	0.1592	0.0075	0.0090
30	50	0.7	0.1	0.1420	0.1408	0.1396	0.1475	0.0065	0.0069
40	50	0.2	0.1	0.1531	0.1514	0.1511	0.1572	0.0066	0.0087
40	50	0.5	0.3	0.1434	0.1424	0.1418	0.1429	0.0057	0.0067
40	50	0.7	0.1	0.1409	0.1399	0.1387	0.1597	0.0047	0.0052
10	70	0.2	0.1	0.1691	0.1623	0.1633	0.1708	0.0252	0.0281
10	70	0.5	0.3	0.1605	0.1548	0.1563	0.1576	0.0224	0.0227
10	70	0.7	0.1	0.1539	0.1510	0.1503	0.1650	0.0186	0.0190
20	70	0.2	0.1	0.1594	0.1562	0.1558	0.1589	0.0107	0.0140
20	70	0.5	0.3	0.1494	0.1420	0.1464	0.1512	0.0081	0.0098
20	70	0.7	0.1	0.1457	0.1378	0.1426	0.1489	0.0076	0.0086
30	70	0.2	0.1	0.1557	0.1535	0.1532	0.1592	0.0061	0.0076
30	70	0.5	0.3	0.1451	0.1437	0.1430	0.1601	0.0053	0.0067
30	70	0.7	0.1	0.1426	0.1400	0.1403	0.1484	0.0044	0.0047
40	70	0.2	0.1	0.1527	0.1510	0.1508	0.1570	0.0048	0.0063
40	70	0.5	0.3	0.1425	0.1415	0.1408	0.1410	0.0037	0.0045
40	70	0.7	0.1	0.1407	0.1397	0.1389	0.1421	0.0032	0.0035

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0, 0.08)$; $v_{it} \sim N(0, 0.05)$; $u_{it} \sim N(0, 0.05)$; $e_{it} \sim N(0, 0.05)$.

Exogenous variable β : RMSE

T	N	α	ρ	SMLE	SDMLE	SDQMLE	LSDV	DIF-GMM	SYS-GMM
10	20	0.2	0.1	0.0061	0.0053	0.0051	0.0055	0.0131	0.0386
10	20	0.5	0.3	0.0051	0.0047	0.0044	0.0049	0.0119	0.0169
10	20	0.7	0.1	0.0046	0.0044	0.0042	0.0047	0.0355	0.0335
20	20	0.2	0.1	0.0049	0.0041	0.0040	0.0028	0.0050	0.0120
20	20	0.5	0.3	0.0031	0.0027	0.0025	0.0028	0.0050	0.0101
20	20	0.7	0.1	0.0024	0.0021	0.0019	0.0018	0.0053	0.0095
30	20	0.2	0.1	0.0044	0.0038	0.0037	0.0042	0.0034	0.0072
30	20	0.5	0.3	0.0026	0.0023	0.0022	0.0030	0.0032	0.0066
30	20	0.7	0.1	0.0021	0.0018	0.0017	0.0014	0.0027	0.0075
40	20	0.2	0.1	0.0039	0.0034	0.0034	0.0036	0.0024	0.0055
40	20	0.5	0.3	0.0023	0.0021	0.0020	0.0058	0.0024	0.0059
40	20	0.7	0.1	0.0018	0.0016	0.0015	0.0010	0.0024	0.0054
10	30	0.2	0.1	0.0056	0.0045	0.0043	0.0055	0.0080	0.0142
10	30	0.5	0.3	0.0037	0.0031	0.0031	0.0050	0.0082	0.0177
10	30	0.7	0.1	0.0029	0.0024	0.0025	0.0029	0.0088	0.0128
20	30	0.2	0.1	0.0042	0.0035	0.0034	0.0041	0.0033	0.0080
20	30	0.5	0.3	0.0022	0.0019	0.0017	0.0021	0.0034	0.0071
20	30	0.7	0.1	0.0022	0.0019	0.0017	0.0016	0.0033	0.0076
30	30	0.2	0.1	0.0039	0.0034	0.0033	0.0023	0.0021	0.0044
30	30	0.5	0.3	0.0021	0.0019	0.0017	0.0017	0.0019	0.0059
30	30	0.7	0.1	0.0018	0.0016	0.0014	0.0010	0.0020	0.0048
40	30	0.2	0.1	0.0038	0.0034	0.0033	0.0028	0.0016	0.0027
40	30	0.5	0.3	0.0020	0.0018	0.0016	0.0010	0.0015	0.0031
40	30	0.7	0.1	0.0017	0.0015	0.0014	0.0008	0.0014	0.0044
10	50	0.2	0.1	0.0043	0.0032	0.0032	0.0043	0.0042	0.0070
10	50	0.5	0.3	0.0027	0.0020	0.0020	0.0019	0.0043	0.0097
10	50	0.7	0.1	0.0021	0.0017	0.0017	0.0021	0.0047	0.0088
20	50	0.2	0.1	0.0037	0.0030	0.0029	0.0033	0.0020	0.0051
20	50	0.5	0.3	0.0020	0.0016	0.0015	0.0020	0.0018	0.0045
20	50	0.7	0.1	0.0017	0.0013	0.0012	0.0016	0.0019	0.0050
30	50	0.2	0.1	0.0036	0.0030	0.0030	0.0014	0.0012	0.0029
30	50	0.5	0.3	0.0019	0.0016	0.0015	0.0014	0.0012	0.0032
30	50	0.7	0.1	0.0015	0.0013	0.0011	0.0008	0.0011	0.0030
40	50	0.2	0.1	0.0035	0.0030	0.0030	0.0018	0.0009	0.0023
40	50	0.5	0.3	0.0019	0.0017	0.0016	0.0017	0.0008	0.0020
40	50	0.7	0.1	0.0014	0.0012	0.0011	0.0025	0.0008	0.0027
10	70	0.2	0.1	0.0037	0.0026	0.0025	0.0032	0.0030	0.0050
10	70	0.5	0.3	0.0021	0.0016	0.0015	0.0020	0.0029	0.0071
10	70	0.7	0.1	0.0016	0.0013	0.0012	0.0014	0.0036	0.0075
20	70	0.2	0.1	0.0036	0.0029	0.0028	0.0036	0.0014	0.0035
20	70	0.5	0.3	0.0019	0.0013	0.0013	0.0016	0.0012	0.0037
20	70	0.7	0.1	0.0014	0.0010	0.0010	0.0011	0.0012	0.0041
30	70	0.2	0.1	0.0034	0.0029	0.0028	0.0025	0.0009	0.0018
30	70	0.5	0.3	0.0018	0.0015	0.0013	0.0007	0.0009	0.0022
30	70	0.7	0.1	0.0014	0.0011	0.0010	0.0007	0.0008	0.0021
40	70	0.2	0.1	0.0035	0.0030	0.0030	0.0019	0.0006	0.0017
40	70	0.5	0.3	0.0017	0.0015	0.0014	0.0015	0.0006	0.0015
40	70	0.7	0.1	0.0013	0.0012	0.0010	0.0011	0.0006	0.0020

Notes: The DGP are generated according to the expressions (15)-(18) based on a Rook-type spatial weight matrix of order 1 with $\beta = 1$; $\delta = 0.65$; $\gamma = 0.5$; $\lambda = 0.45$; $\psi = 0.25$; $\theta = 0.6$; $\eta_i \sim N(0,0.08)$; $v_{it} \sim N(0,0.05)$; $u_{it} \sim N(0,0.05)$; $e_{it} \sim N(0,0.05)$.

Chapter 2

Pollution Havens: a Spatial Vector Auto-Regression Approach

1 Introduction

There is an ongoing debate about the environmental impact of multinationals on host countries, in particular developing economies. On the one hand, the pollution haven hypothesis (PHH) states that trade/capital liberalization will lead pollution-intensive industries to relocate to countries with less stringent environmental policies. Competition to attract FDI among economies might then lead to a reduction or status quo in environmental standards. On the other hand, according to the pollution halo hypothesis, relocating industries can improve environmental performance in the countries with laxer environmental standards through new and cleaner technology. Empirical studies designed to test these hypotheses have so far shown little and conflicting evidence. One of the crucial problems inherent in existing studies is that no specific causality analysis of the mutual relationship between FDI, environmental policy and capital/trade openness has been addressed. The most popular approaches consist of either assuming environmental regulation as a function of FDI/trade or vice-versa. Only recently, Ederington and Minier (2003), Cole et al. (2006, 2009), Levinson and Taylor (2008) and Kellenberg (2009) showed the theoretical and empirical importance of endogenizing environmental stringency with respect to FDI/Trade.

Disentangling which relationship(s) is (are) more prevalent is essentially an empirical question. In order to do so, environmental stringency, inward FDI and openness are analyzed through a spatial panel vector autoregression (SpVAR) model for 124 host countries during the period 1993 to 2000. The VAR framework provides a flexible statistical tool where dynamic relations can be determined and quantified based on a minimal theoretical structure. In addition, unlike traditional panel VAR, SpVAR allows to explicitly account for spatial cross-section dependence in the data. This is particularly important for two interrelated reasons. First, environmental policies, FDI allocation as well as trade openness have been shown to be spatially correlated (Eliste and Fredriksson, 2004; Blonigen et al., 2007; Egger and Larch, 2008). Second, taking into account spatial dependence and spillovers can potentially mitigate the variable omissions bias. To my knowledge, this study is the first to examine carefully the temporal and spatial dynamics through which these three variables interact at the world-wide level.

From a methodological perspective, this study presents several innovative features. First, the generalized empirical likelihood (GEL) estimator is applied to a spatial VAR model. Second, because ignoring the possibility of non-stationary parameters can lead to unreliable inference, second generation panel unit root tests, which explicitly account for heterogeneity and cross-section dependence, are implemented. Last but not least, the role of spatial spillovers is highlighted through spatial complements to panel Granger causality and impulse-response functions.

Results highlight the existence of a bidirectional causality between environmental regulation (proxied by SO_2/GDP emissions) and inflows of FDI in the presence of spatial dependence. In addition, I am able to shed some light on the spatial relationships between the variables not found in the classical VAR. Spatial spillovers do play a key role in the relationship between FDI, environmental stringency and trade openness and confirm the presence of high globalization and strategic decisions. Overall, these findings suggest that the analysis of the linkages between FDI and the environment should explicitly account for the potential endogeneity and spatial nature of the variables in order to reduce variable omissions. Once that is done, the pollution have effect and the pollution halo hypothesis are both confirmed, in particular for low income countries.

The organization of the paper is as follows. Section 2 highlights the weaknesses found in the existing empirical literature about pollution havens. Section 3 presents the spatial VAR and the GEL estimator. Section 4 proceeds with the description of the data and the presentation of the panel unit root tests as well as the estimation results. Section 5 concludes.

2 Pollution Haven Survey

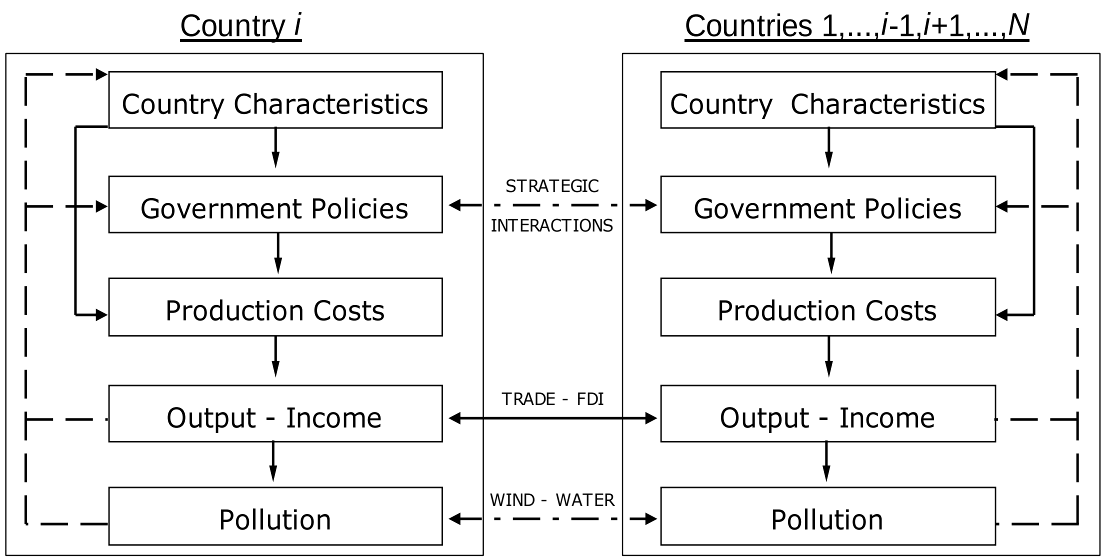
The pollution haven hypothesis refers to the possibility that FDI could be sensitive to weaker environmental regulation, especially when polluting firms want to avoid the costs associated with environmental standards compliance. There is a growing theoretical and empirical literature on the linkages between environmental stringency, FDI/trade and globalization.

Initially, pollution havens were seen as the results of differences in the countries' comparative advantages in the production of pollution intensive goods. Copeland and Taylor (2003) go further and determine the strategic interaction between North and South countries from an autarky situation to free trade when environmental stringency is allowed to be income-induced, thus endogenous. They identify three types of impacts of FDI/trade on the environment. First, the scale (or volume) effect postulates that freer capital and trade flows lead to more production, which in turn means more pollution and harm for the environment. Second, the technique effect assumes that higher income leads to more stringent environmental policies which motivate the adoption of cleaner production technologies. Thus, this effect might mitigate the negative impact of globalization on the environment. Third, the composition effect presumes that some countries, based on their lower level of environmental stringency, are more likely to specialize in pollution-intensive activities than others. This last effect is the mechanism that is argued in the pollution haven hypothesis.

In a theoretical article, Fredriksson (1999) demonstrates that in a perfect competitive market, trade liberalization increases (reduces) both firms and environmental lobby groups' incentive to influence the level of environmental stringency if the country displays a comparative advantage (disadvantage) in the pollution-intensive industry. Consequently, the effect of FDI/trade on the environmental policy-making will depend on the relative changes in political pressures. Cole et al. (2006) show that the effect of capital and trade flows on the environmental standards depends on the host country's level of corruption. In highly corrupted countries, environmental policy is partially determined by bribery and FDI can lead more easily to a reduction in the level of environmental standards or a lack of policies enforcement. In line with this mechanism, Cole and Fredriksson (2009) argue that the impact of FDI on environmental policy-making will be conditioned on the host country's political institutions. In the presence of few (many) legislative units and high (low) corruption, FDI will lead to less (more) stringent environmental policy and thus contribute to sustain (mitigate) the creation of a pollution haven.

Figure 1 depicts the linkages between environmental policy and trade/FDI in an interdependent world. Growing international capital and trade flows have given rise to global interdependency of the economies and have strengthened the globalization process. In particular, government policies, including environmental and trade policies, are not only determined by the country's endowed characteristics but are also the product of strategic interactions among economies (Eliste and Fredriksson, 2004; Egger and Larch, 2008). The production costs are determined by the country's endowments and government policies, and determine themselves the level of domestic production/income which is created. Multinationals enterprises will also shape the domestic production through different channels including higher competition, technological diffusion and agglomeration externalities. In fact, one of the consequences of international capital mobility is that some types of FDI decisions to invest to different host countries are no longer independent of each other. Policy makers may also be tempted to integrate the effects on FDI/trade flows into their regulations, because production and trade, through national and multinational firms, impact the country's environment and endowment through scale, composition and technique effects. This can be further exacerbated by the spatial diffusion of regional/global pollutants emissions through wind and water. In particular, the level of SO_2 and CO_2 emissions in one given country is partially determined by the geographical location and wind patterns of emissions in the neighboring countries.

Figure 1: FDI-Pollution Haven Linkages



Source: Extended version of Kellenberg (2009)

Even though the intuitive idea behind the PHH is rather simple, its empirical analysis has given rise to some difficulties. This is linked to the confusion between two types of empirical findings. On one hand, the pollution haven effect is empirical evidence showing that differences in the level of environmental stringency have an impact on the distribution of pollution industries among countries. In this perspective, the cost of environmental regulation compliance constitutes an additional location determinant. On the other hand, the pollution haven hypothesis is empirical evidence attesting that a reduction in trade and capital restrictions leads to a relocation of dirty industries from countries with stringent environmental regulations to those with laxer standards. Therefore, the underlying assumption of the PHH is that the most important location determinant is environmental regulation and pollution intensity. Without necessarily explicitly stating it, most empirical studies actually measure the pollution haven effect, which is a necessary but not a sufficient condition for the PHH to hold. The main reason is that the PHH is a second order effect, which is by definition more difficult to measure.

Beside this terminology confusion, the controversy regarding the pollution haven effect is generated, in part, by a lack of convincing empirical evidence. Most studies, which analyze inflows of FDI to different countries originating from various home countries at the aggregated or industrial level, fail to disentangle any evidence of a pollution haven effect[21]. The traditional approach, corresponding to the solid arrows in Figure 1, considers the allocation of FDI as a function of environmental regulation (Spatareanu, 2007). The alternative approach considers the dashed arrows depicted in Figure 1 and estimate environmental regulation as a function of the level of FDI (Cole et al., 2006). More recently, a limited number of papers combine both approaches to address the potential endogenous nature of the pollution haven effect (Cole et al., 2007). Hoffmann et al. (2005) study whether FDI/pollution Granger causes pollution/FDI using a meta analysis of Granger causality tests with short time series and panel data. They observe that in low-income countries CO_2 emissions Granger-cause net inflows of FDI, while for middle-income countries FDI Granger-causes the level of CO_2. For high-income countries, there is no Granger causality. Based on these findings, the authors argue that the pollution haven effect is more likely to happen in low-income host countries.

The main drawback of this study is to deliberately ignore the dashed-dotted arrows in Figure 1 and assume no interdependence in economic and policy decisions as well as pollution emissions. As emphasized in the spatial econometrics literature, failure to account for spatial dependence may lead to biased inference. This may be particularly problematic for the empirical analysis of the PHH, because it is essential to identify the direction of causality between inflows of FDI, the level of environmental stringency and trade liberalization.

[21] Two other types of studies can be found in the empirical literature. The first strand considers inflows of FDI to a single country at the regional and/or industrial level. The second strand considers outflows of FDI from a single home country to one or more host countries at the aggregated or industrial level.

3 Spatial Panel Vector Autoregression

The most popular method to examine the relationships of several variables with total flexibility is the vector autoregression analysis (VAR) (Sims, 1980). VAR is a generalized form of an autoregressive model in which the evolution of each variable is a function of its previous values and previous observations of all other series in the system. Theoretically (panel) VAR is able to potentially address the set of possible relationships between variables without relying on a prior theoretical structure beside the assumption of independence between cross-sections. A spatial vector autoregressive model (SpVAR) consists of an extended version of a VAR which includes spatial as well as temporal lags among a vector of stationary state variables. The spatial extension is justified by VAR's inability to explicitly consider the potential impacts of economic events in neighboring countries. In standard VAR, an economic shock that occurs in a given country only influences the economic conditions in that nation. All of the impacts prevail within its frontiers and thus automatically exclude spatial spillovers. In reality, a shock in a given country is likely to affect the economic conditions in the neighboring countries as well. That is, the economic conditions in the neighborhood countries are likely to exhibit co-movement over time.

A limited number of studies have extended spatial econometrics to VAR analysis. LeSage and Pan (1995) apply spatial contiguity as an alternative prior in a Bayesian VAR model. They show that incorporating a spatial contiguity structure dramatically lowers forecast error. Chen and Conley (2001) propose to specify the spatial dependence of the endogenous dependent variables in a semi-parametric framework and estimate the parameters with a two-step sieve least-squares procedure. Di Giacinto (2003) derives a structural VAR model applying spatial relationships to follow prior theoretical beliefs. Recently, Beenstock and Felsenstein (2007) present a flexible SpVAR model that builds on the spatial autoregressive model with a spatial error process. They demonstrate that the inclusion of spatially and temporally correlated disturbances is problematic, because the structural parameters are not fully identified. To avoid this issue, Mutl (2009) extends Binder et al.'s (2005) framework to include a single spatial dependent error variable. Most of these spatial econometrics techniques, however, have been the object of limited applied works.

For simplicity, I restrict the presentation to a spatial panel VAR of order one, SpVAR(1), the extension to higher order being straightforward. Let y_{it} be a $k \times 1$ vector of dependent variables and μ_i a vector of individual effects of the i^{th} cross-sectional unit at time t. The SpVAR(1) is expressed in panel ($t = 2,...,T;\ i = 1,...,N$) as follow:

$$
\begin{aligned}
Y_t &= \tilde{\Phi} Y_{t-1} + \tilde{\Pi} \tilde{W} Y_{t-1} + \mu + \varepsilon_t \\
Y_t &= \tilde{\Psi} Y_{t-1} + \mu + \varepsilon_t
\end{aligned}
\tag{1}
$$

where $Y_t = [y_{1t},\ldots,y_{it},\ldots,y_{Nt}]'$, $\tilde{\Phi} = (I_N \otimes \Phi)$, $\tilde{\Pi} = (I_N \otimes \Pi)$, $\tilde{W} = (W \otimes I_k)$, I_N is a $N \times N$ identity matrix and $\otimes$ is the Kronecker product. When the variables are grouped, $X_{t-1} = \left[Y_{t-1}, \tilde{W} Y_{t-1}\right]$ and $\Psi = \left[\bar{\Phi}, \tilde{\Pi}\right]$. To simplify the notation, each element of the $\tilde{W} Y_{t-1}$ vector corresponds to $\dot{Y}_{t-1}$. The lag coefficients are given by the $k \times k$ matrix Φ, while the lagged spatial coefficients are associated with the $k \times k$ matrix Π. The individual effects are included in the matrix μ. W is a $N \times N$ spatial weight matrix, which defines the spatial arrangement of the observations. Its diagonal elements are set to zero, while its off-diagonal components are a spatial function between any two pair of locations. The theoretical foundation for W is quite general and the particular functional form of any single element in W is, therefore, not prescribed but has to be exogenous to the model. Although I experiment with alternative spatial weights (e.g. contiguity, inverse distance, ...), the main results are based on a negative exponential scheme: $w_{ij} = \exp\left(-d_{i,j}/500\right)$ *if* $i \neq j$, where $d_{i,j}$ represents the distance between locations *i* and *j* (in km). Each row of the spatial weight matrix is standardized so that all elements sum to one. Therefore, the lagged spatial lag parameters in Π represent the averages values of the neighbor countries in the previous period, i.e. lagged "third-country" effects. The inclusion of these spatial variables accounts for the fact that a shock in a given country might need some time to extend over its neighborhood. This spatial process can play a key role in spatial VAR impulse response functions and Granger causality. Note that model (1) does not include a current spatial lag ($\tilde{W} Y_t$), because it is not well suited for forecasting purposes.

Unlike in standard panel VAR framework, the spatio-temporal stationarity of the SpVAR relies on stringer conditions. Assuming the spatial weight matrix is row-standardized, the stability of the system requires that the eigenvalues λ of the matrix $\Xi = \Phi + \Pi$ are smaller than one, namely $\det(\Xi - \lambda I) = 0$ with $|\lambda| \leq 1$. A necessary but not sufficient condition for the spatio-temporal stationarity to hold is that $\left| diag(\Phi + \Pi) \right| \leq 1$. Theoretically, the model would remain stationary if one or more autoregressive terms are explosive as long as they are compensated by their spatial extensions. Appendix 6.1 provides further insight on this issue.

3.1 Generalized Empirical Likelihood

The estimation of a panel VAR similar to (1) without spatial features ($\Pi = 0$) is a well-studied problem. First, when *T* is small and fixed, it is necessary to impose some restrictions on the initial observations. Second, it is important to mitigate the homogeneity bias by specifying the unobserved individual effects. Beenstock and Felsenstein (2007) discuss the choice of fixed *versus* random effects in spatial panels. If the sample happens to be the population (e.g. countries), individual effects should be fixed because each panel cross-section represents unrandomly itself. Despite being subject to the classical incidental parameters problem, models with fixed effects are more likely to be robust to possible misspecification of the distribution of the specific individual effects. This explains why in practice, most empirical panel VAR applications are estimated in a fixed effects setting with a Generalized Method of Moments estimator (GMM) (Love and Zicchino, 2006).

It is convenient to differentiate between the standard (difference) GMM estimators suggested by Arellano and Bond (1991), and their extended versions proposed by Arellano and Bover (1995), and Blundell and Bond (1998). Difference GMM relies on orthogonality conditions between the disturbances in first-difference and the lagged values of the variables:

$$E[g(\Psi, Y_{it})] = E[(\Delta Y_{it} - \Psi \Delta X_{i,t-1}) Y'_{i,t-s}] = 0, \ t = 3,...,T; \ 2 \leq s \leq t-1. \quad (2)$$

$$E[g(\Psi, Y_{it})] = E[(\Delta Y_{it} - \Psi \Delta X_{i,t-1}) \dot{Y}'_{i,t-s}] = 0, \ t = 3,...,T; \ 1 \leq s \leq t-1. \quad (3)$$

Besides being asymptotically inefficient, standard GMM suffers from weak instrument problem when the data is persistent. It also breaks down, when the data is not stationary (Binder et al., 2005). To overcome these drawbacks, several extended GMM estimators have been proposed. In particular, Blundell and Bond (1998) propose a system GMM estimator that relies on the following additional moment conditions:

$$E[g(\Psi, Y_{it})] = E[(Y_{it} - \Psi X_{i,t-1}) \Delta Y'_{i,t-1}] = 0, \ t = 3,...,T \quad (4)$$

$$E[g(\Psi, Y_{it})] = E[(Y_{it} - \Psi X_{i,t-1}) \dot{Y}'_{it}] = 0, \ t = 2,...,T. \quad (5)$$

This system of moment conditions makes extended GMM more efficient and less biased than standard GMM. In addition, it remains consistent in the presence of unit roots. However, one of the main drawbacks of system GMM, and GMM in general, is to be particularly sensitive to the set of moment conditions. In fact, a high degree of over-identification is problematic because GMM estimators assign the same weight to all moment conditions over the sample, although some moments conditions convey more information than others.

To address this issue, Newey and Smith (2004) propose a generalized empirical likelihood estimator (GEL) that assesses the orthogonality conditions by assigning a probability weight to each observation. More specifically, GEL generates a discrete multinomial distribution which yields the highest probability of the observed data subject to a set of moment restrictions:

$$\hat{\Psi},\{\hat{p}_i\} = \arg\max_{\Psi,\{p_i\}} \sum_{i=1}^{N} \log(p_i) \qquad (6)$$
$$s.t. \qquad \sum_{i=1}^{N} p_i g(\Psi, Y_{it}) = 0,\ \sum_{i=1}^{N} p_i = 1,\ \ p_i \geq 0$$

where p_i represents the individual probability associated with the individual observation y_i and $g(\Psi, Y_{it}) = g(\Psi, Y_{it}; W)$ is the orthogonality conditions given in (2) and (3). Newey and Smith (2004) show that although GMM and GEL possess identical first-order asymptotic properties, the non-parametric maximum likelihood estimator is able to remove several sources of GMM biases, including the correlation between the moment function and its derivative. These properties explain why GEL displays a higher order of efficiency than GMM. On a practical note, the underlying optimization process of GEL can sometimes fail to find a solution, which is not the case with the GMM estimators. In addition, traditional GMM estimators are less time consuming than GEL estimators. The drawbacks associated with GEL are, however, expected to be overweighted by its higher order of efficiency and flexibility. When there is no convergence or no possible solution with the GEL, I turn to the Continuously Updated GMM estimator (CUE), which is a special case of the GEL when each individual probability is given the same weight ($p_i = \frac{1}{N}$, i = 1, …, N).

4 Empirical Results

Before displaying the main empirical findings, I implement different panel unit root tests to investigate the stationarity condition necessary to make any inference. But first, I briefly discuss the selection of the variables considered in this study which covers up to 124 countries between 1993-2000.

4.1 Variables Selection

In its simplest form, the pollution haven effect is the result of changes in capital and trade flows due to modifications in environmental regulation. Therefore the focus on the empirical analysis should be on the difference between environment policy instruments across countries and how this alters capital and trade flows. Finding a dynamic measure of environmental stringency is a difficult task. I restricted my search over environmental measures that have both within and between-country variation in order to control for important unobservable factors that may influence the level of stringency. The choices are severely limited, as most environmental regulatory indices at the country level consists of cross-sectional estimates based on one year of data. In the end, I consider sulphur dioxide (SO_2) emissions, whose reductions in emissions may be viewed as proxies for a host country's effective enforcement of environmental policies. Data on SO_2 emissions are taken from Stern (2005) and expressed as (SO_2/GDP) for ease of interpretation. As an output-oriented measure, SO_2 emission should be a better measure of environmental regulation, because it does not only take into account the level of stringency but also its actual enforcement. This is important because the level of stringency is usually less time varying than its actual compliance. In fact, output-oriented measures should, theoretically, be able to account for the existence of subsidies or policy instruments that can offset the effect of stringer regulation. Note that for some countries, the SO_2/GDP ratio is closely associated with the country's economic structure. In particular, Chile and Peru appear as the dirtiest countries, while Bolivia and Argentina are among the cleanest. To address this issue, I estimate the SpVAR without these outliers as a robustness check. In addition, I consider two other measures linked to environmental policies. The CO_2/GDP ratio is taken from the World Development Indicators, while the number of ratified international environmental treaties is constructed based on the Environmental Treaties and Resource Indicators database.

Data on FDI, expressed as the ratio of FDI/GDP, is collected from the UNCTAD website. In order to maximize the number of covered countries with balanced observations, the analysis will focus on FDI flows rather than stock. Given that openness of an economy is difficult to quantify, I follow the literature and use the ratio of exports and imports over GDP given in the World Development Indicators website. Other proxies which measure the country's regulatory degree of capital account openness (*de jure* openness) are available. Unfortunately, for several countries the data is non-time varying enough which leads to multicollinearity issues affecting the estimation of the remaining coefficients. That is why these data are not considered in this study.

4.2 Panel Unit Root Tests

Stationarity in each variable leads to straightforward interpretation and inference of the VAR results. Thus, before estimating the model, careful analysis of the dynamic properties of the variables is needed. It is a well established result that univariate tests such as the Dickey-Fuller (DF) and the Augmented DF (ADF) test (Dickey and Fuller, 1981) are not powerful enough to reject the null of a unit root given the short span of many time series. The applied literature has suggested to employ panel unit root tests in order to increase the power of the tests by exploiting a large number of cross-section units and overcome this issue.

Although most panel unit root tests are based on different underlying specifications, the most general framework one can consider is given by:

$$
\begin{aligned}
y_{it} &= \phi_i y_{i,t-1} + f(x_{it}) + u_{it} \\
\Delta y_{it} &= (\phi_i - 1) y_{i,t-1} + f(x_{it}) + u_{it}
\end{aligned}
\tag{7}
$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T_i$ and $f(\cdot)$ is a (linear) function (constant and/or trend). The presence of a unit root in the panel can be tested in two ways. First, one can assume a homogenous alternative hypothesis, where all individual series are stationary: $H_0 : \phi_i = 1$ for all i *versus* $H_1 : \phi_i = \phi < 1$ for all i. The second possibility is to consider a heterogeneous alternative, where some individual time-series are potentially non-stationary: $H_0 : \phi_i = 1$ for all i *versus* $H_1 : \phi_i < 1$ for some i.

Most of the first generation of panel unit root tests considers a heterogeneous alternative (Im et al., 2003; Maddala and Wu, 1999; Choi, 2001; Hadri and Larsson, 2005; Moon et al., 2007), except Levin et al.'s (2002) and Breitung (2000) tests. Yet, the common point of all these tests is to assume cross-section independence, i.e. independent distribution for each individual time series in the panel. Although this admittedly very strong assumption simplifies the derivation of the asymptotic distributions of the panel unit root and stationarity tests, failing to account for cross-section dependence can seriously bias the test when the degree of dependence is sufficiently large. This is an issue, because a large amount of literature provides evidence of co-movements between economic variables through common observed and unobserved factors, spatial spillovers effects, or general residual correlation that remains after controlling for common influences.

Table 1: Cross-section and Spatial Dependence Tests

		FDI	SO_2	Openness
Pesaran *CD*	No Constant	1.286*	16.876***	19.863***
	Constant	12.053***	-0.838	19.812***
	Trend	4.310***	4.883***	14.390**
Moran *I*	1993	-0.090*	0.088*	0.168***
	1997	0.076*	0.047	0.175***
	2000	0.132***	0.051	0.203***
Geary *C*	1993	-	0.869*	0.891*
	1997	-	0.926	0.884***
	2000	-	0.863*	0.868**

Note: The Pesaran (2007) test checks cross-section independence, while the Moran *I* and Geary *C* check the absence of global spatial correlation. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 1 reports the results of testing for the absence of spatial and cross-sectional dependence. Independently of the specification, the Pesaran's CD test tends to reject the null of cross-sectional independence. The Moran *I* and Geary's *C* statistics also reject the null hypothesis of no global spatial autocorrelation for standard level of confidence for almost each period for FDI and trade openness, but to a less extent for SO_2. Note that Geary test cannot be implemented with FDI inflows because the test is only valid with positive values. Both spatial tests suggest that the variables are characterized by positive spatial correlation through agglomeration effects. In other words, the value taken on by the variables at each location tends to be similar to the values taken on by the variables at spatially close countries. The only exception is for the Moran test for FDI in 1993, where there seems to be negative spatial dependence. Overall, the tests tend to rule out the assumption of strong and weak (spatial) cross-section independence.

The second generation of heterogeneous panel unit root tests relaxes the cross-sectional independence hypothesis (Hurlin and Venet, 2007). Within this new generation of tests, two general streams can be distinguished. The first approach relies on a factor structure specification, where the common and idiosyncratic components influence all the cross-section units in a linear dynamic factor model (Bai and Ng, 2004; Moon and Perron, 2004; Choi, 2006; Pesaran, 2007). The second approach relies on few or no restrictions on the residual covariance matrix (Chang, 2002, 2004; Breitung and Das, 2005; Demetrescu et al., 2006). The most popular second generation panel unit root tests are presented in the next sub-sections. Since most tests are still not available in econometric softwares, I implemented them in Matlab. Note that some tests allow for the presence of a trend, others do not. The remaining tests, including the first generation of panel unit root tests, are reported in Appendices 6.6 and 6.7. The conclusions based on the second generation of panel unit root tests are not clear-cut and thus should be treated carefully.

4.2.1 Bai and Ng (2004)

Bai and Ng (2004) procedure is based on the decomposition of the process y_{it} into three components, a deterministic element, a common element captured by a factor structure and an idiosyncratic error term:

$$y_{it} = \alpha_i + \lambda_i' F_t + u_{it} \tag{8}$$
$$u_{it} = \phi_i u_{i,t-1} + e_{it}$$

where λ_i and F_t represents a $r \times 1$ vector of factor loadings and common factors, respectively, and e_{it} is the i.i.d. idiosyncratic error terms.

Rather than testing for a panel unit root directly in the process y_{it}, the Panel Analysis of Nonstationary and Idiosyncratic Common Components method (PANIC) proposed by Bai and Ng consists of testing for the presence of a unit root in the common factors and the idiosyncratic error terms separately. The PANIC approach begins with formulating the model (8) in first-difference with the assumption that $E(\Delta F_t) = 0$. Once the r common factors and the corresponding loading factors are estimated by the principal component analysis (PCA) method, the unit root tests are performed on the following re-cumulated variables for $m = 1,...,\hat{r}$ and $i = 1,...,N$:

$$\hat{F}_{m,t} = \sum_{s=2}^{t} \Delta \hat{F}_{m,s} \qquad \hat{u}_{i,t} = \sum_{s=2}^{t} \Delta \hat{u}_{i,s}$$

The unit root test on the defactored cumulated idiosyncratic error terms ($\hat{u}_{i,t}$) is based on the average individual ADF t -statistics. Bai and Ng propose two Fisher's type statistics, $P_{\hat{u}}^{c}$ and $Z_{\hat{u}}^{c}$, as suggested by Maddala and Wu (1999) and Choi (2001). Both pooled statistics converge to a standard normal distribution under the null.

The unit root test on the defactored cumulated common factors ($\hat{F}_{m,t}$) depends on the estimated number of common factors $\hat{r}$. If there is a single common factor ($\hat{r}$ = 1), the nonstationarity test is based on a standard ADF tests with a constant. When there are several common factors ($\hat{r}$ >1), Bai and Ng suggest a sequential procedure to test the number of common independent stochastic trends in the common factors. The first test, MQ_f, analyzes the common factors assuming they can be represented by a finite order VAR. The second test, MQ_c, addresses serial correlation by non-parametrically estimating the nuisance parameters. Since the asymptotic distributions of both tests are non-standard, the authors provide 10%, 5% and 1% critical values for different number of common factors.

Table 2 displays the results of the PANIC analysis. The optimal number of the common factors is selected according to the robust procedure proposed by Alessi et al. (2008), who show that the standard procedure suggested by Bai and Ng usually leads to select the highest number of common factors available. There are between four and one common factors such that the idiosyncratic terms are independent across each cross-section. Most tests fail to reject the null of unit root tests. The only exception is the unit root test of the single common factor of FDI. Independently of the test considered, MQ_c or MQ_f, the number of common stochastic trends corresponds to the number of common factors. Overall, the Bai and Ng procedure tends to suggest that the non-stationary property of the three variables is mostly due to the idiosyncratic components rather than the common factors (growth trend).

4.2.2 Moon and Perron (2004)

The method proposed by Moon and Perron (2004) is different from Bai and Ng as it assumes a factor structure in the error term:

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it} \\
x_{it} &= \phi_i x_{i,t-1} + u_{it} \\
u_{it} &= \lambda_i' F_t + e_{it}
\end{aligned}
\tag{9}
$$

where λ_i and F_t represent a $r \times 1$ vector of factor loadings and common factors, respectively. The i.i.d. idiosyncratic component is represented by e_{it}.

Similar to Bai and Ng, the Moon and Perron approach consists of defactoring the data to eliminate the cross-sectional dependence represented by the common factors. The panel unit root is then performed on the defactored estimated idiosyncratic error terms. The authors derive two modified t-statistics, t_a and t_b, which are based on the cross-sectional average of the long-run variances of the idiosyncratic residuals. The authors demonstrate that, under the null hypothesis, both statistics converge to a standard normal distribution.

The results corresponding to the Moon and Perron test are reported in Table 3. Depending on the criteria considered, the number of common factors ranges from four to one. The null hypothesis of non-stationartiy of the idiosyncratic components is rejected for all variables when the model is specified with individual effects. However, this is no longer the case when a time trend is included in the model specification. Since, Moon and Perron show that their test has low power when the model includes a time trend, these last results should be taken cautiously. In any case, these results, ruling in favour of the absence of a unit root, are in contradiction with the conclusions based on the PANIC procedure.

4.2.3 Chang (2002, 2004)

Under few restrictions on the covariance matrix of residuals, the limit distribution of the standard (Wald-type) unit root test depends on nuisance parameters defining cross-section dependence. To overcome this issue, Chang develops a non-linear instrumental variables estimator of the autoregressive parameters in a standard ADF model (7). The instruments are generated according to a non-linear function of the lagged values: $F(y_{i,t-1})$. The Instruments Generating Function (IGF) must meet the features of a regularly integrable function. In particular, it must satisfy $\int xF(x)dx \neq 0,$ ensuring a correlation between the lagged values of the series and the non-linear instruments. Chang demonstrates that the *t*-ratio, denoted Z_i, associated with the non-linear IV estimation of the individual ADF, is asymptotically normally distributed. In addition, the individual Z_i statistics are shown to be independent across cross-section. This leads Chang to construct a pooled IV *t*-statistics, denoted S_N, which converges to a standard normal distribution under the null hypothesis.

Table 4 displays the nonlinear IV *t*-ratio statistics for four different instruments generating functions. The results are mixed depending on the model specification. In the model without constant, the unit root hypothesis is rejected for the three variables for most IGFs. The same findings hold for the individual constant specification. However, these conclusions are totally reversed when the model specification includes a time-trend variable. SO_2 and trade openness reject the unit root hypothesis independently of the IGF considered.

Table 2: Bai & Ng (2004) Test

Model		FDI			SO_2			Openness		
Constant	Optimal factors	4	1		4	1		4	1	
	$P_{\hat{e}}$ (idiosyncratic shocks)	16.271	3.997		-5.048	-0.038		3.532	3.672	
	$t_{\hat{F}}$ (common factors)	-	-3.83***		-	-2.27		-	-1.683	
	MQ_c (common factors)	-11.317*	-		-9.872	-		-12.239*	-	
	MQ_f, (common factors)	-0.423	-		-9.327	-		-4.224	-	
Constant and Trend	Optimal factors	4	2	1	4	2	1	4	3	1
	$P_{\hat{e}}$ (idiosyncratic shocks)	39.288	19.399	9.599	-0.758	4.761	14.464	2.124	13.279	7.315
	$t_{\hat{F}}$ (common factors)	-	-	-1.775	-	-	-0.716	-	-	-0.984
	MQ_c (common factors)	-11.858*	-5.752	-	-11.19*	-0.13	-	-11.812*	-2.3	-
	MQ_f, (common factors)	-4.17	-6.022	-	-9.884	-2.548	-	-5.441	-5.437	-

Notes: *P* denotes the inverse chi-square (Fisher) test statistic based on the defactored idiosyncratic component of the variable. *t* corresponds to the standard ADF *t*-statistic when there is a single optimal factor. When the number of optimal factor is higher, two other tests, MQ_c and MQ_f, are performed. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 3: Moon & Perron (2004) Test

Model		FDI			SO_2			Openness		
Constant	Optimal factors	4	3	1	4	1	4	2	1	3
	t_a^{*BCNW}	-0.14	-0.133	-0.001	-7E+3***	-497***	-4E+5***	-9E+3***	-7E+3***	-0.133
	t_b^{*BCNW}	-4.73***	-4.94***	-23.7***	-1.28	-7.48***	-5.70***	-14***	-20***	-4.94***
	t_a^{*BINW}	-0.133	-0.131	-0.001	-8E+3***	-469***	-5E+5***	-9E+3***	-8E+2***	-0.131
	t_b^{*BINW}	-4.60***	-4.78***	-22.3***	-1.17	-7.26***	-5.6***	-14.6***	-19***	-4.78***
	t_a^{*QSA}	-0.134	-0.124	0	-7E+3***	-470***	-5E+5***	-9E+3***	-8E+0***	-0.124
	t_b^{*QSA}	-4.85***	-5.02***	-24.4***	-1.33*	-7***	-6***	-15***	-21***	-5.02***
Constant and Trend	Optimal factors	4	3	1	4	1	4	3	1	3
	t_a^{*BCNW}	0	0	0	0.013	11.43	58.98	38.63	3.63	0
	t_b^{*BCNW}	0.084	-0.301	-0.125	0.008	1.14	0.84	0.587	0.068	-0.301
	t_a^{*BINW}	0	0	0	0.013	12.18	64.62	41.14	3.83	0
	t_b^{*BINW}	0.085	-0.31	-0.139	0.008	1.14	0.85	0.597	0.06	-0.31
	t_a^{*QSA}	0	0	0	0.003	5.08	21.12	14.72	1.80	0
	t_b^{*QSA}	0.064	-0.256	-0.107	0.006	0.90	0.608	0.43	0.055	-0.256

Notes: *t* denotes the *t*-statistic based on the defactored data. *BCNW* stands for the long-run variance computed according to Bartlett kernel function with common Newey and West truncation lag. *BINW* denotes the long-run variance computed according to Bartlett kernel function with individual Newey and West truncation lag. *QSA* stands for the long-run variance computed according to a Quadratic Spectral function and individual lag. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 4: Chang (2002) Test

		FDI	SO_2	Openness
No Constant	$S_{N(F(x)=1\{\|x\|\leq K\})}$	6.831	0.654	9.328
	$S_{N(F(x)=sign(x)\cdot 1\{\|x\|\leq K\})}$	0.593	0.721	0.301
	$S_{N(F(x)=x\cdot 1\{\|x\|\leq K\})}$	0.588	0.717	0.299
	$S_{N(F(x)=x\cdot \exp(-\|x\|)1\{\|x\|\leq K\})}$	6.327	-10.51***	6.763
Constant	$S_{N(F(x)=1\{\|x\|\leq K\})}$	3.148	2.519	5.083
	$S_{N(F(x)=sign(x)\cdot 1\{\|x\|\leq K\})}$	0.702	1.145	1.827
	$S_{N(F(x)=x\cdot 1\{\|x\|\leq K\})}$	1.165	0.874	2.108
	$S_{N(F(x)=x\cdot \exp(-\|x\|)1\{\|x\|\leq K\})}$	22.102	-9.219***	-2.201**
Constant and Trend	$S_{N(F(x)=1\{\|x\|\leq K\})}$	-0.841	-4.282***	-7.245***
	$S_{N(F(x)=sign(x)\cdot 1\{\|x\|\leq K\})}$	0.105	-2.668***	-3.372***
	$S_{N(F(x)=x\cdot 1\{\|x\|\leq K\})}$	-0.785	-2.179**	-3.642***
	$S_{N(F(x)=x\cdot \exp(-\|x\|)1\{\|x\|\leq K\})}$	1.215	-12.09***	-13.163***

Notes: S_N denotes the average individual nonlinear IV *t*-ratio statistics based on a specific IGF. *** significant at 1%, ** significant at 5%, * significant at 10%.

Overall, the different tests are not very conclusive about the stationarity properties of these variables. This comes partially from the fact that each test relies on specific assumptions and a different underlying specification. What seems to be clear is the presence of at least one common factor in the data. Since the GEL estimator based on the system of moments conditions does not break down in the presence of a unit root, the coefficient of the lagged dependent variable can ultimately be used to test for the homogenous non-stationarity of each variable.

4.3 Estimation Results

In order to investigate the pollution haven effect, I estimate two different panel VAR models: a non-spatial and spatial trivariate model with FDI, SO_2 and trade openness. The main reason to proceed sequentially is to highlight potential collinearity problems after the introduction of lagged spatial covariates. The order of the VAR is set to one in order to avoid dealing with a too small time period (T = 8) and severe multicollinearity. All specifications include individual time dummies to account for global time related shocks and reduce potential bias due to the presence of cross-section dependence in the error term. However, following Love and Zicchino (2006), these individual time effects are removed before estimation by demeaning the data, in order to mitigate potential additional collinearity.

All estimations are performed with the panel fixed effect GEL estimator, unless specified otherwise. The different equations in the system are estimated at once based on the moment conditions suggested by the system GMM estimator (see equations (2)-(5)). More specifically, the instruments for the endogenous variables rely on the second and third lags of the dependent variables, while the lagged spatial lags are treated as exogenous. To reduce the number of moment conditions, the instruments are also collapsed (see Appendix 6.2, Roodman, 2009). Note that GEL and more generally spatial VAR are not available in standard econometric software packages, that is why I implement the different estimators in Matlab[22]. Table 5 reports the results associated with the aspatial and spatial VAR for the full sample, while Table 6 displays the results for low and high income subsamples but based on the whole sample spatial weight matrix[23]. Note that dealing with subsamples is problematic, because small sample size increases considerably the presence of collinearity and convergence issues. This is particularly the case of the high income sub-sample, that is why I turn to the continuously update GMM estimator, which, despite assigning the same probability weight to each cross-section, remains more efficient than System GMM and is less sensitive to convergence issues.

The results confirm the dynamic nature of FDI inflows, pollution emissions and openness to trade with highly positive significant autoregressive parameters. The estimates suggest the presence of a simultaneous pollution haven and pollution halo effects. More pollution has a strong enhancing effect on inflows of FDI, while FDI seems to lead to lower pollution emissions. Taking together, both results are in line with the existence of potential endogeneity of the pollution haven/halo effect. However, these opposing effects are only significant for the low income sample. For high income countries, it seems that inflows of FDI do not have any significant impact on SO2 emissions. This can be explained by the higher prevalence of cleaner technologies in high income countries, which gives further insight on the smaller order of

[22] The optimization algorithms are based on John Zedlewski's *matElike* and Matlab's *fmincon* routines.
[23] Appendix 6.3 reports the high and low income classification.

magnitude of the pollution haven effect in developed economies. In fact, it seems that FDI allocation to developing countries is slightly more sensitive to environmental stringency in low income countries. These findings remain valid when outlier countries in terms of SO2/GDP are dropped (e.g. Chile and Peru). The relationship between FDI inflows and trade-to-openness is also characterized by opposite effects. On the one hand, countries with large trade openness have lower FDI inflows, on the other hand, an increase in inflows of FDI increases the level of trade-openness. Yet, the sub-sample results suggest a single positive link between FDI and openness in developed economies, but a negative relationship between trade openness and FDI inflows in developing countries.

Table 5: Full Sample Spatial VAR(1) Results

	GEL Results			GEL Results		
	FDI	SO_2	Openness	FDI	SO_2	Openness
FDI (*t*-1)	0.735	-0.026	0.378	0.637	-0.017	0.214
	[0.042]***	[0.007]***	[0.079]***	[0.055]***	[0.008]**	[0.079]***
SO_2 (*t*-1)	0.355	0.739	0.546	0.205	0.670	0.016
	[0.094]***	[0.042]***	[0.193]***	[0.039]***	[0.022]***	[0.109]
Openness (*t*-1)	-0.065	-0.002	0.560	-0.159	-0.009	0.454
	[0.022]***	[0.006]	[0.050]***	[0.028]***	[0.007]	[0.055]***
Spatial FDI (*t*-1)				-0.051	-0.021	-0.079
				[0.058]	[0.014]	[0.109]
Spatial SO_2 (*t*-1)				-0.199	0.034	-0.194
				[0.047]***	[0.019]*	[0.134]
Spatial Openness (*t*-1)				0.108	0.007	0.357
				[0.020]***	[0.005]	[0.047]***
Countries	124	124	124	124	124	124
Time	8	8	8	8	8	8

Notes: The model includes time and country fixed effects, but time effects are removed prior to estimation. The estimation is performed using GEL with 2nd and 3rd lags as collapsed instruments. Standard errors are reported in brackets. *** significant at 1%,** significant at 5%, * significant at 10%.

Taking into account spatial interactions leads to additional interesting findings. In particular, the negative impact of the lagged spatial lag of SO2 on the allocation of FDI confirms the evidence of the pollution haven effect. FDI flows in a given country tend to decrease if SO2 is higher in the neighbouring countries. This effect is quantitatively stronger for high income countries. More generally, most spatial covariates tend to be of a higher order of magnitude for developed countries, while temporal covariates are larger for developing countries. This confirms the view that high income countries are economically more integrated than developing economies. In fact, FDI inflows, pollution emission and trade openness in high income countries are characterized by a strong complementary strategic relationship in terms of spatial covariates.

Table 6: Sub-sample Spatial VAR(1) Results

	Low Income GEL Results			High Income CUE Results		
	FDI	SO_2	Openness	FDI	SO_2	Openness
FDI (*t*-1)	0.580	-0.045	0.010	0.217	-0.007	0.168
	[0.053]***	[0.014]***	[0.113]	[0.029]***	[0.008]	[0.044]***
SO_2 (*t*-1)	0.212	0.683	0.202	0.186	0.179	0.183
	[0.052]***	[0.028]***	[0.145]	[0.068]***	[0.063]***	[0.233]
Openness (*t*-1)	-0.149	-0.026	0.417	1.78e-04	-0.012	0.103
	[0.019]***	[0.008]***	[0.060]***	[0.024]	[0.009]	[0.057]*
Spatial FDI (*t*-1)	-0.075	-0.002	-0.120	0.773	-0.019	0.133
	[0.049]	[0.018]	[0.118]	[0.123]***	[0.009]**	[0.102]
Spatial SO_2 (*t*-1)	-0.207	0.002	-0.304	-0.218	0.811	-0.129
	[0.036]***	[0.011]	[0.095]***	[0.062]***	[0.167]***	[0.711]
Spatial Openness (*t*-1)	0.093	0.007	0.354	0.019	0.001	0.887
	[0.020]***	[0.005]	[0.049]***	[0.029]	[0.014]	[0.077]***
Countries	83	83	83	41	41	41
Time	8	8	8	8	8	8

Notes: The model includes time and country fixed effects, but time effects are removed prior to estimation. The estimation is performed using GEL with 2nd and 3rd lags as collapsed instruments. Standard errors are reported in brackets. *** significant at 1%,** significant at 5%, * significant at 10%.

Putting together, these findings confirm the importance of potential endogeneity as well as the presence of dynamic interdependence in the analysis of the pollution haven effect. Failing to distinguish between high and low income countries might also lead to misleading results.

Before investigating the spatial causality of these results, I re-estimate the SpVAR with the CO_2/GDP ratio and the number of ratified international environmental treaties as alternative indirect measures of environmental policy. As highlighted by Table 7, the estimates associated with CO_2 emissions support most of the earlier findings. Inflows of FDI to developing countries is larger when pollution intensity is higher, while now lagged FDI leads to higher CO_2/GDP ratio in high income countries. However, according to Table 8, this is no longer the case with the number of signed international environmental treaties. The absence of significant relationship between inflows of FDI and environmental treaties could be explained by the fact that the ratification of an international environmental treaty does not necessarily translate into stringer environmental regulation. In some cases, countries with an incentive to ratify the treaty already comply with the environmental standards implied by the agreement. In other cases, sanctions against non-complying members are almost inexistent. In any case, the results with treaties should be treated very carefully, because the number of treaties is a count process and should, therefore, be estimated based on a count model specification.

Table 7: CO_2 Spatial VAR(1) Robustness Check Results

	Low Income GEL Results			High Income CUE Results		
	FDI	CO_2	Openness	FDI	CO_2	Openness
FDI (*t*-1)	0.532	0.037	0.019	0.460	0.027	0.424
	[0.048]***	[0.029]	[0.121]	[0.032]***	[0.014]*	[0.031]***
CO_2 (*t*-1)	0.083	0.831	-0.152	0.030	0.533	0.266
	[0.026]***	[0.052]***	[0.069]**	[0.052]	[0.045]***	[0.072]***
Openness (*t*-1)	-0.112	0.083	0.604	-0.005	-0.031	0.544
	[0.024]***	[0.027]***	[0.065]***	[0.012]	[0.007]***	[0.019]***
Spatial FDI (*t*-1)	0.087	0.098	0.218	0.530	0.008	-0.155
	[0.059]	[0.053]*	[0.126]*	[0.105]***	[0.048]	[0.052]***
Spatial CO_2 (*t*-1)	-0.156	0.150	0.022	-0.069	0.457	-0.074
	[0.033]***	[0.050]***	[0.075]	[0.022]***	[0.108]***	[0.090]
Spatial Openness (*t*-1)	0.048	-0.038	0.204	0.017	0.040	0.446
	[0.014]***	[0.017]**	[0.040]***	[0.011]	[0.009]***	[0.044]***
Countries	91	91	91	42	42	42
Time	8	8	8	8	8	8

Notes: The model includes time and country fixed effects, but time effects are removed prior to estimation. The estimation is performed using GEL with 2nd and 3rd lags as collapsed instruments. Standard errors are reported in brackets. *** significant at 1%,** significant at 5%, * significant at 10%.

Table 8: Treaties Spatial VAR(1) Robustness Check Results

	Low Income GEL Results			High Income CUE Results		
	FDI	Treaties	Openness	FDI	Treaties	Openness
FDI (*t*-1)	0.559	0.018	-0.016	0.508	0.005	0.246
	[0.051]***	[0.016]	[0.109]	[0.039]***	[0.023]	[0.038]***
Treaties (*t*-1)	0.004	0.933	0.091	-0.056	0.501	-0.020
	[0.034]	[0.021]***	[0.099]	[0.041]	[0.094]***	[0.085]
Openness (*t*-1)	0.010	-0.007	0.736	0.007	-0.003	0.533
	[0.017]	[0.008]	[0.060]***	[0.029]	[0.022]	[0.036]***
Spatial FDI (*t*-1)	0.151	0.006	0.545	0.482	-0.179	-0.086
	[0.067]**	[0.021]	[0.159]***	[0.125]***	[0.120]	[0.127]
Spatial Treaties (*t*-1)	0.034	0.059	0.035	0.110	0.489	0.056
	[0.021]	[0.014]***	[0.066]	[0.041]***	[0.102]***	[0.132]
Spatial Openness (*t*-1)	-0.004	0.007	0.065	0.024	0.014	0.457
	[0.008]	[0.004]*	[0.027]**	[0.041]	[0.042]	[0.072]***
Countries	95	95	95	40	40	40
Time	8	8	8	8	8	8

Notes: The model includes time and country fixed effects, but time effects are removed prior to estimation. The estimation is performed using GEL with 2nd and 3rd lags as collapsed instruments. Standard errors are reported in brackets. *** significant at 1%,** significant at 5%, * significant at 10%.

4.4 Spatial Granger Causality Test

In order to check the relevance of the different variables in the system, it is useful to investigate whether each variable is determined, in part, by the remaining spatial and non-spatial variables. A variable *x* is said to Granger cause *y*, if the information embodied in previous values of *x* is useful for forecasting *y*. Yet, Granger causality does not imply in causality *per se*. One shortcoming of the standard panel VAR is its inability to distinguish between time and spatial causality. Spatial causality refers to the predictability associated with the previous spatial distribution of variable *x* to improve understanding the current spatial distribution of variable *x* or *y*. The (spatial) panel Granger test is implemented through a likelihood ratio test, where the null hypothesis assumes no (spatial) causality (Mur and Paelinck, 2009). The (spatial) test compares the log likelihood of the full model with the one associated with the system of equations subject to a zero restriction on a given parameter. Under the null hypothesis, the statistics follows a chi-square distribution of *k* degree(s) of freedom, *k* being the number of parameter constraints: $2\left(L_{H_a} - L_{H_0}\right) \overset{as}{\sim} \chi^2(k)$.

Table 9: Full Sample Panel Granger Causality Tests

$\longrightarrow$	Full Sample			Low Income Sample		
Granger causes	FDI	SO_2	Openness	FDI	SO_2	Openness
FDI	-	2.35	4.52**	-	7.41***	0.15
Spatial FDI	-2.56	-1.99	-2.56	0.10	3.33*	0.09
FDI + Spatial FDI	-	13.85***	4.61*	-	8.07**	0.15
SO_2	16.56***	-	0.03	39.80***	-	0.43
Spatial SO_2	4.26**	-2.80	4.26	29.71***	6.92***	2.97
SO_2 + Spatial SO_2	16.52***	-	3.53	41.58***		0.67
Openness	11.71***	15.78***	-	16.37***	0.35	-
Spatial Openness	11.45***	22.24***	11.45**	13.17***	4.00**	13.17**
Openness + Spatial Openness	11.74***	22.27***	-	24.16***	5.99**	-

Notes: Likelihood ratio test with no Granger causality as the null hypothesis. *** significant at 1%, ** significant at 5%, * significant at 10%.

4.5 Spatial Impulse Response Function

Unlike in traditional VAR, the spatial IRF can decompose the response of a shock (ς) into a (direct) time (Φ) and (indirect) spatial (Π) effect. Instead of analyzing how a specific shock in a given country propagates itself to the remaining countries, I consider a common global shock affecting initially the entire set of countries (e.g. global economic downturn). As shown in Appendix 6.1, the cumulative average effect of the shock ς at time *t* + *s* is given by $[\Phi + \Pi]^s I_m \varsigma$.

Figure 2: FDI's Impulse-Response Functions

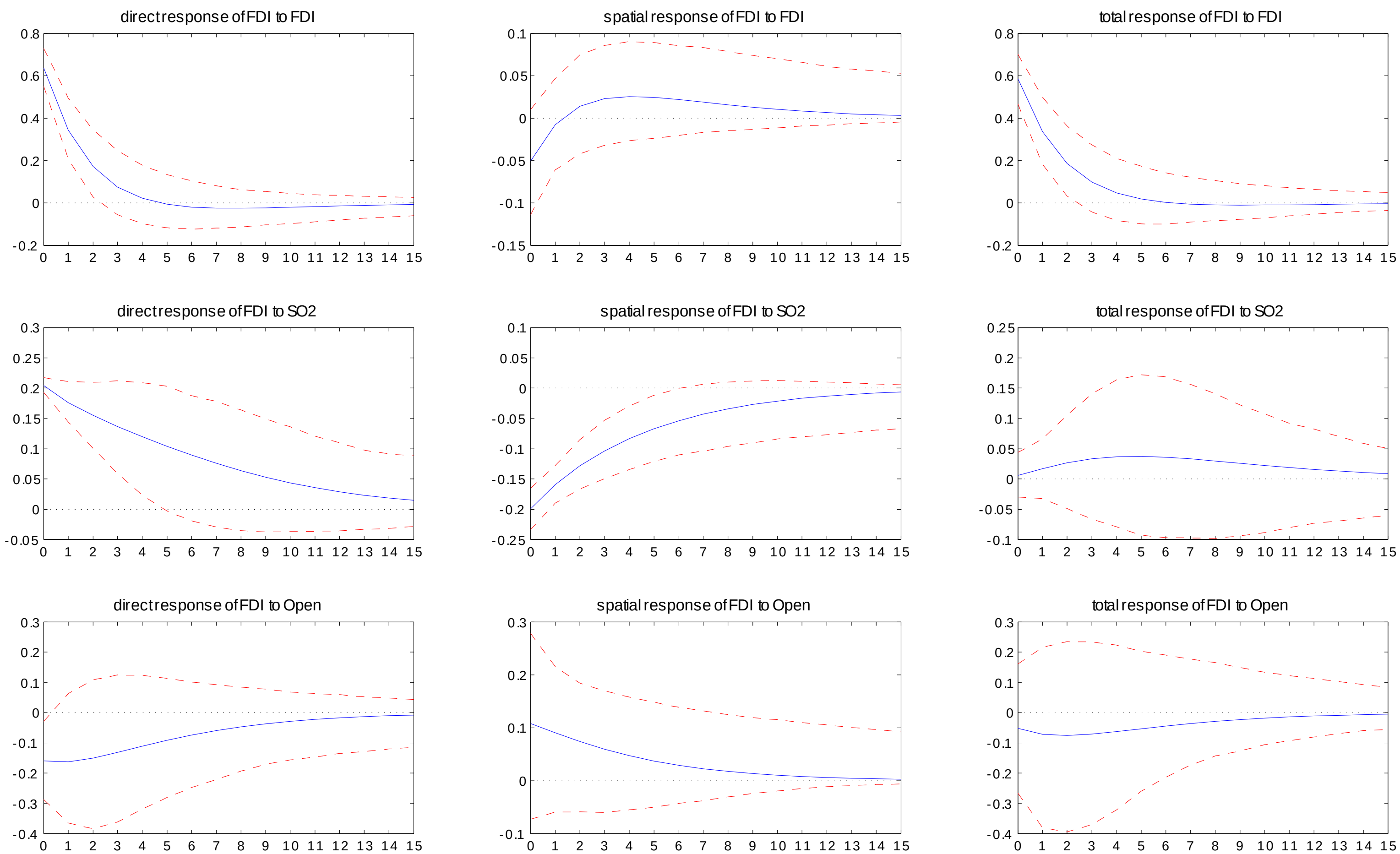


Figure 3: SO_2's Impulse-Response Functions

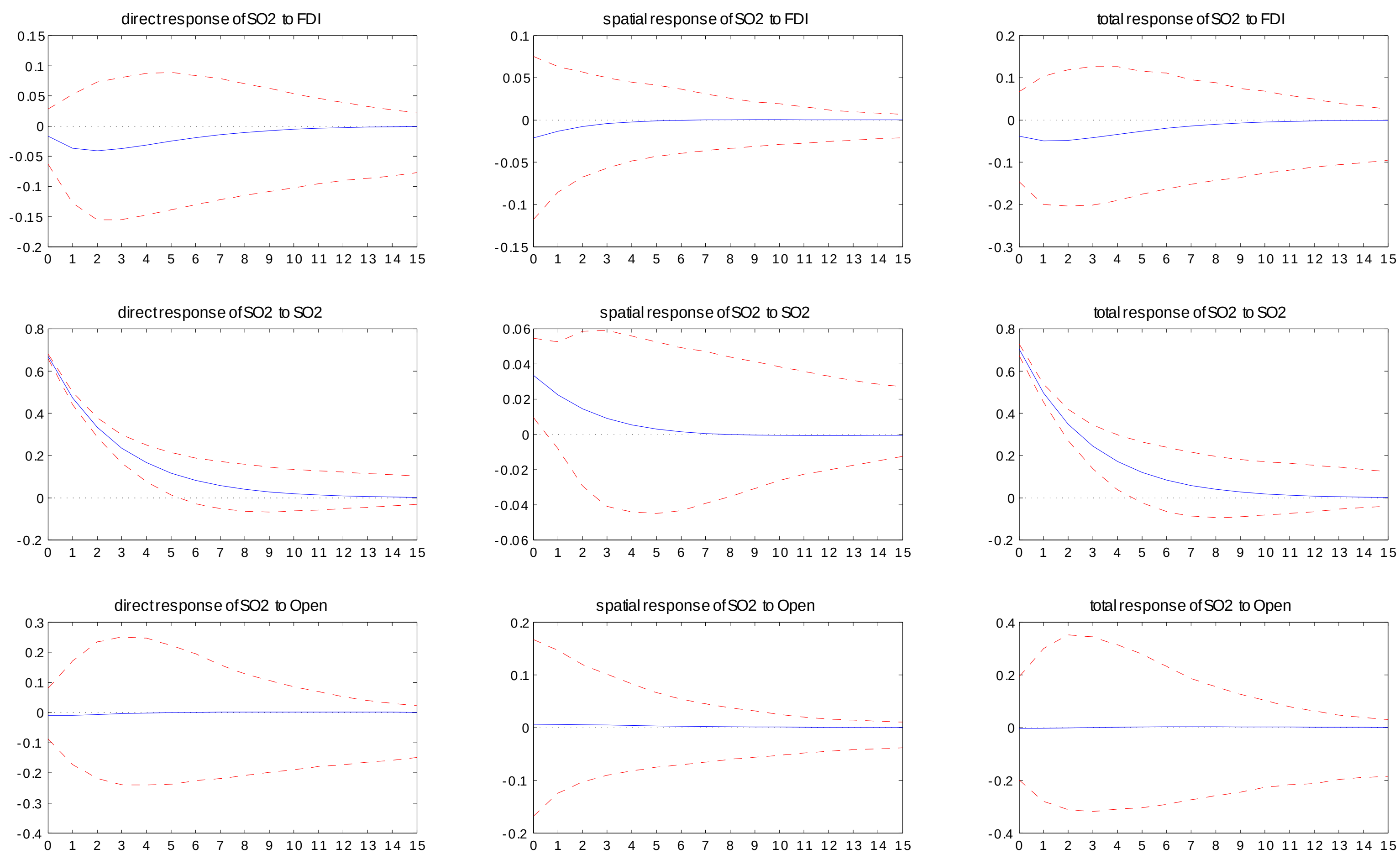


Figure 4: Openness's Impulse-Response Functions

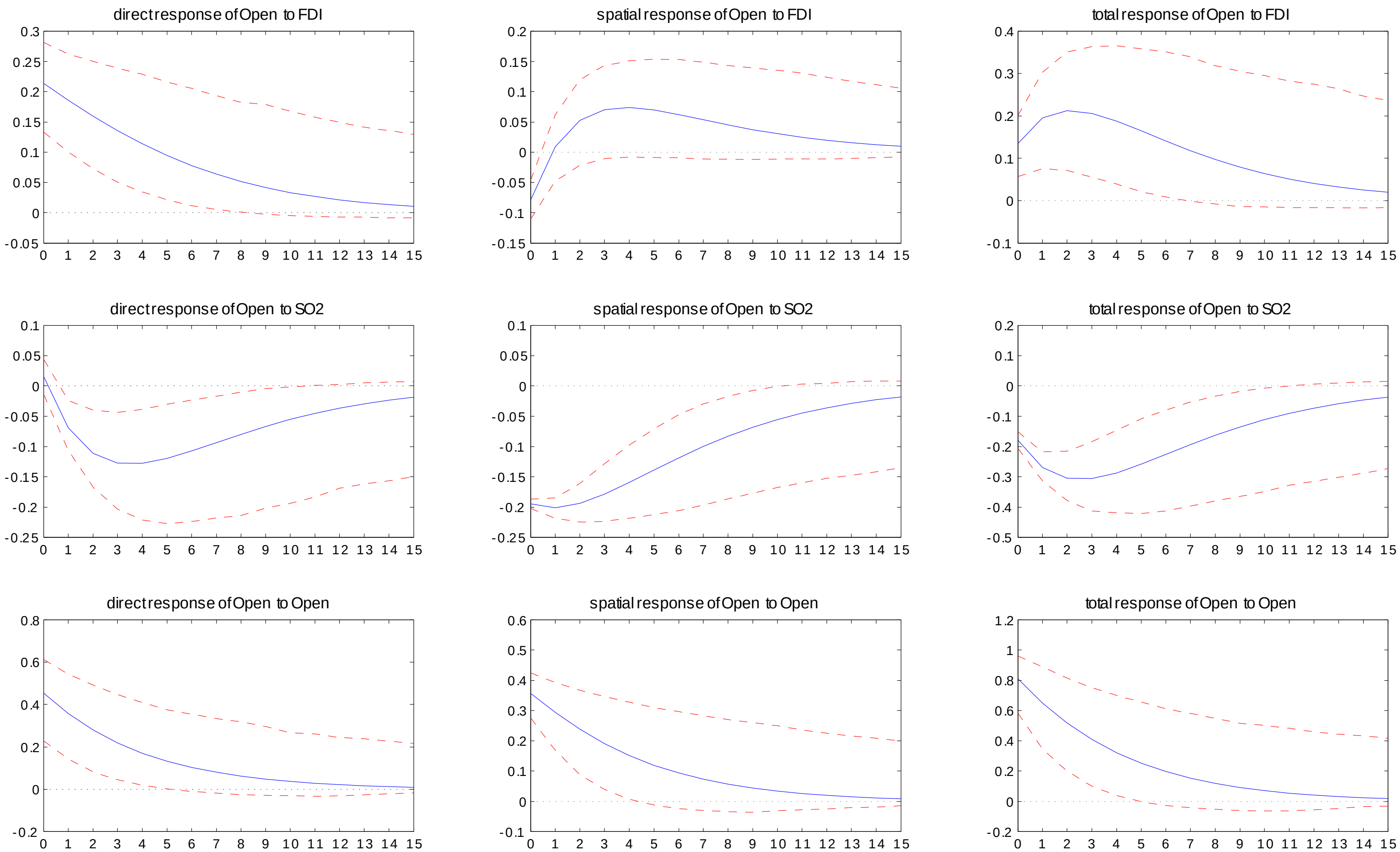


Figure 2 displays the impulse-response functions associated with inflows of FDI and the 5% error bands. Following Love and Zicchino (2006), the construction of the confidence intervals is done by randomly generating draws based on the estimated VAR coefficients and standard errors and re-computing the IRFs. This procedure is performed 1000 times in order to compute the 5th and 95th percentile of the distribution. For each row, the first plot reports the IRF of the time effect of the shock, while the second one presents the spatial effect. The third plot considers the total effect as the sum of the direct and indirect effects. Figure 3 and 4 report the similar plots for a unitary shock in SO_2 emission and trade openness, respectively.

The previous findings are corroborated by the impulse response functions (IRFs). A shock in SO_2 emissions on the level of FDI inflows is globally positive but can be decomposed into a positive and negative impact through time and spatial dynamic, respectively. The opposite holds for a shock in FDI on the level of pollution, where the spatial dependence amplifies the impact of the shock. Both results support the potential existence of a bidirectional relation between FDI and environmental regulation and highlight the potential endogeneity of the pollution haven/halo effect. For the remaining IRFs, most of the spatial spillovers exacerbate the initial shock through a complementary relationship. One of the exceptions is the response of FDI after a shock in trade openness, where spatial dependence mitigates the impact.

5 Conclusion

This paper investigates the relationship between FDI, environmental stringency (proxied by SO_2/GDP emissions) and trade openness of 124 countries during the nineties by taking explicitly into account spatial dependence. Prior to the estimation, I perform second generation panel unit root tests. Although the results of the tests are not clear-cut, they provide confirmation of the existence of at least a single common factor. Based on a spatial panel VAR framework, I find evidence of a pollution haven effect in particular in developing countries. Furthermore, the panel Granger causality tests suggest that there is a potential bidirectional causality between FDI and pollution emissions once spatial dependence is taking into account, confirming in this sense the pollution halo hypothesis. The potential endogeneity of the relationships is further illustrated by the spatial impulse response functions. In addition, the SpVAR is able to highlight the role of spatial interaction in the determination of the linkages between FDI, openness and environmental regulation through amplification or mitigation mechanisms.

The results of this paper complement recent theoretical and empirical works, which highlight the importance of accounting for the likely endogeneity of environmental stringency with respect to FDI. Another important methodological feature of this study is the importance of accounting for spatial dynamic interactions. Any empirical study that omits both econometric issues might face biased and unreliable results. In addition, the analysis should be performed by distinguishing between developed and developing countries in order to mitigate the homogenous bias. Ultimately, the assessment of the pollution haven effect should be done through a structural specification which is able to capture all the determinants of the relationship between environmental policy and economic integration through trade and investment.

From a policy perspective, the results definitively rule in favour of coordination between trade/investment and environmental policies at the national level through multilateral cooperation. The presence of positive and negative spatial spillovers, which amplify or smooth shocks, could then be fully captured or mitigated. More generally, any international agreement with environmental provisions addressing the risk of pollution haven effect should include different dispositions between high and low income countries as their relationship with respect to economic activity and pollution is significantly different from each other. In particular, developing economies cannot be expected to make changes right away. In fact, any agreement should encourage the transition to stringer environmental policies without impeding productivity and growth.

6 Appendices

6.1 Spatio-Temporal Stationarity in SpVAR

To illustrate the spatio-temporal stationarity condition, I consider a SpVAR with N = 2 assuming W is row-standardized such that

$$W = \begin{bmatrix} 0 & d_{12} \\ d_{21} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Then it follows that a common shock ς associated with variable *x*, which occurs at time *t*-1, yields:

$$\begin{bmatrix} x_{1t} \\ z_{1t} \\ x_{2t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & 0 & 0 \\ \phi_{21} & \phi_{22} & 0 & 0 \\ 0 & 0 & \phi_{11} & \phi_{12} \\ 0 & 0 & \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \varsigma \\ 0 \\ \varsigma \\ 0 \end{bmatrix} + \begin{bmatrix} \rho_{11} & \rho_{12} & 0 & 0 \\ \rho_{21} & \rho_{22} & 0 & 0 \\ 0 & 0 & \rho_{11} & \rho_{12} \\ 0 & 0 & \rho_{21} & \rho_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 & d_{12} & 0 \\ 0 & 0 & 0 & d_{12} \\ d_{21} & 0 & 0 & 0 \\ 0 & d_{21} & 0 & 0 \end{bmatrix} \begin{bmatrix} \varsigma \\ 0 \\ \varsigma \\ 0 \end{bmatrix}$$
$$= \begin{bmatrix} \phi_{11} + \rho_{11} d_{12} \\ \phi_{21} + \rho_{21} d_{12} \\ \phi_{11} + \rho_{11} d_{21} \\ \phi_{21} + \rho_{21} d_{21} \end{bmatrix} \varsigma = \begin{bmatrix} \phi_{11} + \rho_{11} \\ \phi_{21} + \rho_{21} \\ \phi_{11} + \rho_{11} \\ \phi_{21} + \rho_{21} \end{bmatrix} \varsigma$$

Thus, the average effect of ς on the variables *x* and *y* at time *t* is given by $[(\phi_{11} + \rho_{11}),(\phi_{21} + \rho_{21})]' \varsigma$, and more generally at time $t + s$ by:

$$\begin{bmatrix} \phi_{11} + \rho_{11} & \phi_{12} + \rho_{12} \\ \phi_{21} + \rho_{21} & \phi_{22} + \rho_{22} \end{bmatrix}^s \begin{bmatrix} \varsigma \\ 0 \end{bmatrix}.$$

This result remains valid for N > 2, as long as the spatial matrix W is row standardized. This is also the approach used to construct the impulse response functions of the system.

6.2 Extended GMM in SpVAR

Let $\varepsilon_i = Y_i - X_{i,-1}\Psi' = [\varepsilon_{i2}, \ldots, \varepsilon_{iT}]'$, where ε_{it} is a $k \times 1$ vector. To simplify the notation, each element of the $\tilde{W}Y_t$ vector corresponds to $\dot{Y}_{i,t}$. The moments conditions given in (2)-(5) can be stacked as $E[P_i' \varepsilon_i] = 0$, where $P_i' = [P_{L_i}', P_{D_i}']'$.

$$P_{L_i}=\begin{bmatrix} -y_{i1} & y_{i1} & 0 & \cdots & 0 & 0 \\ 0 & -y_{i2} & y_{i2} & \cdots & 0 & 0 \\ 0 & -y_{i1} & y_{i1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & -y_{i,T-2} & y_{i,T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -y_{i1} & y_{i1} \\ -\dot{y}_{i2} & \dot{y}_{i2} & 0 & \cdots & 0 & 0 \\ -\dot{y}_{i1} & \dot{y}_{i1} & 0 & \cdots & 0 & 0 \\ 0 & -\dot{y}_{i3} & \dot{y}_{i3} & \cdots & 0 & 0 \\ 0 & -\dot{y}_{i2} & \dot{y}_{i2} & \cdots & 0 & 0 \\ 0 & -\dot{y}_{i1} & \dot{y}_{i1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & -\dot{y}_{i,T-2} & \dot{y}_{i,T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\dot{y}_{i1} & \dot{y}_{i1} \end{bmatrix} \quad P_{D_i}=\begin{bmatrix} 0 & y_{i2} & 0 & \cdots & 0 \\ 0 & 0 & y_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i,T-1} \\ \Delta\dot{y}_{i2} & 0 & 0 & \cdots & 0 \\ 0 & \Delta\dot{y}_{i3} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta\dot{y}_{iT} \end{bmatrix}$$

In order to restrict the number of instruments and avoid over-fitting, the moments conditions are collapsed as follows (Roodman, 2009):

$$P_{L_i}=\begin{bmatrix} -y_{i1} & (y_{i1}-y_{i2}) & (y_{i2}-y_{i3}) & \cdots & (y_{i,T-3}-y_{i,T-2}) & y_{i,T-2} \\ 0 & -y_{i1} & (y_{i1}-y_{i2}) & \cdots & (y_{i,T-4}-y_{i,T-3}) & y_{i,T-3} \\ 0 & 0 & -y_{i1} & \cdots & (y_{i,T-5}-y_{i,T-4}) & y_{i,T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -y_{i1} & y_{i1} \\ -\dot{y}_{i2} & (\dot{y}_{i2}-\dot{y}_{i2}) & (\dot{y}_{i3}-\dot{y}_{i4}) & \cdots & (\dot{y}_{i,T-2}-\dot{y}_{i,T-1}) & \dot{y}_{i,T-1} \\ -\dot{y}_{i1} & (\dot{y}_{i1}-\dot{y}_{i2}) & (\dot{y}_{i2}-\dot{y}_{i3}) & \cdots & (\dot{y}_{i,T-3}-\dot{y}_{i,T-2}) & \dot{y}_{i,T-2} \\ 0 & -\dot{y}_{i1} & (\dot{y}_{i1}-\dot{y}_{i2}) & \cdots & (\dot{y}_{i,T-4}-\dot{y}_{i,T-3}) & \dot{y}_{i,T-3} \\ 0 & 0 & -\dot{y}_{i1} & \cdots & (\dot{y}_{i,T-5}-\dot{y}_{i,T-4}) & \dot{y}_{i,T-4} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\dot{y}_{i1} & \dot{y}_{i1} \end{bmatrix},$$

$$P_{D_i}=\begin{bmatrix} 0 & y_{i2} & y_{i3} & \cdots & y_{i,T-1} \\ \Delta\dot{y}_{i,2} & \Delta\,\dot{y}_{i3} & \cdots & \cdots & \Delta\dot{y}_{iT} \end{bmatrix}$$

6.3 Country Lists

Low Income Countries

Albania; Angola; Argentina; Armenia; Bangladesh; Belarus; Belize; Benin; Bolivia; Botswana; Brazil; Bulgaria; Burundi; Cambodia; Cape Verde; Chile; China; Colombia; Congo, Rep.; Costa Rica; Croatia; Côte d'Ivoire; Dominica; Dominican Republic; Ecuador; Egypt, Arab Rep.; El Salvador; Ethiopia; Fiji; Gabon; Ghana; Grenada; Guatemala; Guinea; Guyana; Honduras; India; Indonesia; Iran, Islamic Rep.; Jamaica; Jordan; Kenya; Kyrgyz Republic; Latvia; Lebanon; Lesotho; Lithuania; Madagascar; Malawi; Malaysia; Maldives; Mali; Mauritius; Mexico; Mongolia; Morocco; Mozambique; Myanmar; Namibia; Nicaragua; Nigeria; Pakistan; Panama; Papua New Guinea; Paraguay; Peru; Philippines; Poland; Romania; Rwanda; Senegal; Seychelles; Solomon Islands; South Africa; Sri Lanka; St. Kitts and Nevis; St. Lucia; Sudan; Suriname; Swaziland; Syrian Arab Republic; Tajikistan; Tanzania; Thailand; Togo; Tonga; Tunisia; Turkey; Turkmenistan; Uganda; Uruguay; Venezuela, RB; Yemen, Rep.

High Income Countries

Antigua and Barbuda; Aruba; Australia; Austria; Bahamas, The; Bahrain; Barbados; Belgium-Luxembourg; Canada; Cyprus; Czech Republic; Denmark; Estonia; Finland; France; Germany; Greece; Hungary; Iceland; Ireland; Israel; Italy; Japan; Korea, Rep.; Kuwait; Malta; Netherlands; Netherlands Antilles; New Zealand; Norway; Oman; Portugal; Saudi Arabia; Singapore; Slovak Republic; Slovenia; Spain; Sweden; Switzerland; Trinidad and Tobago; United Kingdom; United States.

6.4 Descriptive Statistics

	Mean	Std-Error	Min	Max
FDI	3.02	4.34	-19.11	43.53
SO_2	0.49	0.73	0.01	7.62
Openness	42.28	23.181	0	167
Spatial FDI	3.15	2.35	-3.70	16.20
Spatial SO_2	0.48	0.26	0.09	1.97
Spatial Openness	43.32	14.79	14.05	133.23

6.5 Correlation Matrix

	FDI	SO_2	Openness	Spatial FDI	Spatial SO_2	Spatial Openness
FDI	1					
SO_2	-0.04	1				
Openness	0.39***	0.07**	1			
Spatial FDI	0.23***	-0.1***	0.14**	1		
Spatial SO_2	-0.14***	0.18**	-0.06*	-0.29***	1	
Spatial Openness	0.08***	-0.08***	0.28***	0.48***	-0.06*	1

Notes: *** significant at 1%, ** significant at 5%, * significant at 10%.

6.6 First Generation Panel Unit Root Tests

Levin, Lee & Chu (2002) Test

Model		FDI	SO_2	Openness
No Constant	t^{*BCNW}	-1.352*	-6.72***	5.161
	t^{*BINW}	-1.352*	-6.72***	5.161
	t^{*QSIA}	-1.35*	-6.72***	5.162
Constant	t^{*BCNW}	-1.355*	-6.723***	5.16
	t^{*BINW}	0.183	-19.763***	-16.964***
	t^{*QSIA}	-1.793***	-20.666***	-16.896***
Constant and Trend	t^{*BCNW}	1.833***	-19.33***	-16.143***
	t^{*BINW}	-34.986***	-35.586***	-10.678***
	t^{*QSIA}	-36.458***	-34.839***	-15.704***

Notes: t^* denotes the adjusted t-statistic. *BCNW* stands for the long-run variance computed according to Bartlett kernel function with common Newey and West truncation lag. *BINW* denotes a long-run variance computed according to Bartlett kernel function with individual Newey and West truncation lag. *QSA* stands for a long-run variance computed according to a Quadratic Spectral function and individual lag. *** significant at 1%, ** significant at 5%, * significant at 10%.

Maddala & Wu (1999) and Choi (2001) Tests

Model		FDI	SO_2	Openness
No Constant	$P_{FiniteN}$	169.732	385.347***	-130.284
	$Z_{FiniteN}$	4.912	-4.61***	7.8
	$L_{FiniteN}$	99.955	-96.017	155.349
	$L^*_{FiniteN}$	0.246	-0.236	0.382
	$Pm_{InfiniteN}$	-3.514***	6.167***	-5.286***
	$Z_{InfiniteN}$	4.912	-4.61***	7.8
	$L^*_{InfiniteN}$	0.245	-0.235	0.381
Constant	$P_{FiniteN}$	191.51	390.015***	295.239**
	$Z_{FiniteN}$	4.909	-4.531***	0.259
	$L_{FiniteN}$	104.164	-93.616	5.880
	$L^*_{FiniteN}$	0.256	-0.23	0.014
	$Pm_{InfiniteN}$	-2.536***	6.377***	2.121**
	$Z_{InfiniteN}$	4.909	-4.531***	0.259
	$L^*_{InfiniteN}$	0.255	-0.229	0.014
Constant and Trend	$P_{FiniteN}$	275.561	389.488***	272.808
	$Z_{FiniteN}$	2.057	-2.364***	1.451
	$L_{FiniteN}$	40.789	-52.386	29.084
	$L^*_{FiniteN}$	0.1	-0.129	0.072
	$Pm_{InfiniteN}$	1.238	6.353***	1.114
	$Z_{InfiniteN}$	2.057	-2.364***	1.451
	$L^*_{InfiniteN}$	0.1	-0.128	0.071

Notes: *P* denotes the inverse chi-square (Fisher) test statistic, while *Z* is the inverse normal (Choi) and L^* is the logit test. Following Choi, there are two cases considered: when *N* is assumed to be finite or not to avoid degenerate distribution in the limit. *** significant at 1%, ** significant at 5%, * significant at 10%.

Breitung (2000) Test

Model		FDI	SO_2	Openness
Constant	t_{UB}	-1.785**	-1.728***	0.613
Constant and Trend	t_{UB}	0.577	0.564	2.612

Notes: *t* denotes the robust *t*-statistics. *** significant at 1%, ** significant at 5%, * significant at 10%.

Im, Pesaran & Shin (2003) Test

Model		FDI	SO_2	Openness
Constant	$\tilde{Z}_t$	-0.123	-20.902***	-10.823***
	$\tilde{W}_t$	4.72	-6.976***	-1.303*
Constant and Trend	$\tilde{Z}_t$	-23.554***	-51.945***	-20.949***
	$\tilde{W}_t$	-23.554***	-15.236***	0.464

Notes: $\tilde{Z}$ and $\tilde{W}$ denote the standardized *t*-statistic under the assumption of serially correlated errors. *** significant at 1%, ** significant at 5%,* significant at 10%.

Hadri & Larsson (2005) Test

Model		FDI	SO_2	Openness
Constant	$Z^{BCNW}_{InfiniteT}$	9.585***	10.536***	11.317***
	$Z^{BINW}_{InfiniteT}$	15.132***	16.122***	17.086***
	$Z^{QSA}_{InfiniteT}$	12.756***	13.159***	12.685***
	$Z^{BCNW}_{FiniteT}$	9.575***	10.709***	11.64***
	$Z^{BINW}_{FiniteT}$	16.19***	17.37***	18.499***
	$Z^{QSA}_{FiniteT}$	13.356***	13.837***	13.272***
Constant and Trend	$Z^{BCNW}_{InfiniteT}$	21.556***	21.446***	21.321***
	$Z^{BINW}_{InfiniteT}$	69.143***	61.464***	73.123***
	$Z^{QSA}_{InfiniteT}$	13.941***	14.423***	12.605***
	$Z^{BCNW}_{FiniteT}$	22.53***	22.385***	22.221***
	$Z^{BINW}_{FiniteT}$	85.173***	75.064***	90.412***
	$Z^{QSA}_{FiniteT}$	12.505***	13.139***	10.747***

Notes: *Z* reports the stationarity test assuming correlated errors and heterogeneous individuals. Two cases are considered: when *T* is infinite or not. *BCNW* stands for the long-run variance computed according to Bartlett kernel function with common Newey and West truncation lag. *BINW* denotes the long-run variance computed according to Bartlett kernel function with individual Newey and West truncation lag. *QSA* stands for the long-run variance computed according to a Quadratic Spectral function and individual lag truncation parameters. *** significant at 1%, ** significant at 5%, * significant at 10%.

Moon, Perron & Phillips (2007) Test

Model		FDI	SO_2	Openness
No Constant	$CPO_{(c=1)}$	0.097	-3.839	4.717
	$CPO_{(c=1.5)}$	0.879	-6.44	11.476
	$CPO_{(c=2)}$	2.223	-7.927	20.152
Constant	$CPO_{(c=1)}$	-2.672	9.195	4.581
	$CPO_{(c=1.5)}$	-4.972	19.917	9.999
	$CPO_{(c=2)}$	-7.026	32.041	16.129
Constant and Trend	$CPO_{(c=1)}$	0.746	1.812	0.834
	$CPO_{(c=1.5)}$	-13.447	-9.184	-13.095
	$CPO_{(c=2)}$	-24.202	-14.609	-23.411

Notes: *CPO* denotes the common-point optimal statistics which depends on the exogenous parameter *c*, whose values between 1 and 2 seems to provide a good balance between *N* and *T*. *** significant at 1%, ** significant at 5%, * significant at 10%.

6.7 Second Generation Panel Unit Root Tests

Choi (2006) Test

Model		FDI	SO_2	Openness
Constant	*Pm*	-0.926	0.491	1.849*
	Z	1.465	-0.755	-1.572*
	L^*	0.78	-0.376	-0.848
Constant and Trend	*Pm*	-0.889	-0.889	-0.889
	Z	1.255	1.255	1.255
	L^*	0.652	0.652	0.652

Notes: *P* denotes the inverse chi-square (Fisher) test statistic, while *Z* is the inverse normal (Choi) and L^* is the logit test. *** significant at 1%, ** significant at 5%, *significant at 10%.

Pesaran (2007) Test

Model		FDI	SO_2	Openness
No Constant	*CIPS*	-16.787***	-3.351***	-4.431***
	$CIPS^*$	-1.909***	-1.753***	-2.162***
Constant	*CIPS*	-8.7E+05***	-9.6E+05***	-14E+0.5***
	$CIPS^*$	-1.451	-3.164***	-2.437***
Constant and Trend	*CIPS*	-0.46	3.394	0.115
	$CIPS^*$	-0.382	-0.094	-0.34

Notes: *CIPS* (*CIPS**) denotes the average of the (truncated) individual cross-sectionally augmented CADF statistics. *** significant at 1%, ** significant at 5%, * significant at 10%.

Breitung & Das (2005) Test

Model		FDI	SO_2	Openness
Constant	t_{rob}	-1.451*	0.652	0.089
Constant and Trend	t_{rob}	-0.476	-0.44	0.364

Notes: t_{rob} denotes robust ADF t-statistics. *** significant at 1%, ** significant at 5%, * significant at 10%.

Demestru, Hassler & Tarcolea (2006) Test

Model		FDI	SO_2	Openness
No Constant	$t_{k_1}(\lambda=1)$	2.773	-2.603***	4.404
	$t_{k_2}(\lambda=1)$	3.891	-3.652***	6.178
	$t_{k_1}(\lambda=\sqrt{i})$	2.366	-2.389***	4.01
	$t_{k_2}(\lambda=\sqrt{i})$	3.261	-3.292***	5.527
Constant	$t_{k_1}(\lambda=1)$	2.772	-2.558***	0.146
	$t_{k_2}(\lambda=1)$	3.888	-3.589***	0.205
	$t_{k_1}(\lambda=\sqrt{i})$	2.272	-2.912***	0.111
	$t_{k_2}(\lambda=\sqrt{i})$	3.132	-4.013***	0.153
Constant and Trend	$t_{k_1}(\lambda=1)$	1.161	-1.335*	0.819
	$t_{k_2}(\lambda=1)$	1.629	-1.872**	1.149
	$t_{k_1}(\lambda=\sqrt{i})$	2.366	-2.389***	4.01
	$t_{k_2}(\lambda=\sqrt{i})$	3.261	-3.292***	5.527

Notes: t_ϕ denotes the weighted linear combination of individual probits. *** significant at 1%, ** significant at 5%, * significant at 10%.

Chapter 3

Complex FDI and Environmental Regulation[24]

1 Introduction

The growing role of multinational enterprises associated with the progressive liberalization of foreign direct investment (FDI) regimes has brought attention toward the environmental impact on host countries. One of the most controversial debate today is whether pollution-intensive industries relocate to countries with less stringent environmental policies, turning these countries into pollution havens.

In this context, the pollution haven effect (PHE) refers to the possibility that FDI could be sensitive to weaker environmental regulation, especially when polluting firms want to avoid the costs associated with environmental standards compliance. While the intuitive logic behind the PHE is straightforward, it is fair to say that empirical studies designed to test this hypothesis have so far shown little evidence and suffer potentially from omitted variable bias, specification and measurement errors. Most empirical studies rely on a two-country framework, which ignores spatial dependence in multilateral FDI decisions. In particular, one of the underlying assumption of the existence of a PHE lies in the vertical structure of FDI. In fact, the consensus is that vertical FDI results from difference in relative endowments, while horizontal FDI is determined by similarity in countries' size and relative factor endowments. Accordingly, MNEs, which locate their production plant in the country whose environmental regulation is laxer and import back the product to be serve in the domestic market, engage in vertical FDI. This explains why there is a large empirical literature that analyzes indirectly the PHE by using trade data (Levinson and Taylor, 2007). From this perspective, the lack of PHE evidence could be explained by a greater prevalence of horizontal FDI. MNEs mainly motivated by the local market access will only be interested in duplicating the production plant in another country and serve its market, where cost differences (including compliance costs of environmental regulation) between the home and host country are of second-order importance.

[24] Written in collaboration with Madina Kukenova, PhD Candidate, University of Lausanne.

Yet, the inclusion of third-country effects, which capture the effect of proximity of other neighborhood host countries to a particular host country, is necessary to explain the emergence of new types of integrated FDI. In particular, complex FDI emerges when multinationals engage FDI in several host countries simultaneously in order to exploit the comparative advantages of each country with respect to a given part of the production process. This hybrid form of FDI allows MNEs to allocate the most pollution part of the production process to a country with laxer environmental regulation and the remaining of the production chain to other countries. Once processed, the product can be shipped to the home country to be consummed. As emphasized in the spatial econometrics literature, failure to account for third-country effects may lead to biased inference. This may be particularly problematic in the context of empirical studies of the PHE for three interrelated reasons. First, the impact of environmental stringency is not homogenous across different types of FDI. Vertical and complex FDI are expected to be more sensitive to environmental stringency than other forms of FDI. Second, environmental policies have been shown to be spatially correlated. Finally, it has proven extremely difficult to find credible instrumental variables for environmental regulation. As a consequence, the conclusions drawn from studies which fail to address those issues, remain questionable.

This paper investigates the pollution haven effect at the world-wide level taking into account the so-called third-country effects. Since the pollution haven effect is more likely to be the result of differences in environmental stringency between developed and developing countries, I examine bilateral FDI flows using an extended OECD investment database which covers a large number of developed and developing host countries for a long sample period (1981-2000). The conceptual framework and methodology applied in this study are intended to address a number of difficulties discussed above. First, I consider a framework that is able to distinguish between types of FDI by accounting for the presence of spatial dependence in FDI decisions. Second, I recognize the strategic nature of environmental stringency between countries in order to reduce potential variables omission in the model specification. Third, I use different complementary measures of environmental stringency, namely SO_2 emission per capita, CO_2 emission per capita and international environmental treaties. Each proxy used in this study relies on different underlying assumptions, which allows to take into account the different facets of environmental stringency. Last but not least, I provide a thorough treatment to simultaneity, endogeneity bias and spatial characteristics of the data, by applying system GMM to the spatial dynamic gravity-like panel model.

To our knowledge, this paper is the first attempt to analyze the pollution haven effect with bilateral FDI at the worldwide level from a third-country perspective. This is also the first time that the most prevalent type of FDI inflows is being investigated for more than a single parent country (e.g. USA). The empirical findings confirm the existence of a robust negative relationship between FDI and environmental stringency (proxied by SO_2), when endogeneity and spatial dependence are addressed properly. In particular, failing to account for spatial dependence of environmental stringency can mask the presence of a pollution haven effect. Furthermore, results suggest the prevalence of complex vertical integrated FDI from high income OECD countries to less developed countries which is in line with the existence of a pollution haven effect.

The remainder of this paper is organized as follows. Section 2 discusses the different reasons that lie behind the inclusion of third-country effects and briefly presents an overview of the previous empirical literature on FDI-PHE linkages. Section 3 describes the model specification, the data selection and the econometric procedure. Section 4 and 5 report the empirical analysis of the model and robustness check, respectively. Finally, section 6 concludes.

2 Spatial Linkages between FDI and Environmental Stringency

Until recently, most multinational enterprises' (MNE) motivations analyzed in the literature relied on a two-country framework, where FDI between home country *i* and host country *j* was only affected by both countries' characteristics. However, new types of FDI have emerged in the last twenty years. Multinationals can allocate FDI in a host country but can also engage simultaneously in trade or FDI in a third country. The location decision of these new type of FDI will not only depend on the home and host countries' determinants, but also on the factors of the host's neighborhood countries. These more complex, integrated FDI, are embedded in a multilateral decision-making process, which means that FDI decisions across various host countries are not spatially independent. Other elements may also lead to interdependent FDI decisions across host countries, including imperfect capital markets and agglomeration externalities (spillovers) which limit the necessary funds a multinational company has to commit abroad.

Based on the theoretical work by Markusen and Maskus (2002), Yeaple (2003) and Baltagi et al., (2007), Blonigen et al. (2007) propose a partial equilibrium model and an empirical framework to assess four different types of FDI: horizontal, vertical, export-platform and complex FDI. In order to distinguish which type of FDI is the most prevalent, they propose to estimate a spatial autoregressive model which includes two spatial variables: (1) a spatial lag dependent variable representing the spatially-weighted sum of bilateral FDI from a given host country to other host countries and (2) a market potential variable corresponding to the spatially-weighted sum of other host countries' market size (see the next section for further details). Table 1 summarizes the expected sign of both variables for the different categories of FDI at the firm level, under the assumption that the firm is already a MNE[25]. The distinction between types of FDI is also important in the analysis of the pollution haven effect, because these four forms of FDI respond differently to the host and neighboring countries' environmental stringency. I, therefore, extend Blonigen et al.'s analysis to account for the MNE's behavior in terms of spatial and environmental stringency responsiveness.

Table 1: Spatial and Environmental Linkages in MNE Choices

FDI Motivation	Horizontal FDI	Vertical FDI	Export Platform FDI	Complex FDI
Spatial Lag	0	–	–	+
Market Potential	0	0	+	+/0
Environmental Stringency	–/0	–	–/0	–
Spatial Environmental Stringency	0	+	+/0	+

Source: Blonigen et al. (2007, p.1308), Baltagi et al. (2007, p. 263)

Horizontal FDI relies on the search of opportunities to sell in foreign markets. A MNE in home country *i*, which wants to serve foreign markets *j* and *k*, can export the products or launch a production unit in both host countries. Horizontal FDI is more likely to happen with sufficiently high trade costs between the home country *i* and countries *j* and *k*. In terms of third-country effects, the decision to undertake FDI in country *j* is more likely to be independent to the decision regarding country *k* and its market potential. In terms of environmental stringency, horizontal FDI is *a priori* not especially sensitive to differences in environmental costs. This is because horizontal activities of MNEs are mainly motivated by market access rather than a reduction in production costs. In other words, multinationals engage in horizontal FDI when

[25] In other words, the trade-off between engaging in FDI and exporting is no longer relevant.

duplicating a production plant in a foreign country outweighs the loss of scale economies associated with a single plant. This would no longer be the case if the multinational has to meet some environmental/health standards to be authorized to sell the products on the host market.

Vertical FDI aims at obtaining low price resources or access to critical resources not available in the home economy. Since resource-seeking FDI is driven by factor cost differences between the home and host countries and not by market potential considerations, vertical FDI from home country *i* to country *j* will be undertaken at the expense of vertical FDI from *i* to another host country *k*. High environmental standards in country *j* will affect negatively FDI from home country *i* to country *j*, while higher environmental stringency in neighboring country *k* will have a positive effect on FDI from *i* to host *j*, other things being equal. In the end, the MNE might choose location *j* over location *k*.

Export platform FDI is used to serve regional export markets through a platform for production and sales (Yeaple, 2003; Bergstrand and Egger, 2007). This type of FDI has elements of both vertical and horizontal FDI. The specific location within a region is defined by cost considerations, as in vertical investments, while the sales in an integrated market respond to horizontal FDI considerations. In terms of environmental stringency, export platform FDI can be affected in two different ways. On the one hand, if the purpose of FDI is to serve export markets in developed countries through a platform production in a developing country and the access to these markets depends on the environmental/health product standards of the developed countries, the host country *j*'s environmental regulation is no longer important. In this case, FDI to country *j* is likely to be associated with new techniques, including the latest abatement technologies (California effect), making the environmental stringency in the host country *j* no longer relevant. On the other hand, if the multinational wants to serve neighboring developing countries, the environmental stringency in the host country still matters. The production-location decision in host country *j* will be negatively affected by higher environmental stringency in host *j*, but environmental standards between close countries (*j* and *k*) should be relatively close so the MNE can serve both markets using the same production process. If the company expects in the future an increase in environmental stringency in the host country *j* and its neighboring countries, it may choose today a production process that will meet higher standards in the future. In this case, the environmental regulation in country *j* would no longer matter, since the MNE no longer experiences a comparative advantage when locating in country *j* with lower environmental regulation.

Complex (vertical) FDI is characterized by a multinational firm from home country *i* which owns not only a production plant in host country *j* but also one in third country *k*, in order to exploit the comparative advantages of various locations. This type of FDI is associated with exports of intermediate inputs from affiliates (*j* and *k*) to another third market for further (or final) processing, before being dispatched to its final destination. The search for (low cost) suppliers in multiple (close) countries leads to the slicing-up of the value-chain of the production process. If both adjacent host countries *j* and *k* present similar supply network characteristics, the MNE may find it advantageous to launch production in host *k* given that it already owns production plants in (contiguous) host country *j*. Thus, complex FDI are characterized by a complementary relationship: complex FDI from home country *i* to third country *k* constitutes a complement for FDI from home country *i* to host country *j*. This positive relationship is strengthened if *j* and *k* are neighboring countries. Market potential in this type of FDI should not matter, although the level of industrial production in neighborhood countries should be positively correlated with higher opportunities for vertical suppliers (e.g. agglomeration incentives). This last category of FDI is particularly sensitive to environmental stringency in a given host and its neighboring countries, because the most polluting stage of production is more likely to be located in the host country characterized by less environmental stringency. The intermediate input will then be exported to one or more third-country for further processing, in order for the final good to be shipped to its final destination (e.g. home market).

This study is based on two different types of literature. The first one is related to the application of spatial econometrics in empirical studies of the determinants of FDI, while the second focuses on the linkages between FDI and the pollution haven effect. The next sections attempt to shed some light on both literatures, with an emphasis on the most recent works.

2.1 FDI Literature Review

Empirical FDI studies allowing for the impact of third-country effects and applying spatial econometrics are sparse. Despite mixed evidence, these studies highlight the importance of spatial interdependence. Coughlin and Segev (1999) are the first to apply spatial econometric techniques to study the geographic distribution of FDI. Their results indicate that FDI into one location within China is positively associated with FDI into other close Chinese locations. Baltagi et al. (2007) study the third-country effects associated with US outbound FDI for seven manufacturing industries across both developed and less-developed destinations. They find substantial evidence of spatial interactions, though they cannot definitively conclude whether export-platform or complex vertical FDI is more prevalent. Blonigen, et al. (2007) focus on aggregate U.S. outward FDI to OECD countries at the country level. While the estimated relationships of traditional determinants of FDI are robust to the inclusion of spatial interdependence, export-platform FDI seems to have greater prevalence, although the results are quite sensitive to the sample of countries examined. Blonigen et al. (2008) consider also inbound FDI from OECD countries to the US. They find strong and robust evidence for parent market proximity effects but it mainly depends on the sample selection. More recently, following Blonigen et al. (2007), Garretsen and Peeters (2009) estimate a spatial lag model for Dutch FDI to 18 host countries. Based on maximum likelihood estimations, third-country effects do matter, but are also sensitive to sample and model selection.

2.2 Pollution Haven Effect Literature Review

The existing empirical research on environmental regulations and FDI linkages displays mixed results depending on the type of studies. The first type of papers, which focus on inflows of FDI to a single country at the regional and/or industrial level, find some evidence of pollution haven effect (see for instance List and Co, (2001) for the United States and Dean et al. (2008) for China). Drukker and Millimet (2007) assess the presence of third-country effects in the determination of inbound US FDI by estimating a spatial error model with spatially weighted covariates. The authors find that many neighboring states attributes, including environmental stringency, influence FDI location.

The second strand of studies, which analyze outflows of FDI from a single home country to one or more host countries at the aggregated or industrial level, display mixed findings (Xing and Kolstad, 2002; Eskeland and Harrison, 2003; Cole and Elliot, 2005). More recently, Wagner and Timmins (2008) find that the German chemical industry is the only pollution-intensive sectors in Germany to have relocated to countries with less stringent regulation once agglomeration effects (through a lagged dependent variable) are taken into account.

The third category of papers considers inflows of FDI to different countries originating from various home countries at the aggregated or industrial level. Most of these studies fail to find evidence of a negative pollution haven affect (Javorcik and Wei, 2001; Eskeland and Harrison, 2003). Cole et al. (2009) argue that the lack of evidence might come from the fact that these papers fail to treat environmental policy as endogenous with respect to FDI. In particular, the effects of FDI on the environmental policy might be conditional on the government's political institutions sensitivity to corruption and lobbies pressure.

Overall, the existing FDI-pollution haven effect empirical literature faces a number of limitations, which may partially explain the ambiguity in the results obtained. These drawbacks include different conceptual frameworks, data sources and proxies as well as different econometric methodologies. They are addressed in the next section.

3 Model Specification

This section presents the spatial dynamic panel model in a gravity setting to account for third-country effects. The selected variables and the different spatial weight matrices considered in this study are presented. Finally, the estimation procedure is discussed.

3.1 Spatial Dynamic Gravity Model

Given the relative success of the traditional gravity equation in explaining the trade flow between countries, recent theoretical models (Head and Ries, 2008) suggest that location and size of bilateral FDI flows depend on country characteristics such as country size, population and factor endowments. According to Evenett and Keller (2002), the gravity model can support both assumptions of product differentiation (increasing returns to scale) and homogenous good production. Since I am interested in the detection of a substitutive or complementary relationship in the allocation of FDI between host countries, I follow Blonigen et al. (2007) and model bilateral FDI flows in a spatial autoregressive gravity panel model[26]. In addition, I acknowledge the fact that FDI decisions are part of a dynamic process and extend the model to a dynamic panel framework:

$$Y_{ijt} = \delta Y_{ijt-1} + P_{it}\alpha + H_{jt}\beta + X_{ijt}\gamma + \rho[W_t Y_t]_{ijt} + [W_t H_t]_{jt}\eta + \varphi_{ij} + \mu_t + u_{ijt} \quad (1)$$

[26] Another possibility would be to introduce spatial dependence in the error term and estimate a spatial error model. However, as mentioned by Blonigen et al. (2007), FDI theory provides no real guidance whether or not to expect positive or negative spatial autocorrelation in a spatial error model, while the inclusion of a spatial lag dependent variable has theoretical foundations. Moreover, from an econometric point of view, the omission of spatial error dependence does not affect the consistency of the results.

where Y_{ijt} is FDI flow from home country *i* to host country *j* at period *t*. P'_{it} is a vector including parent country variables, H'_{jt} includes host variables and X'_{ijt} represent bilateral control variables. W_t denotes the spatial weight matrix, which is assumed to be non-stochastic, row-standardized and exogenous to the model[27]. The diagonal elements of W_t are set to zero so that no observation predicts itself. Thus, $[W_tY_t]_{ijt}$ represents the spatially weighted average of FDI flow from home country *i* to the neighborhood countries of country *j*. The associated coefficient ρ, called the spatial lag, allows to determine if FDI flow from country *i* to country *j* is (positively/negatively) affected by FDI from country *i* to other neighboring host countries. The parameter ρ is assumed to lie between -1 and +1 to ensure spatial stationarity. The vector $[W_tY_t]_{ijt}$ includes spatially weighted host variables to account for potential third-country effects (e.g. market potential). φ_{ij} is the individual effect, that captures unobserved characteristics related to country-pair, which do influence bilateral FDI but are fixed in the short and medium terms. The time fixed effect, μ_t, captures the business cycle common to all countries. Finally, u_{ijt} is the error term.

The spatial dynamic panel model is a flexible framework that allows to capture the complicated dependence structure between economies. However, the coefficients estimate represents the (unobservable) pre-interdependent dynamic impact of each explanatory variables. In order to measure the long-run, steady-state, equilibrium of the model, one can compute spatial multipliers. To see this point more formally and assuming the steady-state prevails ($Y_t = Y_{t-1}$), the reduced form of equation (1) can be rewritten in matrix notation as follows:

$$Y_t = \left[I - \rho W_t + \delta I_N\right]^{-1}\left(P_t\alpha + H_t\beta + X_t\gamma + W_tH_t\eta + \varphi + \mu_t + u_t\right) \qquad (2)$$

where *I* is the $N \times N$ identity matrix. $S = \left[I - \rho W_t + \delta I_N\right]^{-1}$ represents the $N \times N$ spatial multipliers effects matrix through which I can analyze the impact of a shock of a given explanatory variable in a country on the FDI inflows to this country and other contiguous economies. Following Franzese and Hayes (2006), the standard errors of these spatial effects are computed using the delta method (i.e. a first-order Taylor linear approximation to the non-linear expression *S* around the estimated parameters).

[27] As distances are time-invariant, it will generally be the case that $W_t = W_{t-1}$. However, when dealing with unbalanced panel data, this is no longer true. Missing neighborhood observations are problematic because they are treated as zeros. This may likely induce bias in the estimation and interpretation of the results. According to Baltagi et al. (2005), this constitutes an important issue for future research. However this "border problem" should be smaller with a distance-based weighting scheme and large averages distances between locations than for contiguity-based weighting schemes.

3.2 Variables Selection

The dependent variable of the model is bilateral flows of FDI from 26 OECD countries to 79 host countries for the period 1981 through 2000[28]. The data, expressed in current US dollars, are taken from the OECD International Direct Investment Statistics website. Working on flows rather than stocks provides several advantages including more available data and less error measurements, although stocks present less volatility than flows and constitute a better measure of capital ownership.

I specify our model in log-linear form (except for the dummies variables), because such a model is more likely to lead to well-behaved residuals given the skewness of most FDI data samples. A problem that arises when using a log-linear specification is how to deal with observations with negative and zero values. This is the case for FDI inflows which are negative when the home country repatriates previous investments made in the host country. In order to handle the presence of zero/negative FDI flows I apply the following log transformation to variables with negative or zero values: $z = \ln(x + \sqrt[2]{x^2 + 1})$ (Busse et al., 2010).

The variables selection of the determinants of FDI is mainly dictated by data availability. Among the control variables, I include GDP, GDP growth, the share of the dirty industries in total value-added, a risk index, a measure of capital openness (Chinn and Ito, 2007), a natural resource endowment dummy as well as bilateral distance between capitals and a colonial links dummy. All monetary variables are expressed in US dollars. Appendix 7.2 lists the variables considered and their sources.

In addition, I also include a lagged dependent variable to account for agglomeration effects[29]. More FDI in a host country seems to attract more FDI in this same host country. These persistence effects are partly due to the fact that FDI is often accompanied by physical investments that are irreversible in the short run. Furthermore, Wagner and Timmins (2008) demonstrate that failing to account for agglomeration effects can mask the evidence of a pollution haven effect. The lagged FDI can also partially capture infrastructure in the host

[28] Appendix 7.1 lists the host and source countries. Note that countries with a population lower than one million of inhabitants have been dropped in order to exclude tax haven countries (e.g. Bahamas).

[29] Agglomeration effect is measured by lagged FDI inflows, because the inclusion of FDI stocks leads to multicollinearity problems.

country. High physical (e.g. roads and power) and social (e.g. health and education) infrastructure as well as urbanization influence positively FDI inflows. Therefore, I expect the autoregressive coefficient to be positive. However, its sign could be negative because of congestion, when firms compete with one another through price bidding to downstream industries in the region.

3.3 Environmental Stringency

Measuring environmental stringency is the key problem in the PHE literature, because environmental policy can take many forms: environmental fees or taxes, permitting costs, regulatory delays, emissions limits that require installation of costly technology, the threat of lawsuits, product or process redesign, forgone output, and so forth (Levinson and Taylor, 2008). Unfortunately, this type of information is not available for most countries. Reductions in emissions may be viewed as proxies for a host country's effective enforcement of environmental policies. Therefore, I rely on three complementary measures of the relative level of stringency based on the actual level of pollution emission and the number of international environmental treaties[30].

The level of sulphur dioxide SO_2 emissions, taken from Stern (2005), constitutes a good measure of air pollution. First, SO_2 per capita is one of the most significant pollutants worldwide. Milner et al. (2007) show that the reduction in SO_2 emission is highly correlated with the level of environmental funds in former Soviet Union countries. Second, it is highly correlated with other pollutants (Xing and Kolstad, 2002). Third, to the extent that pollution reduction is a public good, it suffers less from a free-ride problem and is available for a large number of host countries. That is the reason why, environmental stringency is proxied by the level of SO_2 per capita emissions multiplied by -1.

Air pollution can also be measured by the level of carbon dioxide CO_2 emissions, provided by the World Development Indicators (Hoffman et al., 2005). Just like SO_2 emissions, CO_2 series are constructed from fuel consumption data, rather than directly observed. The use of CO_2 emissions as a proxy relies on strong assumptions. Some critics argue that CO_2 emissions do not

[30] Other environmental proxies can unfortunately not be considered because of their lack of availability in terms of time period or/and covered countries (e.g. lead-content per gallon gasoline, pollution intensities; water pollution; fertilizer; WEF's environmental sustainability index, index of environmental sensitivity performance; number of deaths related to pollution; number of environmental NGOs, number of ISO 14001 licenses).

reflect only environmental stringency, but also the energy intensity of production. Another problem is that the pollution consequences of CO_2 emissions are subject to the free-ride problem because the damages caused by CO_2 emissions are global. There are fewer incentives for a government to modify its environmental policy, which makes it difficult to use CO_2 emissions as a proxy for environmental stringency. Therefore, I expect to find less evidence of a PHE with CO_2 per capita emissions.

The number of participation in international environmental treaties, taken from the Environmental Treaties and Resource Indicators website, can also constitute a cross-country proxy, although its use relies on several strong assumptions. First, one implicitly assumes that a ratified treaty will automatically translate into stringer environmental stringency. This would probably be true only if there were sanctions for the non-respect of the treaty. Second, each country which signs a treaty will implement the exact same instruments in terms of cost and mechanism to comply with the regulation, which is unrealistic. Therefore, I expect to find less evidence of a pollution haven effect when considering environmental treaties. Following Xing and Kolstad (1998) as well as Javorcik and Wei (2001) I construct a variable of environmental treaties that reports the number of signed treaties.

Independently of the proxy used, I hypothesize a negative relationship between environmental stringency and FDI. Higher environmental standards (ambient quality standards, emission standards, production process standards and products standards) leading to higher environmental costs can deter inflows of FDI. However, it is also possible that environmental stringency can attract FDI in order to gain a competitive advantage through higher standards, which would validate the so-called pollution halo hypothesis. As mentioned in the literature review, previous studies (Ederington et al., 2005; Cole et al., 2009; Kellenberg, 2009) suggest that not only FDI is sensitive to environmental stringency, but also that environmental policy can be affected by FDI, when the level of corruption and lobby pressures are high or when environmental standards are used as a strategic trade policy. In both cases, this could lead the government to set higher or lower level of environmental policy than it is socially efficient. As a result, environmental stringency has to be treated as (potentially) endogenous.

3.4 Third-Country Effects

In order to account for the role of spatial dependence and interaction in the data, I include spatially weighted variables. Each spatial variable is computed using the same spatial weight matrix, except the spatial lag variable because of the bilateral nature of the data. In the construction of the weights themselves, the theoretical foundation is quite general and the particular functional form of any single element in W_t is not prescribed. In fact, the determination of the proper specification of W_t is one of the most difficult and controversial methodological issues in spatial data analysis. Prior to discussing the weighting scheme used, it is important to note that a miss-specification of the weighting matrix can bias the results. In practice, different weight matrices should be compared to find the most proper one.

A general spatial matrix W_t can be defined by a symmetric binary contiguity matrix. This weighting scheme assigns a weight of zero to non-contiguous countries and a weight of one to all contiguous countries. Since the contiguity matrix cannot differentiate the degree of spatial linkages between adjacent locations, some more complex spatial weighting matrix can be used. For example, one can choose a simple inverse distance function, where each pair of location *j* and *k* declines to $1/d_{j,k}$ if $j \neq k$. However, the inverse distance matrix has the disadvantage of always giving some positive weight to very remote countries (with weaker cultural, political and economic ties). A compromise can be reached by allowing the third-country effects to decay at a faster rate by giving more weight to locations within the same region and almost zero to locations outside the region[31]. This is done by dividing the distance between locations *j* and *k* by the minimum distance within the region *r* (where location *j* lies within region *r*):

$$w_t(d_{j,k}) = \exp\left(\frac{-d_{j,k}}{MIN_{r,j}}\right) \quad if\ j \neq k.$$

Following Blonigen et al. (2007), the estimation of complex integration strategies of multinationals is done by including a spatial lag bilateral FDI, $[W_t Y_{it}]_{ijt} = \sum_{k \neq j} w_t(d_{k,j}) \cdot Y_{ikt}$, and a variable capturing market potential in neighboring host countries. According to Head and Mayer (2004), which apply different measures of host country market proximity in their analysis of Japanese outbound FDI into Europe, a distance-weighted sum of close countries' GDPs yields the best fit for the data. Thus, for a given host country *j* in year *t,* the market potential variable is defined as the spatially weighted sum of GDPs of all other countries: $[W_t GDP_t]_{jt} = \sum_{k \neq j} w_t(d_{k,j}) \cdot GDP_{kt}$.

[31] Six geographical regions are considered in this study: (1) North America; (2) Latin America and Caribbean; (3) Europe and Central Asia; (4) East Asia, Pacific and South Asia; (5) Sub-Saharan Africa; (6) Middle East and North Africa.

To capture the fact that countries do not define environmental regulation independently (Eliste and Fredriksson, 2001), I also include a spatially weighted environmental stringency that captures the spatial interdependence of environmental stringency: $\left[W_t Env.Stringency_t\right]_j = \sum_{k \neq j} w_t(d_{k,j}) \cdot Env.Stringency_{kt}$. The inclusion of this additional spatial variable is particularly important when one considers SO_2 per capita emission as a proxy. According to the geographical location and prevailing wind patterns, levels of acid deposition from SO_2 in one country is partially determined by emissions in neighboring countries. As a consequence, any country will have an incentive to behave strategically with respect to emissions originating in neighboring countries. Empirical evidence suggests also that the propensity of a given host country to ratify treaties is positively correlated with the number of treaties signed in the neighborhood countries (Davies et al., 2006).

Based on the prevailing type of FDI, the expected sign of the spatially weighted variables will be different (see Table 1). I estimate the model with contiguity, inverse distance and negative exponential distance spatial weight matrices. But, for the sake of brevity, I only present the results associated with negative exponential distance. Note that the use of data at the country level can only capture net effects. For instance, the spatial lag coefficient may, on average, be no different from zero but this could simply be the result of export-platform and complex vertical FDI effects cancelling out.

3.5 Econometric Issues

The spatial dynamic lag model (1) faces simultaneity and endogeneity problems, which makes OLS estimation biased and inconsistent. In fact, one can analytically demonstrate that the spatial lag term $W \cdot Y$ is correlated with the disturbances, even if u are independently and identically distributed. Each element of the dependent variable depends on a linear combination of all of the error terms. Therefore, the spatial lag term must be treated as an endogenous variable and proper estimation methods must account for this endogeneity. The spatial dynamic panel model is usually estimated using spatial (dynamic) maximum likelihood (ML) or spatial two stage least squares (2SLS). However, the main drawback of applying those spatial estimators is that, while the spatial lagged variable is considered endogenous, other explanatory variables are not. If other FDI determinants are endogenous or potentially endogenous, which is likely the case in our study, no instrumental treatment is applied to control for this econometric problem.

Kukenova and Monteiro (2009) suggest to apply system GMM (Blundell and Bond, 1998) in a spatial dynamic framework. System GMM corrects for the endogeneity of the time and spatial autoregressive variables and other potentially endogenous explanatory variables by using lagged values of the dependent, endogenous and exogenous variables. In addition, system GMM allows to address other econometrics problems such as measurement error and weak instruments. It controls for time-invariant country-specific fixed effects such as culture and political structure. On a practical ground, it also avoids the inversion of the high dimension spatial weights matrix W and the computation of its eigenvalues, which can be sometimes computationally unfeasible to estimate the model or involve accuracy problems (here the spatial weights matrix W is 10'657 $\times$ 10'657). Based on extensive Monte Carlo simulations, Kukenova and Monteiro demonstrate that system GMM estimates consistently the spatial lag variable when N is large and T is moderate, as it is the case here.

The system GMM estimator consists of estimating equation (1) as a system of two equations, one in levels and the other one in first-differences. Lagged first-differences are treated as instruments for equations expressed in levels, while lagged levels are used as instruments for equations in first-differences. The consistency of system GMM relies on the validity of the moment conditions, which depends on the assumption of absence of serially correlation of the level residuals and the exogeneity of the explanatory variables. Therefore, it is necessary to apply specification tests to ensure that these assumptions are justified. Two specification tests are available to verify the consistency of the GMM estimator. First, the overall validity of the moment conditions is checked by the Sargan/Hansen test. The null hypothesis is that instruments are not correlated with the residuals. Aware that too many instrument variables (exceeding the number of cross-section units) tend to validate fallacious results through the Hansen J test for joint validity of those instruments, as well as the difference-in-Sargan/Hansen tests for subsets of instruments, I apply system GMM by collapsing the moments conditions, i.e. avoid separating the instruments for each period (Roodman, 2009).

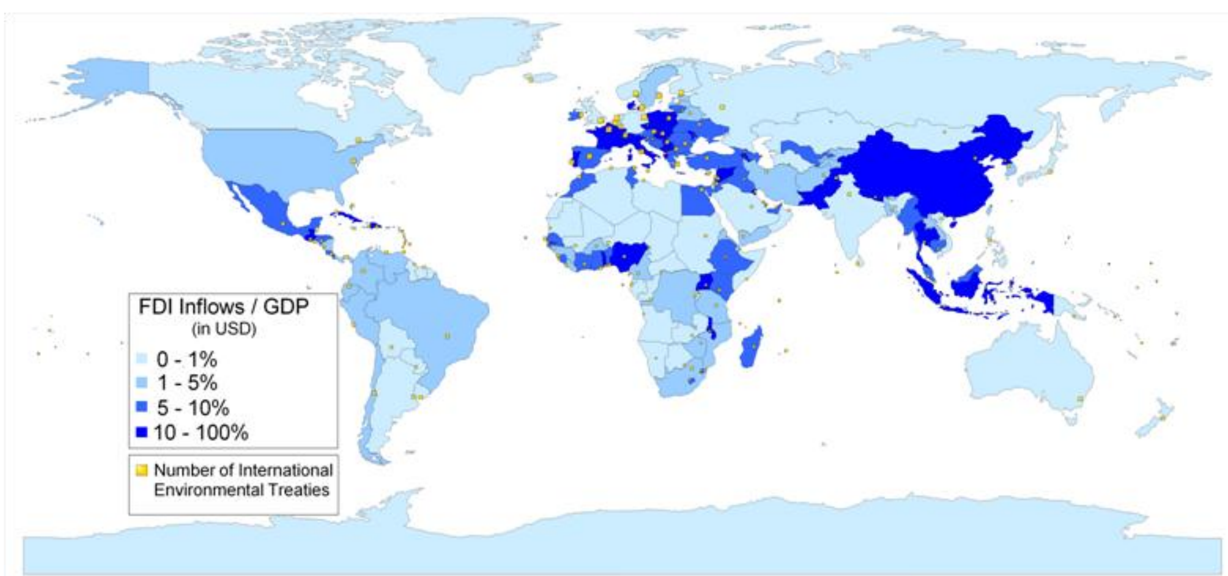
In this study, the system GMM estimation relies on the following instruments structure. The variables considered as endogenous (lagged FDI, spatial lagged FDI and environmental stringency) are instrumented by their second and third lag values. The variables potentially endogenous and predetermined (host GDP, growth of GDP, country's risk) receive the same treatment. To account for the fact that the capital openness index is relatively constant over time, I instrument it using its 5 first lags only for the equation expressed in first-difference.

The capital-labour ratio and manufacturing variables are used as additional external variables and treated as predetermined. GDP of the source country is treated as strictly exogenous. The time fixed effects, the natural resources dummy as well as the gravity variables (distance and colonial links) are also treated as strictly exogenous, but used only in the first-difference specification. Following Xing and Kolstad (2002), I also include population density as an external exogenous variable. The latter is an indicator of congestion and the ability of pollutants to naturally disperse away from population centers. This instrument is unlikely to be correlated with the error term, that is why it is used in level and first-difference. I also correct the standard errors for small sample bias by applying the robust Windmeijer correction to the two-step GMM estimates.

4 Panel Estimation Results

Before estimating the model, it is always interesting to proceed at a graphical exploratory analysis of the relationship between inflows FDI and environmental stringency. Figure 1 depicts the spatial location of OECD's FDI outflows to the different host countries and their level of environmental stringency proxied by the number of international environmental treaties. First, it is interesting to note that most OECD's FDI takes place among OECD countries. Only a small fraction of FDI is allocated to less developed countries. Among these countries, some countries like Brazil or China can be considered as potential pollution haven countries, since they display a relatively low environmental stringency but attract a large amount of FDI from OECD countries.

Figure 1: Average FDI Inflows and Environmental Stringency (1981-2004)



As Figure 1 suggests it, FDI flows and environmental stringency seem to be spatially correlated. To check the presence of spatial correlation (i.e. coincidence of value similarity and locational similarity), I perform two of the most popular global spatial indicators: Moran *I* (1948) and Geary's *C* (1954) statistics[32]. Appendix 7.5. provides the results of the Moran and Geary tests for bilateral FDI, GDP and environmental stringency (proxied by SO_2 per capita and international treaties). Using different spatial weight schemes, the null hypothesis of no spatial correlation is rejected by both tests for most years for each variable. In particular, the results suggest the presence of positive spatial autocorrelation for each variable. This definitively calls for a systematic treatment of spatial interdependencies through a spatial dynamic panel framework.

I first present the results associated with the full sample to determine potential biases caused by omitting the spatial structure of the data and by not instrumenting for environmental stringency (proxied by SO_2 per capita emissions) and other potential endogenous variables. Then, since the motivation behind investment to developed and developing countries may be quite different, I estimate models for OECD and non-OECD host countries samples. The baseline model deliberately includes a limited number of explanatory variables in order to avoid multicollinearity problems and a decrease in the number of countries covered. The robustness check in the next section investigates the inclusion of additional FDI determinants and the use of other proxies for environmental stringency.

4.1 Baseline Results

Table 6 reports the results for the full sample (OECD and non-OECD host countries). The first three columns display the results without any spatial features for fixed effects, random effects and system GMM respectively. The other four remaining columns present the system GMM estimations including third-country effects (based on a negative inverse exponential distance). More precisely, I first add the spatial lag dependent variable, I then include separately the spatially weighted environmental stringency and market potential variable, to finally consider the three spatially weighted variables together. This sequence of specifications allows us to study the sensitivity of the results to the assumption of exogenous environmental stringency and the inclusion of third-country effects.

[32] The Getis and Ord (1992) test cannot be computed for all variables of interest in our study, since it can only be applied to positive values (FDI inflows can be negative for instance).

Despite the fact that most variables are significant and display the expected sign in the random effect regression, the Hausman specification test suggests that the fixed effect model is preferred over the random effect specification. Overall, the estimates associated with fixed effect are relatively poor. Only a few variables are significant. This should not come as a surprise since the fixed effect estimator does not correct for the endogeneity of some explanatory variables, including the lagged dependent variable and environmental stringency. When this issue is taken into account, as system GMM allows it, the results clearly improve. The estimated coefficients usually lie between the random and fixed effects results. The autoregressive parameter becomes significant, while the coefficient of environmental stringency switches from positive to negative. This finding suggests a positive endogenous bias in the pollution haven effect, when endogeneity is not corrected.

According to the Hansen test, the different specifications associated with system GMM are appropriately specified. As discussed in Roodman (2009), the Hansen test of overidentifying restriction is robust to heteroskedasticity but can suffer from instruments proliferation in the regression. Too many instruments can overfit the model and, at the same time, weaken the power of the Hansen test to detect overidentification. Since I deliberately limited the number of instruments to be always significantly smaller than the number of groups (general rule of thumb), I rely on the diagnostic of the Hansen test. Although the Sargan test of over-identification tests is not weakened by too many instruments, it is inconsistent in the presence of heteroskedasticity. That is why I pay less attention to this test statistic. The Arellano-Bond test for second-order serial correlation in the residuals (m_2) cannot reject the null of no correlation by any standard levels of significance for all specifications, which is additional support of the correct instrument specification of the model.

The results show that the most important classical determinants of FDI have the expected sign and are significantly different from zero. In particular, the measure of agglomeration effects (lagged bilateral FDI) is significant in all system GMM specifications and confirm a persistence effect in the FDI decision. The host and source countries' GDP are significant across all specifications which is fairly intuitive and in line with the claim that market size is the single most important factor in the investment location decision. In particular, large home countries are more likely to invest abroad, while large host countries are more likely to receive FDI. The standard gravity variables (bilateral distance and colonial links) are also significant across specifications, which confirms the positive impact of close cultural environment and proximity on the FDI allocation.

Table 2: Full Sample Results

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.012	0.198***	0.062***	0.056**	0.056**	0.054**
	[0.012]	[0.011]	[0.022]	[0.022]	[0.022]	[0.022]
GDP Host	0.850*	0.557***	0.684***	0.649***	0.576***	0.599***
	[0.450]	[0.033]	[0.137]	[0.126]	[0.142]	[0.144]
GDP Source	4.637***	0.802***	0.987***	0.602***	0.568***	0.593***
	[0.824]	[0.033]	[0.064]	[0.120]	[0.123]	[0.129]
GDP growth	0.05	0.089***	0.071	0.012	0.019	0.003
	[0.033]	[0.031]	[0.070]	[0.072]	[0.071]	[0.070]
V.A. Dirty Shares	-0.172	-0.122	0.409	-0.14	-0.242	-0.084
	[0.218]	[0.135]	[0.348]	[0.365]	[0.354]	[0.369]
Risk Index	-2.110***	-0.999***	-2.135**	-1.348	-0.735	-0.982
	[0.546]	[0.340]	[0.849]	[0.863]	[0.789]	[0.838]
Capital Openness	-0.018	0.152***	0.122	0.043	0.076	0.112
	[0.089]	[0.048]	[0.144]	[0.144]	[0.142]	[0.143]
Oil Resources		-0.119	-0.284	-0.434	-0.319	-0.288
		[0.111]	[0.290]	[0.265]	[0.290]	[0.286]
Colonial Links		0.872***	1.027***	0.980***	0.917***	0.872***
		[0.148]	[0.259]	[0.270]	[0.272]	[0.272]
Distance		-0.577***	-0.649***	-0.497***	-0.476***	-0.561***
		[0.039]	[0.070]	[0.084]	[0.085]	[0.107]
Environmental Stringency	0.348**	-0.305***	-0.679***	-0.428**	-0.466***	-0.500***
	[0.173]	[0.044]	[0.168]	[0.175]	[0.176]	[0.178]
Spatial Lag FDI				0.515***	0.541***	0.519***
				[0.145]	[0.153]	[0.157]
Market Potential					0.128	0.063
					[0.165]	[0.175]
Spatial Stringency						0.389
						[0.294]
Observations	8678	8678	8678	8678	8678	8678
Groups	699	699	699	699	699	699
R-squared	0.033	.	.	.	.	.
Instruments	.	.	145	148	151	151
Arrellano-Bond AR(2)	.	.	0.779	0.715	0.708	0.739
Sargan test	.	.	0.327	0.685	0.726	0.716
Hansen Test	.	.	0.488	0.663	0.619	0.642

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

By explicitly accounting for spatial dependence, the last three specifications allow us to highlight the most prevailing type of FDI originating from OECD countries (see Table 1). The fact that the spatial lag coefficient is positive and significant and the market potential is not significant rule in favour of a higher prevalence of complex FDI. However, one should be careful with the interpretation of these results. As noted by Blonigen and Davies (2004) combining rich and poor countries in FDI data can mask evidence or, in the worst case scenario, lead to implausible coefficient estimates. The next section takes this issue into account.

4.2 OECD vs. Non-OECD Host Countries

Since the motivation behind investment to developed and developing countries may be quite different, especially in the case of the PHE (Hoffman et al., 2005), combining FDI allocated to developed and developing countries may introduce undesirable noise into the data. In line with this consideration, I reestimate separately the different model specifications for two subsamples. Table 3 and 4 display the results for FDI to high-income OECD and to non OECD countries[33], respectively. Note that when I consider the restricted sample or OECD and non-OECD host countries, the third country variables are based on all countries.

Overall, the comments made in the previous section can be reiterated for both tables. Based on the Hausman test, the fixed effect model is preferred over the random effect model. The classical determinants of FDI are in most cases significant with the expected sign. Yet, failing to correct for potential endogeneity of environmental stringency yields a different conclusion in each sub-sample. In the OECD host sample, the environmental stringency has initially a positive effect, but after correction it is no longer significant. In the non-OECD host sample, the endogeneity problem tends to mask the presence of a negative and significative pollution haven effect. Note that the evidence of a pollution haven effect for less developed countries is in line with the theory. Following Edgerington et al. (2005), the non-significant coefficient of the value-added share of dirty industries as a percentage of total GDP might be attributed to the fact that the most polluting intensive industries are not necessarily the most likely to react to stringer environmental regulation because of high sunk costs impending them to be geographically mobile.

[33] Only high income OECD countries are dropped. Mexico, Turkey, Hungary, Poland, Czeck Republic and Slovak Republic remain in the sample, since they can be considered as potential pollution haven (Cole and Elliott, 2005).

Table 3: OECD Host Sample Results

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.01	0.186***	0.063**	0.054**	0.058**	0.055**
	[0.015]	[0.014]	[0.025]	[0.026]	[0.026]	[0.026]
GDP Host	1.189	0.562***	0.599***	0.655***	0.622***	0.616***
	[0.814]	[0.046]	[0.117]	[0.103]	[0.107]	[0.107]
GDP Source	5.905***	0.813***	0.947***	0.626***	0.631***	0.639***
	[1.075]	[0.042]	[0.071]	[0.141]	[0.145]	[0.143]
GDP growth	-0.052	0.009	-0.076	-0.094	-0.09	-0.089
	[0.062]	[0.061]	[0.092]	[0.094]	[0.094]	[0.094]
V.A. Dirty Shares	-0.385	0.053	0.025	-0.426	-0.449	-0.502
	[0.334]	[0.264]	[0.463]	[0.509]	[0.506]	[0.512]
Risk Index	-2.183	-1.709	-1.959	-1.52	-1.708	-0.822
	[1.552]	[1.175]	[2.571]	[2.661]	[2.509]	[2.437]
Capital Openness	-0.051	0.05	0.23	0.18	0.177	0.314
	[0.143]	[0.111]	[0.228]	[0.224]	[0.217]	[0.227]
Oil Resources		-0.002	0.131	-0.389	-0.339	0.085
		[0.172]	[0.379]	[0.392]	[0.395]	[0.394]
Colonial Links		0.984***	1.063**	0.890**	0.838*	1.019**
		[0.211]	[0.415]	[0.445]	[0.463]	[0.438]
Distance		-0.711***	-0.758***	-0.685***	-0.682***	-0.730***
		[0.051]	[0.096]	[0.095]	[0.096]	[0.096]
Environmental Stringency	0.872***	-0.376***	-0.319	-0.420**	-0.384*	-0.181
	[0.243]	[0.076]	[0.204]	[0.196]	[0.202]	[0.202]
Spatial Lag FDI				0.489**	0.459**	0.442**
				[0.193]	[0.199]	[0.191]
Market Potential					0.015	-0.085
					[0.186]	[0.192]
Spatial Stringency						0.787**
						[0.337]
Observations	5178	5178	5178	5178	5178	5178
Groups	341	341	341	341	341	341
R-squared	0.032	.	.	.	.	.
Instruments	.	.	144	147	150	153
Arrellano-Bond AR(2)	.	.	0.744	0.883	0.829	0.836
Sargan test	.	.	0.016	0.0922	0.0771	0.105
Hansen Test	.	.	0.444	0.49	0.462	0.423

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 4: Non-OECD Host Sample Results

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	-0.003	0.202***	0.088**	0.074*	0.075**	0.078**
	[0.019]	[0.017]	[0.037]	[0.038]	[0.038]	[0.040]
GDP Host	-0.006	0.781***	0.800***	0.581***	0.592***	0.465**
	[0.567]	[0.063]	[0.196]	[0.190]	[0.211]	[0.228]
GDP Source	2.085	0.806***	0.860***	0.448***	0.435***	0.387***
	[1.277]	[0.054]	[0.092]	[0.130]	[0.131]	[0.145]
GDP growth	0.117***	0.056	0.098	0.059	0.057	0.06
	[0.039]	[0.037]	[0.072]	[0.071]	[0.070]	[0.070]
V.A. Dirty Shares	-0.371	-0.162	-0.115	-0.498	-0.558	-0.659
	[0.300]	[0.156]	[0.348]	[0.346]	[0.383]	[0.417]
Risk Index	-0.829	-1.681***	-0.351	-0.593	-0.744	-1.141
	[0.648]	[0.490]	[1.005]	[0.962]	[0.883]	[0.901]
Capital Openness	0.153	0.261***	0.236*	0.163	0.159	0.127
	[0.117]	[0.062]	[0.127]	[0.124]	[0.131]	[0.133]
Oil Resources		-0.589***	-0.691*	-0.549	-0.582	-0.379
		[0.171]	[0.416]	[0.390]	[0.430]	[0.459]
Colonial Links		0.792***	1.013***	0.701**	0.695**	0.710**
		[0.202]	[0.337]	[0.294]	[0.293]	[0.301]
Distance		-0.486***	-0.421***	-0.225*	-0.230*	-0.157
		[0.079]	[0.130]	[0.129]	[0.131]	[0.178]
Environmental Stringency	-0.407	-0.342***	-0.734***	-0.559***	-0.536***	-0.455**
	[0.302]	[0.067]	[0.188]	[0.198]	[0.198]	[0.202]
Spatial Lag FDI				0.549***	0.561***	0.608***
				[0.138]	[0.135]	[0.148]
Market Potential					-0.021	0.023
					[0.215]	[0.230]
Spatial Stringency						-0.187
						[0.344]
Observations	3500	3500	3500	3500	3500	3500
Groups	358	358	358	358	358	358
R-squared	0.062	.	.	.	.	.
Instruments	.	.	145	148	151	154
Arrellano-Bond AR(2)	.	.	0.756	0.462	0.451	0.424
Sargan test	.	.	0.283	0.292	0.338	0.355
Hansen Test	.	.	0.544	0.746	0.767	0.753

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

A positive and significant spatial lag and an insignificant market potential lead us to conclude that complex FDI is the most prevailing type of FDI from OECD countries toward less developed host countries. The OECD sub-sample results indicate also a high vertical integrated complementary relationship in the allocation of FDI to other high income host countries. Note that these results differ from the findings of Blonigen et al. (2005). They found strong evidence of vertical FDI from the United States to non-OECD countries and export platform FDI to developed European countries. They also notice that the spatial lag is no longer significant once country-fixed effects are controlled, which is definitively not the case here. Our findings are in line with Garresten and Peeters (2009), who highlight the presence of complex FDI for Dutch outbound FDI to developed and developing host countries. In any case, one should be careful with comparisons, since I do not consider a single but several parent countries (i.e. OECD countries).

The estimated coefficients given in the last three columns of Table 4 measure only the pre-interdependent dynamic impact from changes in the explanatory variables. In particular, I am interested in investigating the long-run steady-state equilibrium effect of a counterfactual shock in environmental stringency in a given country on FDI inflows to this country and its neighbor economies. Table 5 illustrates the steady-state spatial effect of a permanent increase in the stringency of environmental regulation in China, Brazil and Hungary on FDI flows based on the system GMM estimates given in the last column from Table 4.

Table 5: Steady-State Spatial Effects of Environmental Stringency Shock on US bilateral FDI

Counterfactual Shock in China		Counterfactual Shock in Brazil		Counterfactual Shock in Hungary	
China	-.685***	Brazil	-.674***	Hungary	-.567***
South Korea	-.358***	Argentina	-.285***	Austria	-.112***
Japan	-.297***	Chile	-.244***	Romania	-.107***
Hong-Kong	-.173***	South Africa	-.155***	Poland	-.102***
Philippines	-.136***	Colombia	-.052***	Turkey	-.098***

Note: Standards errors based on the delta method. *** significant at 1%.

To save space, I only present the results for US FDI flows as they represent the largest share among OECD's FDI. I also only display the results for the subset of countries that are the mostly affected (in absolute value) by the counterfactual shock. A permanent 1% increase in environmental stringency in China, Brazil and Hungary leads to a permanent significant decrease in US FDI flows to these countries. However, because of the complementary relationship in FDI allocation, the countries within the same region are also negatively (but to a lesser extent) affected by a lasting and stringer environmental regulation.

5 Robustness Check

It is relevant to ask to what extent our results are sensitive to the use of SO_2 per capita as a proxy for environmental stringency[34]. Therefore, I re-estimate the model using CO_2 emission per capita and the number of international environmental treaties as complementary proxies for the environmental stringency. In addition, I consider the initial model specification and include additional FDI determinants to make sure that the level of SO_2 per capita emission does not capture other factors.

5.1 Alternative Environmental Stringency Proxies

As previously mentioned, the use of CO_2 per capita and international environmental treaties as a proxy relies on stronger assumptions than SO_2 per capita. Therefore, I expect to find less evidence of a pollution haven effect when considering these two additional proxies. Table 6 and 7 display the estimation results for the non-OECD host country sample[35]. Note that the same instruments and lags structure as in the main results are used to estimate the spatial dynamic panel model.

Most results found previously are also confirmed in Table 6 and 7. One major difference in Table 7 is the fact that the country risk index is now significant. In both tables, the results suggest the prevalence of complex FDI from OECD countries to less developed countries. As Table 6 documents it, the environmental stringency is negative and significant, only when I take into account the spatial structure of FDI inflows. This important finding suggests that ignoring spatial dependence can mask the pollution haven effect. The reason why this happens for CO_2 and not SO_2 emissions is probably related to the fact that carbon dioxide emissions are a global air pollutant, while sulfur dioxide emissions are more local/regional. The same kind of pattern

[34] I also estimated the model using SO_2 per GDP to account for the country's economic activity. The results were qualitatively similar (i.e. significant pollution haven effect), despite multicollinearity problems with host and source's GDP.

[35] Appendix 7.6 and 7.7 report the results for OECD countries.

happens with international treaties. The PHE becomes significant only when the spatial model includes the spatially weighted environmental stringency variable. More precisely, the allocation of FDI to a given host country is not only determined by the environmental regulation in the host economy, but also by the environmental stringency in the neighborhood countries. This finding is in line with the results of Davies and Naughton (2006). They show that the participation to international environmental treaties for non-OECD countries depends, among other factors, on the participation of geographically close governments. These results confirm once again the prevalence of complex vertical FDI among OECD home countries. In order to integrate the different intermediate production processes, the MNE considers a system of close countries rather than a single host economy.

The fact that there is less evidence of pollution haven effect using international environmental treaties should not come as a surprise. As mentioned previously, this last proxy relies on stronger assumptions, which are less likely to hold for less developed countries. In particular, the enforcement of the environmental treaties is weak because monitoring is lax or non-existent. Recent empirical evidence suggests that MNEs are more sensitive to the enforcement of environmental policy than the level of stringency itself (Kellenberg, 2009). This finding might even be more relevant, if I take into account the fact that decentralized local governments tend to set or implement lower environmental standards in order to attract mobile capital. This could partially explain why the correlation between SO_2 per capita emissions and the number of international treaties is not robust across the different subsamples. For instance, the correlation is positive for the OECD sample, but weakly negative for the non-OECD and full sample. The use of environmental treaties leads also to high collinearity with several variables (the capital openness index and country risk index among others), which could explain why environmental stringency is not significant. It is interesting to note that the capital openness and the number of ratified international environmental treaties are related, especially for small countries, in terms of diplomacy and willingness to comply with international standards.

Table 6: Non-OECD Host Sample Results with CO_2 per capita

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.004	0.193***	0.086**	0.071**	0.070**	0.071**
	[0.017]	[0.016]	[0.034]	[0.034]	[0.034]	[0.034]
GDP Host	1.016*	0.718***	0.810***	0.514**	0.413*	0.342
	[0.550]	[0.059]	[0.191]	[0.200]	[0.215]	[0.232]
GDP Source	1.222	0.789***	0.838***	0.361***	0.385***	0.434***
	[1.154]	[0.050]	[0.100]	[0.138]	[0.136]	[0.139]
GDP growth	0.126***	0.083**	0.08	0.057	0.05	0.045
	[0.037]	[0.035]	[0.075]	[0.076]	[0.074]	[0.073]
V.A. Dirty Shares	-0.121	-0.04	0.003	-0.453	-0.199	-0.117
	[0.279]	[0.144]	[0.351]	[0.356]	[0.392]	[0.475]
Risk Index	-1.442**	-2.013***	-2.026*	-1.221	-1.065	-1.466
	[0.613]	[0.472]	[1.053]	[1.032]	[0.976]	[0.976]
Capital Openness	0.209**	0.152***	0.105	0.067	0.02	-0.023
	[0.099]	[0.058]	[0.143]	[0.131]	[0.131]	[0.138]
Oil Resources		-0.514***	-0.697	-0.323	-0.14	-0.02
		[0.161]	[0.431]	[0.425]	[0.448]	[0.479]
Colonial Links		0.887***	0.968***	0.606**	0.616**	0.579*
		[0.189]	[0.287]	[0.301]	[0.300]	[0.336]
Distance		-0.492***	-0.544***	-0.215	-0.174	-0.148
		[0.075]	[0.134]	[0.142]	[0.139]	[0.238]
Environmental Stringency	0.038	-0.317***	-0.355	-0.417	-0.468*	-0.454*
	[0.390]	[0.069]	[0.242]	[0.256]	[0.253]	[0.267]
Spatial Lag FDI				0.636***	0.612***	0.560***
				[0.144]	[0.140]	[0.139]
Market Potential					0.231	0.253
					[0.223]	[0.255]
Spatial Stringency						-0.031
						[0.386]
Observations	4078	4078	4078	4078	4078	4078
Groups	374	374	374	374	374	374
R-squared	0.065	.	.	.	.	.
Instruments	.	.	145	148	151	154
Arrellano-Bond AR(2)	.	.	0.994	0.759	0.778	0.783
Sargan test	.	.	0.345	0.635	0.681	0.686
Hansen Test	.	.	0.232	0.37	0.46	0.499

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 7: Non-OECD Host Sample Results with Environmental Treaties

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.015	0.185***	0.078**	0.054*	0.051	0.054
	[0.016]	[0.015]	[0.031]	[0.032]	[0.033]	[0.033]
GDP Host	0.898*	0.732***	0.842***	0.567***	0.648***	0.678***
	[0.489]	[0.058]	[0.190]	[0.193]	[0.235]	[0.227]
GDP Source	1.992*	0.785***	0.880***	0.391***	0.383***	0.423***
	[1.057]	[0.048]	[0.094]	[0.133]	[0.133]	[0.128]
GDP growth	0.118***	0.071**	0.1	0.057	0.055	0.03
	[0.036]	[0.034]	[0.080]	[0.084]	[0.084]	[0.081]
V.A. Dirty Shares	-0.069	-0.01	-0.269	-0.495	-0.524	-0.536
	[0.258]	[0.140]	[0.352]	[0.364]	[0.394]	[0.376]
Risk Index	-2.065***	-2.604***	-1.738**	-1.340*	-1.483**	-1.488**
	[0.430]	[0.335]	[0.835]	[0.719]	[0.702]	[0.686]
Capital Openness	0.133	0.222***	0.123	0.058	0.099	0.18
	[0.091]	[0.059]	[0.151]	[0.138]	[0.143]	[0.138]
Oil Resources		-0.529***	-0.668	-0.246	-0.362	-0.245
		[0.159]	[0.425]	[0.405]	[0.448]	[0.456]
Colonial Links		0.920***	0.823***	0.719***	0.744***	0.714***
		[0.185]	[0.298]	[0.266]	[0.258]	[0.247]
Distance		-0.530***	-0.672***	-0.308**	-0.352**	-0.314**
		[0.070]	[0.130]	[0.140]	[0.158]	[0.155]
Environmental Stringency	-0.392	0.064	-0.201	-0.2	-0.232	-0.332*
	[0.684]	[0.070]	[0.179]	[0.153]	[0.177]	[0.175]
Spatial Lag FDI				0.648***	0.658***	0.613***
				[0.138]	[0.140]	[0.132]
Market Potential					-0.098	-0.086
					[0.264]	[0.249]
Spatial Stringency						3.113**
						[1.542]
Observations	4576	4576	4576	4576	4576	4576
Groups	374	374	374	374	374	374
R-squared	0.061	.	.	.	.	.
Instruments	.	.	157	160	163	166
Arrellano-Bond AR(2)	.	.	0.637	0.686	0.723	0.687
Sargan test	.	.	0.116	0.389	0.439	0.514
Hansen Test	.	.	0.217	0.251	0.225	0.3

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

5.2 Additional Control Variables

Since uncovering evidence of a pollution haven effect is subject to potential variables omission if spatial dependence is not addressed properly, I have to make sure that the use of the level of SO_2 per capita emission does not capture other FDI determinants. The next table reports the system-GMM estimations for the non-OECD sample by adding one by one additional control to the set of initial regressors[36]. This way I can detect any multicollinearity problem. Note that the inclusion of some explanatory variables decreases significantly the sample size of the panel. Therefore, one should be cautious when comparing the different results. Overall, the inclusion of the remaining additional FDI determinants leaves the results unchanged. Most of these additional explanatory variables have the expected sign but are not significantly different from zero. This is additional evidence that environmental stringency, proxied by SO_2 per capita emission does not encompass any other determinants.

More specifically, the first variable considered is GDP per capita to control for FDI allocation to higher income countries. However, the variable is not significant. I also include the number of telephone mainlines (per 1000 people) to measure the infrastructure level of the host economy. Its inclusion does not alter the results. In order to account for the importance of human capital, I include the proportion of secondary school enrollment. Since the measure is only available every 5 years, I interpolated it. Its estimated coefficient is not significantly different from zero. This might be related to the fact that I indirectly already account for the labour skill in the host economy by using the capital-labor ratio as a GMM instrument. In any case, the pollution haven effect remains negative and significant. This is also the case with the inclusion of a dummy variable for the presence of bilateral investment treaties and capital tax agreements between countries to account for the governments' marketing efforts to attract foreign investments. Openness to trade is considered in the literature as an important FDI determinant, that is why I consider several measures. The first proxy is a dummy variable for the existence of regional agreements trade which takes the value of 0 previous to the conclusion of the trade agreement and 1 afterwards. The ratio of exports to GDP as well as an index to assess the level of free trade of the economy and another one for the importance of trade tariff are also considered. In all these specifications, the associated coefficient is not significant. Although the

[36] While Appendix 7.2 reports the sources of the additional control variables, Appendix 7.8 displays the results for OECD countries.

Table 8: Non-OECD Host Sample Results with Additional FDI Determinants

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.086**	0.083**	0.096**	0.064*	0.080**	0.072*
	[0.038]	[0.037]	[0.043]	[0.037]	[0.039]	[0.039]
GDP Host	0.897***	0.939***	0.820***	0.931***	0.739***	0.750***
	[0.209]	[0.178]	[0.214]	[0.184]	[0.188]	[0.195]
GDP Source	0.874***	0.865***	1.021***	0.900***	0.822***	0.917***
	[0.092]	[0.092]	[0.105]	[0.099]	[0.096]	[0.105]
GDP growth	0.131*	0.129*	-0.006	0.086	0.122*	0.128*
	[0.070]	[0.070]	[0.103]	[0.078]	[0.074]	[0.076]
V.A. Dirty Shares	0.101	0.048	0.217	-0.142	-0.331	-0.258
	[0.334]	[0.337]	[0.349]	[0.309]	[0.296]	[0.355]
Risk Index	-1.355	-0.508	-1.21	-0.988	0.161	-0.457
	[0.930]	[0.977]	[1.187]	[1.039]	[1.106]	[1.044]
Capital Openness	0.127	0.118	0.143	0.234*	0.185	0.291**
	[0.152]	[0.148]	[0.142]	[0.121]	[0.138]	[0.140]
Oil Resources	-0.883**	-0.859**	-0.421	-0.861**	-0.562	-0.529
	[0.438]	[0.393]	[0.465]	[0.387]	[0.393]	[0.414]
Colonial Links	1.007***	1.106***	1.211***	1.042***	0.946**	0.865**
	[0.317]	[0.310]	[0.434]	[0.378]	[0.372]	[0.365]
Distance	-0.463***	-0.446***	-0.526***	-0.435***	-0.463***	-0.525***
	[0.131]	[0.123]	[0.144]	[0.145]	[0.140]	[0.166]
Environmental Stringency	-0.615***	-0.451**	-0.767***	-0.563***	-0.657***	-0.708***
	[0.225]	[0.217]	[0.186]	[0.183]	[0.178]	[0.209]
GDP per capita Host	0.099					
	[0.239]					
Phone		0.291				
		[0.221]				
School Enrollment			-0.068			
			[0.657]			
Bilateral Invest. Treaties				-0.07		
				[0.278]		
Capital Tax Agreements					-0.582*	
					[0.301]	
Regional Trade Agreement						-0.243
						[0.519]
Observations	3500	3500	2631	3500	3500	3500
Groups	358	358	347	358	358	358
R-squared	148	148	121	233	227	199
Instruments	0.796	0.795	0.251	0.973	0.827	0.904
Arrellano-Bond AR(2)	0.162	0.41	7.03E-05	0.125	0.055	0.0861
Sargan test	0.437	0.496	0.336	0.708	0.614	0.767
Hansen Test	0.086**	0.083**	0.096**	0.064*	0.080**	0.072*

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 8: Non-OECD Host Sample Results (*continued*)

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.107***	0.081**	0.090**	0.034	0.090**	0.087**
	[0.036]	[0.037]	[0.037]	[0.051]	[0.037]	[0.037]
GDP Host	0.772***	0.817***	0.809***	0.544**	0.791***	0.782***
	[0.193]	[0.185]	[0.194]	[0.222]	[0.198]	[0.202]
GDP Source	0.867***	0.869***	0.873***	1.019***	0.839***	0.862***
	[0.099]	[0.092]	[0.089]	[0.105]	[0.093]	[0.092]
GDP growth	0.121*	0.086	0.081	0.094	0.1	0.096
	[0.073]	[0.073]	[0.071]	[0.107]	[0.072]	[0.072]
V.A. Dirty Shares	0.263	0.164	0.045	0.095	-0.157	-0.157
	[0.360]	[0.371]	[0.361]	[0.373]	[0.351]	[0.352]
Risk Index	0.072	0.053	-0.437	-0.498	-0.32	-0.35
	[1.021]	[1.005]	[0.995]	[1.493]	[1.010]	[1.006]
Capital Openness	0.161	0.143	0.206	0.102	0.219*	0.242*
	[0.130]	[0.145]	[0.145]	[0.138]	[0.128]	[0.127]
Oil Resources	-0.59	-0.670*	-0.717*	-0.029	-0.744*	-0.663
	[0.390]	[0.402]	[0.430]	[0.540]	[0.423]	[0.423]
Colonial Links	1.040***	1.044***	1.022***	1.022**	0.941***	1.288***
	[0.349]	[0.338]	[0.329]	[0.489]	[0.335]	[0.409]
Distance	-0.404***	-0.467***	-0.462***	-0.423***	-0.348**	-0.409***
	[0.130]	[0.132]	[0.140]	[0.141]	[0.137]	[0.130]
Environmental Stringency	-0.789***	-0.673***	-0.662***	-0.729***	-0.702***	-0.744***
	[0.188]	[0.184]	[0.211]	[0.194]	[0.186]	[0.190]
Openness	0.035					
	[0.235]					
Free Trade Index		1.051				
		[0.750]				
Tariff Index			0.349			
			[0.549]			
Corruption Index				-0.47		
				[0.295]		
Contiguity					1.190**	
					[0.541]	
Common Language						-0.396
						[0.445]
Observations	3312	3481	3480	2422	3500	3500
Groups	343	357	357	349	358	358
Instruments	148	148	148	97	146	146
Arrellano-Bond AR(2)	0.555	0.82	0.739	0.798	0.74	0.759
Sargan test	0.0904	0.314	0.227	0.0217	0.274	0.284
Hansen Test	0.478	0.5	0.529	0.607	0.549	0.551
Lagged FDI	0.107***	0.081**	0.090**	0.034	0.090**	0.087**

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

baseline model already accounts for the level of corruption, through a country's (financial, economic and political) risk index, I include an index of corruption to capture the level of enforcement of environmental stringency. Although the inclusion of this index decreases the sample size significantly, results are unchanged except for the lagged dependent variable, which is no longer significant. Finally, the inclusion of a contiguity and common language dummy further confirms that the evidence of a pollution haven effect is robust to the inclusion of additional control variables.

6 Conclusion

This paper analyzes the effect of environmental stringency on FDI inflows in a multi-country setting. In addition to traditional determinants of FDI, the model included spatially weighted variables in order to account for third-country effects. The estimations is carried out on a sample of bilateral FDI from OECD countries to a large number of host countries over the period 1981-2000.

I show the importance of distinguishing between inflows of FDI to high income versus low income countries and correcting for the endogeneity of the environmental regulation using system GMM. In fact, most specifications yield a significant negative relationship between environmental stringency and inflows of FDI, once endogeneity and spatial dependence are taken into account. This finding is robust across specifications but also using different environmental stringency proxies. In particular, depending on the global nature of the proxy, failing to account for spatial dependence can mask the pollution haven effect. Furthermore, I am able to highlight the prevalence of complex FDI between OECD countries and lower income countries. In other words, MNEs tend to allocate each main production step to different countries in order to minimize the production process costs.

The policy implications are not necessarily straightforward. Environmental stringency can be used as an instrument to attract manufacturing multinationals. However, the host country should be aware that they will mainly attract intermediate goods production through vertical integrated complex FDI. In other words, they will only be part of a small part of the production process, the most polluting one. In terms of economic development, this is not necessarily the best way to ensure long term sustainable economic growth.

Although the findings are largely plausible and robust across specifications, they should be taken with cautious. The proxies for environmental stringency are unfortunately still imperfect and scarce. The use of a more reliable proxy would further reduce the bias associated with the potential omission of variables. Another interesting extension of this study would be to estimate the specification using bilateral data at the industry level in order to control more accurately for the polluting intensive composition effect. Unfortunately, disaggregated bilateral FDI data are scarce and only available for a limited number of countries and years. These are the main challenges the study of the linkages between FDI and pollution haven at the world-wide level faces. In any case, an inter-country analysis remains relevant to get the big picture in terms of the spatial allocation of FDI and can be of particular interest to policymakers in developing countries who compete to attract FDI.

7 Appendices

7.1 Country Lists

Host Countries:

Algeria; Argentina; Bangladesh; Benin; Bolivia; Brazil; Bulgaria; Cameroon; Central African Republic; Chad; Chile; China; Colombia; Costa Rica; Côte d'Ivoire; Czech Republic; Ecuador; Egypt; El Salvador; Ethiopia; Finland; France; Gabon; Germany; Ghana; Greece; Guatemala; Honduras; Hong Kong; Hungary; India; Indonesia; Iran; Israel; Jordan; Kenya; Korea Rep.; Kuwait; Latvia; Malaysia; Mauritius; Mexico; Morocco; Nigeria; Oman; Pakistan; Panama; Peru; Philippines; Romania; Russian Federation; Senegal; Singapore; Slovak Republic; Slovenia; South Africa; Sri Lanka; Tanzania; Thailand; Trinidad and Tobago; Tunisia; Uruguay; Venezuela.

Source/Host Countries (OECD countries):

Australia; Austria; Belgium; Canada; Denmark; Finland; France; Germany; Greece; Ireland; Italy; Japan; Korea, Rep.; Mexico; Netherlands; New Zealand; Norway; Poland; Portugal; Spain; Sweden; Switzerland; Turkey; United; Kingdom; United States.

High Income OECD countries:

Australia; Austria; Belgium; Canada; Denmark; Finland; France; Germany; Greece; Ireland; Italy; Japan; Korea, Rep.; Netherlands; New Zealand; Norway; Portugal; Spain; Sweden; Switzerland; United Kingdom; United States.

7.2 Data Description

FDI Determinant	Theory/Hypothesis	Proxy Variable (Source)	Effect
Market demand	Market size hypothesis	GDP (World Development Indicator)	+
Growth rate	Differential rates of return	Real GDP Growth (World Development Indicator)	+
Agglomeration effect	Other	Lagged FDI (OECD Database)	+
Industrial Structure	Other	Share of Manufacturing in total GDP (World Development Indicator)	+/-
Factor Endowments	Factor endowments hypothesis	Capital/ Labour ratio (World Penn Tables)	+
Trade Barriers	Tariff jumping hypothesis	Dummy for Regional Trade Agreement (WorldTradeLaw.net)	+/-
Capital Openness	Other	Ito-Chinn index	+
Investment Promotion	Other	Dummy for Bilateral Investment Treaties (UNCTAD)	+
Country Risk	Other	ICRG index (World Development Indicator)	-
Natural Resources	Other	Dummy for largest oil rich countries (www.eia.doe.gov)	-
Proximity	Gravity hypothesis	Capitals distance Contiguity, Colonies, Common Language (CEPII)	- +
"Third-Country" Effect	Spillovers hypothesis	Spatially weighted variables (Negative exponential distance matrix)	+/-
Environmental Stringency	Pollution haven hypothesis	SO_2 emission per capita (Stern (2005)) International environmental treaties (ENTRI)	- -

7.3 Descriptive Statistics

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
FDI	13555	3.29	4.06	-10.97	12.75
Lagged FDI	12211	3.39	4.02	-10.97	12.75
Spatial Lag	13555	3.14	2.29	-7.93	11.4
GDP	13415	26.41	1.47	21.16	30
Market Potential	13415	25.22	0.9	21.94	27.79
GDP Growth	13487	2.2	4.29	-50.49	89.83
V.A. Dirty Shares	12074	-1.21	0.32	-5.59	-0.1

Notes: All variables are expressed in logarithm, except for the dummies variables.

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
Manufacturing Share (%GDP)	10791	2.94	0.36	-0.54	3.77
Risk Index	13386	-4.29	0.2	-4.62	-2.67
Capital Openness	12723	0.64	1.13	-1.33	1.68
Bilateral Investment Treaties	13555	0.26	0.44	0	1
Regional Agreement Trades	13555	0.38	0.49	0	1
Capital Tax Agreements	13555	0.42	0.49	0	1
Oil Resources	13555	0.2	0.4	0	1
Colonial Links	13555	0.08	0.27	0	1
Distance	13555	8.17	1.11	4.09	9.88
Environmental Stringency (SO_2 pc)	11299	11.36	1.03	5.68	15.53
Spatial Stringency (SO_2 pc)	11299	11.53	0.58	9.55	14.18
Environmental Stringency (Treaties)	13549	4.4	0.81	0	5.82
Spatial Stringency (Treaties)	13549	4.04	0.48	2.16	4.88
Phone	13474	5.26	1.32	-0.16	6.64
School Enrolment	11168	4.4	0.42	1.61	5.18
Openness	13106	3.36	0.6	1.19	5.26
Exports	13050	24.65	1.53	18.22	27.77
Free Trade Index	13182	1.96	0.2	0.43	2.28
Tariff Index	13110	1.42	0.24	-2.07	1.63
Corruption Index	9282	-2.07	0.19	-2.29	-1.13

Notes: All variables are expressed in logarithm, except for the dummies variables.

7.4 Correlation Matrix

	FDI	Lagged FDI	Spatial Lag	GDP	Market Potential	GDP Growth	V.A. Dirty	Manuf. Share	Risk Index
FDI	1								
Lagged FDI	0.28	1							
Spatial Lag	0.25	0.23	1						
GDP	0.16	0.18	-0.02	1					
Market Potential	0.04	0.04	0.16	0.24	1				
GDP Growth	0.03	0.02	0.07	-0.1	0.11	1			
V.A. Dirty Shares	-0.01	-0.01	0.07	-0.07	0	0.01	1		
Manuf. Share	0.06	0.07	0.11	0.15	0.33	0.16	0.05	1	
Risk Index	-0.07	-0.07	-0.09	-0.2	-0.43	-0.26	0	-0.1	1
Capital Openness	0.07	0.08	0.06	0.27	0.3	0.03	0.05	-0.02	-0.64
BIT	-0.06	-0.05	-0.06	-0.2	-0.09	0	-0.02	-0.01	0.42
RAT	0.05	0.04	0	-0.13	0.18	0.04	-0.05	0.09	-0.29
KTA	0.05	0.04	0.01	-0.07	0.13	-0.01	-0.03	0.05	-0.21
Oil Resources	0.04	0.04	0.05	0.34	-0.09	-0.06	0.13	-0.2	0.01
Colonial Links	0.1	0.09	0.07	0.03	-0.05	-0.02	0.05	0	0.03
Distance	-0.12	-0.11	-0.07	0.12	-0.25	-0.03	0.01	-0.08	0.27

Notes: All variables are expressed in logarithm, except for the dummies variables.

	FDI	Lagged FDI	Spatial Lag	GDP	Market Potential	GDP Growth	V.A. Dirty	Manuf. Share	Risk Index
-(SO_2 pc)	-0.05	-0.05	-0.01	-0.1	0.12	-0.08	-0.06	0.05	0.07
Spatial -(SO_2 pc)	-0.02	-0.02	-0.08	-0.1	-0.18	0.08	-0.15	0.01	0.23
Treaties	0.12	0.11	0.05	0.29	0.17	-0.01	0.03	0.17	-0.43
Spatial Treaties	0.08	0.09	0.09	-0.08	0.19	0.08	-0.08	-0.03	-0.32
Phone	0.1	0.11	0.08	0.3	0.4	0.01	-0.03	0	-0.71
School Enrolment	0.11	0.11	0.07	0.25	0.36	0.06	-0.06	0.04	-0.66
Openness	-0.03	-0.04	0.1	-0.57	0.21	0.16	0	0.08	-0.16
Exports	0.17	0.19	0.04	0.85	0.46	-0.02	-0.06	0.17	-0.53
Free Trade Index	0.12	0.12	0.09	0.04	0.28	0.12	0.01	0.06	-0.57
Tariff Index	0.1	0.11	0.05	0.21	0.25	-0.03	-0.02	-0.02	-0.45
Corruption Index	-0.06	-0.07	-0.05	-0.19	-0.35	-0.07	0.17	-0.01	0.67

Notes: All variables are expressed in logarithm, except for the dummies variables.

	Capital Open.	BIT	RAT	KTA	Oil Ress.	Col. Links	Distance	-(SO_2 pc)	Spatial -(SO_2 pc)
Capital Openness	1								
BIT	-0.36	1							
RAT	0.23	-0.12	1						
KTA	0.2	-0.02	0.5	1					
Oil Resources	-0.02	-0.13	-0.24	-0.05	1				
Colonial Links	-0.03	-0.02	-0.12	-0.08	0.03	1			
Distance	-0.18	0.11	-0.83	-0.54	0.19	0.03	1		
-(SO_2 pc)	0.11	0.03	0.09	0.18	-0.21	-0.04	-0.15	1	
Spatial -(SO_2 pc)	-0.31	0.18	-0.23	-0.17	-0.14	0.02	0.26	0.11	1
Treaties	0.37	-0.35	0.44	0.3	0	-0.02	-0.39	-0.02	-0.37
Spatial Treaties	0.25	-0.11	0.51	0.32	-0.17	-0.05	-0.47	-0.01	-0.18
Phone	0.56	-0.33	0.38	0.21	-0.01	-0.06	-0.32	-0.24	-0.41
School Enrolment	0.53	-0.28	0.37	0.21	-0.09	-0.06	-0.27	-0.2	-0.31
Openness	0.1	0.09	0.33	0.17	-0.28	-0.04	-0.32	0.07	0.11
Exports	0.54	-0.31	0.1	0.08	0.23	0	-0.1	-0.05	-0.16
Free Trade Index	0.62	-0.21	0.42	0.23	-0.17	-0.04	-0.33	-0.13	-0.17
Tariff Index	0.43	-0.24	0.34	0.15	-0.07	-0.08	-0.22	-0.13	-0.22
Corruption Index	-0.44	0.28	-0.25	-0.16	0.01	0.02	0.2	0.14	0.07

Notes: All variables are expressed in logarithm, except for the dummies variables.

	Treaties	Spatial Treaties	Phone	School Enrol.	Open.	Exports	F. T, Index	Tariff Index	Corr. Index
Treaties	1								
Spatial Treaties	0.57	1							
Phone	0.49	0.47	1						
School Enrolment	0.54	0.48	0.83	1					
Openness	-0.13	0.23	0.08	0.09	1				
Exports	0.37	0.12	0.59	0.5	-0.14	1			
Free Trade Index	0.3	0.45	0.65	0.6	0.52	0.45	1		
Tariff Index	0.35	0.37	0.71	0.62	0.09	0.44	0.64	1	
Corruption Index	-0.33	-0.34	-0.62	-0.54	-0.14	-0.46	-0.54	-0.31	1

Notes: All variables are expressed in logarithm, except for the dummies variables.

7.5 Spatial Dependence Tests

	Bilateral FDI		GDP		SO_2 per capita		Treaties	
Year	**Moran *I***	**Geary *C***	**Moran *I***	**Geary *C***	**Moran *I***	**Geary *C***	**Moran *I***	**Geary *C***
1981	0.12***	0.89***	0.23***	0.74***	0.08*	0.91*	0.39***	0***
1982	0.07***	0.94**	0.12**	0.86**	0.14***	0.84***	0.36***	0***
1983	0.11***	0.89***	0.1**	0.85**	0.11**	0.91*	0.33***	0***
1984	0.1***	0.93***	0.2***	0.77***	0.21***	0.79***	0.34***	0***
1985	0.19***	0.81***	0.25***	0.72***	0.26***	0.73***	0.33***	0***
1986	0.11***	0.91***	0.31***	0.67***	0.24***	0.74***	0.31***	0***
1987	0.13***	0.88***	0.16***	0.78***	0.22***	0.75***	0.33***	0***
1988	0.15***	0.86***	0.26***	0.71***	0.23***	0.77***	0.29***	0***
1989	0.2***	0.8***	0.32***	0.67***	0.31***	0.7***	0.32***	0***
1990	0.19***	0.84***	0.31***	0.67***	0.29***	0.72***	0.31***	0***
1991	0.14***	0.87***	0.35***	0.63***	0.25***	0.76***	0.18***	0***
1992	0.14***	0.89***	0.34***	0.65***	0.27***	0.75***	0.07**	0.37
1993	0.12***	0.89***	0.3***	0.69***	0.25***	0.76***	0.03	0.21
1994	0.24***	0.76***	0.21***	0.75***	0.2***	0.8***	0.04*	0.15
1995	0.15***	0.84***	0.25***	0.71***	0.21***	0.79***	0.09***	0.2
1996	0.17***	0.84***	0.28***	0.68***	0.27***	0.73***	0.13***	0***
1997	0.18***	0.82***	0.31***	0.66***	0.16***	0.84***	0.15***	0***
1998	0.09***	0.91***	0.28***	0.68***	0.18***	0.82***	0.16***	0***
1999	0.12***	0.89***	0.26***	0.69***	0.18***	0.81***	0.18***	0***
2000	0.15***	0.86***	0.26***	0.69***	0.16***	0.84***	0.17***	0***
2001	0.14***	0.86***	0.28***	0.68***	0.33***	0.66***	0.18***	0***
2002	0.14***	0.86***	0.33***	0.65***	0.2**	0.83*	0.22***	0***

Notes: Both tests performed in logarithm with an inverse exponential distance matrix. significant at 10, ** significant at 5, *** significant at 1.

7.6 OECD Robustness Check with CO_2 per capita

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.012	0.190***	0.058**	0.051**	0.055**	0.063**
	[0.015]	[0.014]	[0.025]	[0.026]	[0.026]	[0.026]
GDP Host	0.883	0.534***	0.518***	0.583***	0.550***	0.593***
	[0.851]	[0.046]	[0.113]	[0.106]	[0.117]	[0.115]
GDP Source	5.781***	0.802***	0.947***	0.655***	0.624***	0.588***
	[1.071]	[0.042]	[0.077]	[0.144]	[0.146]	[0.150]
GDP growth	-0.056	0.064	-0.047	-0.074	-0.053	-0.065
	[0.062]	[0.059]	[0.087]	[0.091]	[0.092]	[0.093]
V.A. Dirty Shares	-0.11	-0.053	-0.098	-0.524	-0.676	-0.495
	[0.324]	[0.263]	[0.431]	[0.490]	[0.495]	[0.493]
Risk Index	-2.433	1.261	-2.824	-1.701	-1.199	-1.325
	[1.600]	[1.063]	[2.516]	[2.648]	[2.621]	[2.594]
Capital Openness	-0.087	0.095	0.173	0.114	0.167	0.107
	[0.143]	[0.111]	[0.213]	[0.210]	[0.210]	[0.200]
Oil Resources	0	0.028	0.007	-0.579	-0.611	-0.539
	[0.000]	[0.190]	[0.450]	[0.453]	[0.463]	[0.465]
Colonial Links	0	1.035***	1.120***	0.937**	0.851*	0.878*
	[0.000]	[0.211]	[0.416]	[0.458]	[0.477]	[0.456]
Distance	0	-0.700***	-0.733***	-0.687***	-0.686***	-0.682***
	[0.000]	[0.051]	[0.092]	[0.094]	[0.096]	[0.103]
Environmental Stringency	1.123	-0.602***	-0.558	-0.874*	-0.914*	-0.79
	[0.722]	[0.201]	[0.520]	[0.516]	[0.538]	[0.563]
Spatial Lag FDI				0.444**	0.468**	0.493**
				[0.190]	[0.192]	[0.196]
Market Potential					0.055	-0.035
					[0.192]	[0.225]
Spatial Stringency						0.062
						[0.459]
Observations	5194	5194	5194	5194	5194	5194
Groups	341	341	341	341	341	341
R-squared	0.03	.	144	147	150	153
Instruments	.	.	144	147	150	153
Arrellano-Bond AR(2)	.	.	0.819	0.923	0.884	0.783
Sargan test	.	.	0.0138	0.0862	0.0657	0.0239
Hansen Test	.	.	0.409	0.467	0.455	0.356

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

7.7 OECD Robustness Check with Treaties

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.037***	0.195***	0.060**	0.060**	0.059**	0.056**
	[0.014]	[0.013]	[0.029]	[0.028]	[0.028]	[0.028]
GDP Host	-0.295	0.544***	0.638***	0.635***	0.654***	0.666***
	[0.740]	[0.047]	[0.124]	[0.120]	[0.122]	[0.124]
GDP Source	7.798***	0.820***	0.937***	0.615***	0.614***	0.591***
	[0.996]	[0.041]	[0.084]	[0.137]	[0.136]	[0.132]
GDP growth	-0.012	0.097	0.058	0.009	0.007	-0.021
	[0.062]	[0.060]	[0.093]	[0.096]	[0.094]	[0.093]
V.A. Dirty Shares	0.135	-0.09	-0.454	-0.680*	-0.535	-0.609
	[0.312]	[0.246]	[0.380]	[0.407]	[0.410]	[0.405]
Risk Index	-1.5	1.239	0.83	1.281	-0.147	-0.67
	[1.534]	[1.029]	[2.727]	[2.698]	[2.614]	[2.518]
Capital Openness	-0.048	0.184*	0.611***	0.491**	0.369*	0.347
	[0.143]	[0.108]	[0.219]	[0.210]	[0.217]	[0.216]
Oil Resources	0	0.373**	0.301	-0.092	-0.171	-0.01
	[0.000]	[0.155]	[0.325]	[0.319]	[0.304]	[0.325]
Colonial Links	0	0.981***	0.975**	0.992**	0.884*	0.968**
	[0.000]	[0.209]	[0.447]	[0.481]	[0.489]	[0.481]
Distance	0	-0.622***	-0.887***	-0.742***	-0.749***	-0.675***
	[0.000]	[0.057]	[0.140]	[0.146]	[0.144]	[0.145]
Environmental Stringency	0.061	0.177	-1.191*	-0.742	-0.764	-1.263
	[1.921]	[0.229]	[0.712]	[0.709]	[0.713]	[0.827]
Spatial Lag FDI				0.450***	0.449***	0.488***
				[0.168]	[0.165]	[0.164]
Market Potential					-0.101	-0.147
					[0.186]	[0.181]
Spatial Stringency						3.66
						[2.679]
Observations	5728	5728	5728	5728	5728	5728
Groups	341	341	341	341	341	341
R-squared	0.033	.	.	.	.	.
Instruments	.	.	155	158	161	164
Arrellano-Bond AR(2)	.	.	0.216	0.311	0.317	0.354
Sargan test	.	.	0.225	0.36	0.403	0.467
Hansen Test	.	.	0.311	0.39	0.376	0.424

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

7.8 OECD Robustness Check with Additional FDI Determinants

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.055**	0.062**	0.054*	0.057**	0.047*	0.055**
	[0.026]	[0.026]	[0.029]	[0.025]	[0.025]	[0.028]
GDP Host	0.679***	0.607***	0.718***	0.566***	0.541***	0.530***
	[0.132]	[0.122]	[0.132]	[0.115]	[0.117]	[0.142]
GDP Source	0.951***	0.950***	1.045***	0.958***	0.985***	0.924***
	[0.071]	[0.072]	[0.079]	[0.073]	[0.075]	[0.109]
GDP growth	-0.093	-0.106	-0.151	-0.055	-0.107	-0.071
	[0.092]	[0.095]	[0.110]	[0.091]	[0.082]	[0.083]
V.A. Dirty Shares	0.059	0.031	0.263	-0.201	0.469	0.609
	[0.460]	[0.462]	[0.477]	[0.449]	[0.415]	[0.457]
Risk Index	-2.974	-1.983	0.727	-2.051	-5.120**	-3.848
	[2.580]	[2.519]	[3.603]	[2.413]	[2.277]	[2.363]
Capital Openness	0.256	0.162	0.194	0.377	0.044	-0.088
	[0.234]	[0.235]	[0.230]	[0.238]	[0.211]	[0.217]
Oil Resources	0.239	0.135	-0.136	0.214	0.138	-0.224
	[0.404]	[0.407]	[0.423]	[0.384]	[0.398]	[0.458]
Colonial Links	1.093***	1.128***	0.842*	1.105***	0.980**	1.050*
	[0.424]	[0.417]	[0.472]	[0.407]	[0.443]	[0.538]
Distance	-0.807***	-0.769***	-0.960***	-0.792***	-0.928***	-1.120**
	[0.104]	[0.101]	[0.115]	[0.101]	[0.154]	[0.446]
Environmental Stringency	-0.301	-0.325	-0.668***	-0.282	-0.388*	-0.329
	[0.205]	[0.209]	[0.227]	[0.202]	[0.198]	[0.206]
GDP per capita Host	-0.468					
	[0.550]					
Phone		-0.199				
		[0.618]				
School Enrollment			-0.854			
			[0.905]			
Bilateral Investment Treaties				1.016		
				[1.282]		
Capital Tax Agreements					-0.74	
					[0.569]	
Regional Agreement Trades						-1.131
						[1.321]
Observations	5178	5178	4253	5178	5178	5178
Groups	341	341	340	341	341	341
Instruments	147	147	121	186	218	203
Arrellano-Bond AR(2)	0.837	0.753	0.979	0.813	0.97	0.841
Sargan test	0.0212	0.0162	0.03	0.191	0.118	0.0856
Hansen Test	0.369	0.383	0.163	0.91	0.428	0.0509
Lagged FDI	0.055**	0.062**	0.054*	0.057**	0.047*	0.055**

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

	Fixed	Random	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
Lagged FDI	0.064**	0.060**	0.057**	0.048*	0.064**	0.064**
	[0.025]	[0.026]	[0.025]	[0.027]	[0.025]	[0.025]
GDP Host	0.811***	0.677***	0.663***	0.587***	0.591***	0.619***
	[0.160]	[0.132]	[0.121]	[0.140]	[0.115]	[0.116]
GDP Source	0.934***	0.933***	0.963***	0.971***	0.936***	0.937***
	[0.070]	[0.072]	[0.072]	[0.077]	[0.070]	[0.071]
GDP growth	-0.08	-0.099	-0.037	-0.265**	-0.066	-0.066
	[0.088]	[0.091]	[0.091]	[0.107]	[0.092]	[0.092]
V.A. Dirty Shares	-0.049	0.128	-0.212	0.032	0.056	0.041
	[0.452]	[0.461]	[0.444]	[0.468]	[0.462]	[0.452]
Risk Index	-0.128	1.497	-1.011	-1.242	-1.799	-1.743
	[2.393]	[2.350]	[2.408]	[4.020]	[2.573]	[2.579]
Capital Openness	0.349	0.015	0.278	-0.583*	0.228	0.205
	[0.220]	[0.241]	[0.213]	[0.338]	[0.227]	[0.230]
Oil Resources	0.144	0.125	0.005	0.165	0.117	0.057
	[0.395]	[0.387]	[0.398]	[0.421]	[0.375]	[0.385]
Colonial Links	1.081**	1.100***	1.114**	1.257***	1.034**	0.653
	[0.427]	[0.399]	[0.450]	[0.471]	[0.407]	[0.464]
Distance	-0.658***	-0.734***	-0.861***	-0.950***	-0.687***	-0.731***
	[0.105]	[0.104]	[0.110]	[0.120]	[0.101]	[0.096]
Environmental Stringency	-0.356*	-0.385*	-0.254	-0.430*	-0.303	-0.295
	[0.201]	[0.215]	[0.201]	[0.238]	[0.204]	[0.203]
Openness	1.019**					
	[0.489]					
Free Trade Index		4.881***				
		[1.894]				
Tariff Index			-3.685*			
			[2.045]			
Corruption Index				-0.488		
				[0.771]		
Contiguity					0.495	
					[0.327]	
Common Language						0.799**
						[0.369]
Observations	5178	5161	5161	3420	5178	5178
Groups	341	341	341	339	341	341
Instruments	147	147	147	97	145	145
Arrellano-Bond AR(2)	0.725	0.768	0.807	0.594	0.733	0.735
Sargan test	0.0727	0.0753	0.0201	0.078	0.0202	0.0231
Hansen Test	0.478	0.515	0.452	0.647	0.45	0.426
Lagged FDI	0.064**	0.060**	0.057**	0.048*	0.064**	0.064**

Notes: Exponential inverse distance as spatial weighting scheme. Two-Step System-GMM estimations. Time dummies and constant not reported. Standard errors incorporating the Windmeijer correction in brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Chapter 4

Unequal Spatial Diffusion of Eco-label Programs

1 Introduction

While most industrialized countries have adopted an eco-labelling scheme, African and Latin American countries have yet to decide to implement this type of market-based initiative. This asymmetry has led to trade tensions, as eco-labels are criticized for potentially imposing the environmental concerns of (high income) importing countries on the production methods of (low income) trading partners (Bonsi et al., 2008). This is particularly problematic for developing countries depending heavily on exports to sustain their growth. So far, in spite of these important concerns, most of the literature on eco-labels and its international trade linkages has taken a conceptual or descriptive approach due to lack of data.

This study brings new empirical evidence on the determinants of multi-sector eco-labelling programs. It is based on a large sample, including 141 developing and developed economies and is the first one to examine the extent to which the eco-label decision is dependent upon the behavior adopted by nearby partner countries. A government, which faces pressure on its national and export markets, might consider introducing an eco-labelling scheme in order to restrict market access to foreign products, or to boost its exports earnings. Whatever the case, its decision will depend on how many and which countries have or are expected to adopt an eco-labelling program. Such interdependence may help explain the very unequal diffusion of eco-label programs among countries. As prior studies (Grolleau & El Harbi, 2008) have not controlled for the spatial dependence in the eco-label decision, their results may be biased or inconsistent. In short, the empirical approach in this paper aims at providing an answer to two main questions. First, why some governments have adopted eco-label programs and others did not. Second, do strategic interactions among partners matter in their decision to adopt an eco-labelling scheme?

To tackle the issue of cross-sectional interdependence, I estimate a Bayesian spatial autoregressive probit model by applying a new Markov-Chain-Monte-Carlo algorithm. The standard iterative sampling method proposed by LeSage (2001) suffers from slow mixing and convergence when there is a strong posterior correlation between the parameters and the latent variable. In order to avoid relying on a excessively large number of draws, which can be time consuming, I follow Holmes and Held (2006) and modify the algorithm to update jointly the non spatial regression parameters and the latent variable. To my knowledge, this is the first time that this type of joint sampling is done in a spatial Bayesian framework. A small simulation study confirms the superiority of this new algorithm over the standard one in terms of mixing and convergence.

The empirical evidence confirms the role of the economy's stage of development and innovation capacity in the government's decision to introduce an eco-label. Moreover, results highlight the strategic nature of the eco-labelling decision. In particular, the probability for a country to introduce an eco-labelling scheme depends on the eco-label programs adopted by countries which are spatially close or sharing a strong trade intensity relationship with each other. These results explain why developing countries, which are standards takers, have naturally been put at a disadvantage in terms of eco-labelling adoption.

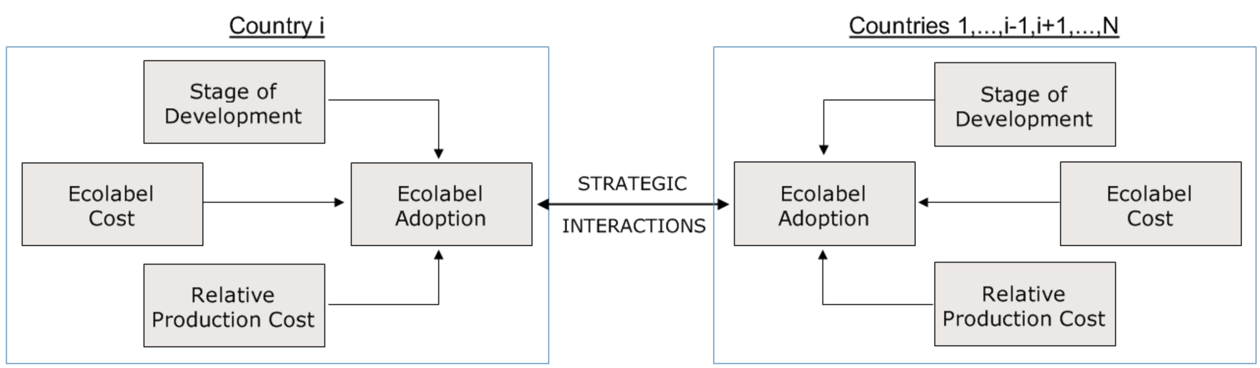
The remainder of the paper is structured as follows. Section 2 reviews the main determinants associated with the decision to introduce an eco-labelling scheme. Section 3 describes the spatial probit model. The joint sampling estimation method and its performance based on a small Monte-Carlo simulation study are also presented. The empirical results are discussed in section 4. Section 5 checks the robustness of the findings. Finally, section 6 concludes.

2 Determinants of Eco-Label Adoption

"Eco-labels" is the short form for ecological labels. They contain information regarding potential impacts on the environment of the production, consumption and waste phases of the products/services consumed. However, eco-labelling schemes are not just messages about a product or a service but claims stating that it has particular properties or features (Mason, 2006). In fact, even the instrument of labelling itself is a claim, as it refers to certain characteristics of the procedure under which the label is awarded. The very own existence of eco-labels arises when firms have information on the environmental impact of the product that consumers value but cannot check. Buyers are unable to verify the environmental consequence of the goods before the purchase or through frequent purchase. This type of information asymmetry, also known as credence feature, is due to the temporal, spatial and non-exclusion characteristics of most environmental impacts. In addition, the environmental degradation (or improvement) generated by the environmental characteristics of the conventional (or eco-friendly) product displays properties of public goods showing either non-excludability or non-rivalry in consumption. This usually implies a free riding problem as well as an assurance problem. Moreover, no market prices prevail to really reflect the value of the production and consumption externalities of the product. These public good features lead to a misallocation of scarce resources because the decision-making process does not take into account all the costs. Therefore, most eco-label programs seek to fulfill two objectives:

- provide consumers with more information about the environmental effects of their consumption, (i.e. transform the product's credence attribute into a search attribute), which should generate a move towards more environmentally friendly consumption patterns;
- encourage economic agents (mainly firms and governments) to increase the environmental standards of products/services by benchmarking environmental performance (i.e. internalize the non-market benefits of the eco-labelled good).

Figure 1: Eco-label Determinants



Several authors have defined the potential determinants that could explain why countries and producers would adopt an eco-label (Grolleau & El Harbi, 2008). Among them, Basu et al. (2004) determine analytically and empirically the economic, trade and environmental variables under which governments and agricultural firms that apply green production methods are favorably selected in the set of countries that adopt an eco-label program. They also investigate the strategic interactions that prevail between trading partners in their decision to adopt the eco-label in a subgame perfect Nash equilibrium in a two stages multi-country setting. Based on their theoretical framework, which can be extended to manufacturing industries, Figure 1 depicts the main incentives behind the government's decision to introduce an eco-label. The perceived gains resulting from the adoption of an eco-label are simultaneously related to (1) the stage of development of the adopting country, (2) the fixed cost of the eco-labelling scheme, (3) the relative production cost advantage of the country in producing the type of products covered by the eco-label program and (4) the strategic interactions between trade competitors.

2.1 Economy's Stage of Development

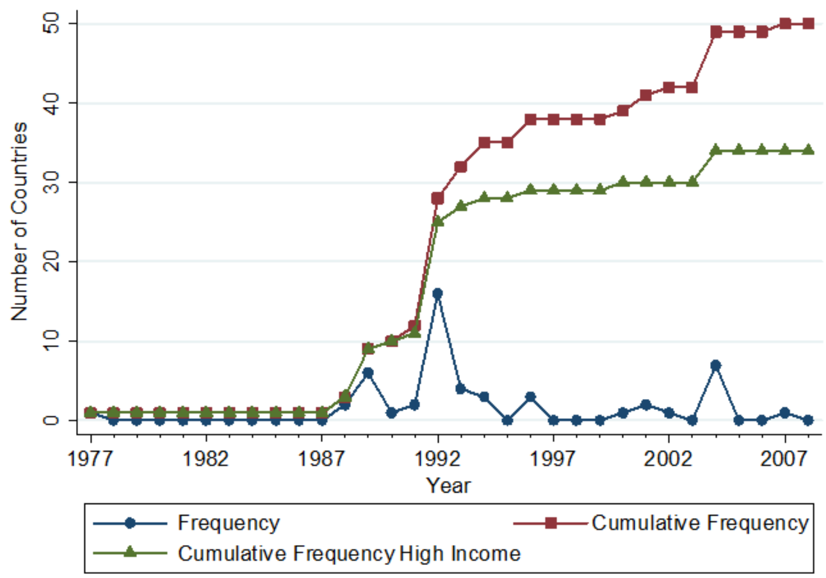
The decision to adopt an eco-label is mainly determined by the stage of development of the economy. This is partially confirmed by Figure 2 which highlights the fact that most high income countries were among the first to introduce an eco-labelling scheme[37]. In particular, several governments decided to introduce an eco-label during the 1980s and early 1990s, coinciding with the trend of market governance and self-regulation in environmental policy instrument (away from command-and-control measures). In recent years, the need to address the issues related to global warming and biodiversity has led to a renewed interest in eco-labelling schemes. One explanation for this trend, which is related to the debate about the environmental Kuznets curve, is that, as the economy becomes richer, individuals are more aware of environmental issues and ask for stringer regulation in order to reduce and reverse the environmental pollution trend resulting from industrialization.

Other features of the stage of development, which are linked with GDP per capita, but not necessarily perfectly correlated with it, are likely to enter the picture. Countries that face more social problems (e.g. high unemployment rate, child mortality rate, wage inequality, ...) are likely to assign a lower degree of concern for environmental issues, reducing the probability of introducing an eco-labelling scheme. Besides, market-based instruments like eco-labelling scheme are more likely to be implemented by efficient governments than command-and-control standards. Thus, the probability to introduce an eco-label program will be lower in countries characterized by high inefficiency and/or corruption, which prevent the private sector from adopting voluntary environmental initiatives. Somehow related to the issue of corruption, democratic governments with high political freedom can support more easily environmental quality improvement measures through voters' preferences (Magnani, 2000) and thus increase

[37] The peak in 1992 and 2004 correspond to the introduction and extension of the European eco-label program, EU(15) and EU(25), respectively. Note that several European countries (e.g. Germany (Blue Angel), Netherlands (Stichting Mileukeur), ...) introduced an eco-label program before the UE Flower ecolabel. For those cases, only the first eco-label scheme introduced is considered (see Appendix 7.2).

the odds of introducing an eco-label program. Firms might also be interested in supporting the introduction of an eco-labelling scheme, when the threat of direct environmental policy regulation on the environmental quality of the product is high. Therefore, countries whose production structure is characterized by a high share of dirty products might be more favourable to eco-label programs.

Figure 2: Time Pattern of Eco-label Adoption



2.2 Cost of Eco-label

Several reasons can lead a firm to opt to be a member of an eco-label program (e.g. improved corporate reputation, risks mitigation and management, competitive advantage, access to new markets, cost reductions in the long run,). Ultimately, the producers' decision is based on evaluating two dimensions (Sedjo et al., 2002):

– the extent to which the eco-label would increase the production costs (i.e. investment costs to comply with the eco-label's standards (indirect costs) and administrative costs associated with the eco-labelling procedure (direct costs));
– the extent to which consumers are willing to pay a price premium for eco-labelled products (i.e. predictability of the producers' future revenues).

Firms in large domestic markets will be able to dampen the fixed cost to be paid to be certified by a third party through economies of scale and improved learning curves. These elements play a key role in the development of differentiated goods (Bruce & Laroiya, 2007). As highlighted by Nadai (1999), the firms' eco-label adoption strategy is partially determined by the degree of heterogeneity between the sets of products sold. In addition, before the market phase, the government might face the opposition of some firms in the industry during the negotiation of the eco-label's criteria. These firms might want to try to block the agreement on the criteria or, if these criteria are nonetheless adopted by the authority, they can deliberately avoid the use of a label on their products.

Consumer sensitivity to the environment, which is related to the country's stage of development, is also essential for eco-labelling programs to be effective. Greaker (2006) demonstrates that the introduction of an eco-labelling scheme becomes preferable than setting a minimum standard only when the willingness to pay for eco-labelled products is sufficient. In fact, the proportion of environmentally concerned consumers in the economy and their willingness to pay for public good characteristics increase the probability to adopt an eco-label. Results from a number of studies (Teisl et al., 1999; Sammer et al., 2006) suggest that two of the major reasons why consumers choose eco-labelled products are consideration for the environment and/or for their own health. Several demographic and economic characteristics play a role in determining eco-friendly behaviors. For instance, younger and more educated individuals usually display a lower information processing cost and thus are assumed to be more proactive in terms of environmental quality requests. In particular, the level of the green premium price is stimulated if consumers are already environmentally conscious and able to express their preferences through their environmentally friendly consumption choices. In addition, countries characterized by larger population densities are usually in need of a better environmental quality, because the lives of more people are affected by pollution. Yet, the relationship between attitudes and behaviors with respect to the environment is not simple and straightforward. The reason is that consumers are often dealing with mixed motives[38].

[38] Several explanations have been provided, such as the "warm glow effect" (i.e. increased utility from the act of giving rather than receiving) or the "Veblen effect" (increase utility associated with the statue value given by the consumption) *versus* an excessive premium price charged or a lack of trust in the eco-labelled product (Peattie, 2001; Pedersen & Neergaard, 2006).

2.3 Relative Production Costs Advantage

The export orientation of the economy's manufacturing industry can partially determine the decision to introduce an eco-label program, when the country, as a net exporter, is able to cover the eco-labelling fixed costs by capturing the green premium in larger export markets. In addition to this trade factor, the comparative cost advantage of an economy is related to the diffusion of innovations among private firms. In order to be awarded an eco-label on their products, firms have to invest more in environmentally sound technologies. This suggests that the criteria of the eco-labelling scheme will implicitly orient firms' R&D. Eco-label regulators usually expect that producers will achieve innovation during the market phase in order to respect the criteria. Therefore, the diffusion of eco-organizational innovations can ultimately improve the relative production cost advantage of an economy in producing different products (Porter & Van der Linde, 1995). One can expect this mechanism to be even stronger, if producers are already familiar with environmental innovations (e.g. standard-setting, certification and accreditation procedures), reducing the cost of the negotiation and market phases of the eco-label program.

However, the introduction of an eco-labelling scheme might also lead to innovation distortions. In particular, during the eco-label criteria negotiation, producers may try to ensure that the standards will rely on the current technology they possess, ensuring the lowest environmental innovation costs as possible. Once the eco-labelling scheme is in place, producers might have no additional incentive to innovate beyond the eco-label's standards, even when they enjoy a larger profit margin in the market. Distortion might be further exacerbated by exporting firms afraid of the reaction of developing countries, which can file complaints over the non-respect of WTO's rules, and thus dissuading them from adopting higher environmental standards.

2.4 Strategic Interactions with Trade Competitors

Although countries are judicially independent, they are economically interdependent because of international trade and capital flows. Each government may face strategic interdependence on its national and export markets. Consequently, a country, which faces competitive challenges from a large number of countries whose level of environmental policy differs, will have some incentives to change its environmental regulation in response to other countries' policies. Since environmental attributes are unobservable, an eco-labelling scheme can serve as a screening mechanism or a signalling device. The governmental decision to introduce an eco-labelling program can thus be seen as a strategic environmental policy. A country's decision will depend on the existence and future existence of eco-label programs in other countries as well as the existence of a green premium in those countries. Therefore, the eco-label adoption interdependence is expected to increase with high economic relationship intensity and decreases with large trade cost. Three mechanisms can explain the strategic environment in adopting an eco-label.

First, a country, which faces competition from imports on its national market, might be willing to introduce an eco-labelling scheme whose criteria may be determined, intentionally or unintentionally, in favour of domestic firms. If domestic producers can adopt the eco-label more easily than foreign firms due to the criteria established, this may cause undesirable trade effects or trade frictions. Grolleau et al. (2007) investigate the implications of a possible strategic manipulation of eco-label programs that lead to raise the foreign rivals' production costs (Salop and Scheffman, 1983) through product categories definition, eco-labelling criteria and monitoring procedures. Beside modifying directly the foreign firms' production cost, the introduction of an eco-labelling program can increase the perceived quality of eligible domestic products and decrease that of noneligible imported product through market signalling[39]. Therefore, one of the "protectionist-like" effects could result from consumers valuing cleaner production, and hence, switching from the imported product to the domestically produced good once the standard becomes known. This change in consumption spending in favour of the eco-label product can lead to a decrease in the price of the non eco-labelled product, which ultimately can reduce the trade volume and worsen the terms of trade of the developing exporters.

Second, the emergence of environmental demands in export markets is more likely to be important in open economies. An exporting country, interested in extending its foreign market share (presumably to high income countries), may be interested in adopting an eco-label program. To promote the trade of the country's environmentally friendly products, the eco-labelling will thus include products and follow the criteria adopted by other national eco-label programs (e.g. China eco-label with respect to EU). Since demand for environmental quality tends to be income elastic, differences in product categories and criteria are expected to be greatest between high and low income countries. In particular, foreign producers may be required to meet criteria that are not relevant in their own country. This is particularly

[39] For instance, the share of eco-labelled paper for notebooks in the swedish and danish market has increased over time to about 80%.

problematic for small and medium enterprises in small, resources-constrained developing countries, as certification procedures and eco-labelling compliance require important funding (Piotrowski & Kratz, 1999). Beside these costs, producers in many low income countries face asymmetric information. They do not necessarily possess information about some eco-label programs and all the certification requirements. These obstacles are further exacerbated by the fact that advanced technologies used to define the standards of the eco-label might be patented and thus difficult to access if not unaffordable. Just like in the first case, it is the determination of the standards associated with the eco-label, rather than the information embodied in it, that explains why eco-labelling can affect the market access to (developing) exporters (UNEP, 2005). This becomes even more problematic when each developed country supports different underlying eco-labelling criteria making it almost impossible for the producers to exploit economies of scales.

Third, eco-labelling designed to reduce global environmental problems (e.g. CO_2 emissions or biodiversity) might face free-riding that can alter a country's decision to implement an eco-labelling scheme. Due to the transboundary nature of environmental degradation, consumers, whose country has implemented an eco-label program, can be tempted to free ride by reducing their own purchases of eco-labelled products. In fact, domestic buyers may expect foreigners customers to purchase eco-friendly products and thus choose to reduce their own purchases. This incentive will increase as the number of countries with an eco-labelling scheme is high, even if the potential green premium is large (Robertson, 2007). Domestic governments aware of this issue may decide to abandon the idea of an eco-label program as the number of countries with eco-labelling increases.

3 Joint Sampling in a Bayesian Heteroskedastic Spatial Probit Model

In this section, I present the limited dependent variable spatial model. The most popular method to estimate this type of interdependent binary models is the standard Metropolis-Hastings-Gibbs sampling approach proposed by LeSage (2000). A potential problem of this standard iterative method is slow convergence. To address this issue, I extend the joint updating approach suggested by Holmes and Held (2006) in a spatial setting. To illustrate the relative efficiency of this new algorithm, I perform a small Monte Carlo study. Finally, I review the different types of spatial marginal effects associated with the spatial probit model.

3.1 Spatial Autoregressive Probit Model

Assuming that each country assesses the costs and benefits of adopting an eco-label, a rational government will introduce an eco-labelling scheme only if it gains in welfare, expressed in monetary terms or net gain in utility[40]. Formally, $\mathbf{Y}_i^*$ denotes country *i*'s welfare associated with the adoption of the eco-label. Note that $\mathbf{Y}_i^*$ is by definition a latent or auxiliary variable and thus cannot be observed directly. What is observable is the binary indicator variable $\mathbf{Y}_i$ with entry 1 if the country has adopted an eco-label $\left(\mathbf{Y}_i^* > 0\right)$ and 0 otherwise $\left(\mathbf{Y}_i^* < 0\right)$. As explained in the previous section, a government's decision to introduce an eco-label may depend on the related decision of other close countries. Yet, assuming incorrectly that the decision of country *i* is independent of the decision of the *N*-1 remaining countries leads to biased as well as inconsistent and inefficient estimates which voids subsequent hypothesis testing (LeSage & Pace, 2009). That is why the latent variable is specified as a spatial autoregressive probit model[41]:

$$
\begin{aligned}
&\mathbf{Y}^* = \rho \mathbf{W}\mathbf{Y}^* + \mathbf{X}\beta + \mathbf{u} \qquad (1)\\
&\mathbf{Y} = \mathbf{1}\left[\mathbf{Y}^* > \mathbf{0}\right]\\
&\mathbf{u} \sim N\left(0, \sigma^2 \mathbf{V}\right)\\
&\mathbf{V} = diag\left(v_1, v_2, \ldots, v_N\right)
\end{aligned}
$$

where $\mathbf{Y}^*$, $\mathbf{Y}$ and $\mathbf{u}$ are $N \times 1$ vectors. The parameter ρ, also known as the spatial lag, is associated with the non-negative row-standardized exogenous $N \times N$ matrix $\mathbf{W}$. This spatial weight matrix, whose diagonal elements are zero, determines the form of the interdependence across country-pairs. This spatial autoregressive parameter can be seen as a reaction function which relates a country's choice about whether to introduce an eco-label to the existence of an eco-label in spatially close economies. Additional *K* explanatory variables are included in the $N \times K$ matrix $\mathbf{X}$. To account for potential spatial heterogeneity and outliers, the variance of the error terms, $\mathbf{V}$, is not constant. Following LeSage (1997, 2000), I introduce a set of variance scalars $\left(v_1, v_2, \ldots, v_N\right)$ as unknown parameters to be estimated. This is important, because if a given country follows a different pattern than the majority of spatial observations, the errors

[40] Most theoretical papers focusing on the labelling procedure consider an authority that maximizes a social surplus which depends on the profits of the firms, the consumers' surplus, the environmental damage associated with the production of the good as well as other potential costs related to the introduction of the eco-label (Greaker, 2006). Alternatively, the objective function of the government may be represented by a weighted average of social welfare and political contributions, to include political economy considerations.

[41] As highlighted by LeSage & Pace (2009), the cross-sectional spatially autocorrelated lag model, which is related to the spatiotemporal model, provides a long term perspective.

would no longer be normally distributed (i.e. fat-tailed errors associated with a Student-*t* distribution). The associated parameter estimates would thus be inconsistent if this was not accounted for.

Methods for properly estimating and analyzing equation (1) have been the topic of a relative small body of research in the spatial econometrics literature. The issue is that the introduction of the spatial lag leads to simultaneity biases as well as additional heteroskedasticity in the error terms. The reduced form of expression (1) highlights this issue:

$$\mathbf{Y}^* = (I - \rho\mathbf{W})^{-1}(\mathbf{X}\beta + \mathbf{u}) \tag{2}$$

The heteroskedasticity as well as the spatial dependence in the error term render standard probit approach inappropriate ($\mathrm{cov}\left((I-\rho\mathbf{W})^{-1}\mathbf{u}\right) = \sigma^2 (I-\rho\mathbf{W})^{-1}\mathbf{V}(I-\rho\mathbf{W}')^{-1}$). In fact, the presence of spatial autocorrelation makes the traditional maximum likelihood method less practical. The main reason is that the likelihood function requires to evaluate the joint distribution of the *N* interdependent outcomes, which is not the product of the *N* marginal distributions, but involves *N*-dimensional integration and the determinant of the $N\times N$ matrix **W**. To see this point more formally, the likelihood function of equation (1) is expressed as follows:

$$\begin{aligned} L(\rho,\beta,\sigma^2,\mathbf{V};\mathbf{Y}^*,\mathbf{W}) &= (\sigma^2)^{-N/2}\left|I_N - \rho\mathbf{W}\right|\left|\mathbf{V}^{-1}\right|\exp\left[-\frac{1}{2\sigma^2}\mathbf{u}'\mathbf{V}^{-1}\mathbf{u}\right] \\ &= \sigma^{-N}\prod_{i=1}^{N}(1-\rho\lambda_i)\prod_{i=1}^{N}v_i^{-½}\exp\left[-\sum_{i=1}^{N}\frac{u_i^2}{2\sigma^2 v_i}\right] \end{aligned} \tag{3}$$

where u_i is the *i*th element of the error vector $\mathbf{u} = (I_N - \rho\mathbf{W})\mathbf{Y}^* - \mathbf{X}\beta$. Note that the determinant of the Jacobian $|I-\rho\mathbf{W}|$ is approximated by $\Pi_{i=1}^{N}(1-\rho\lambda_i)$ with λ_i representing the i^{th} eigenvalue of the matrix **W**.

To avoid the direct calculation of multiple integrals in the likelihood function,which can be analytically intractable, several estimators have been proposed (Fleming, 2004; Franzese & Hays, 2008). McMillen (1992) is the first to suggest an Expectation-Maximization (EM) algorithm. His approach consists of replacing the discrete dependent variable with the expectation of the underlying continuous latent variable and maximizing its likelihood function until convergence is reached. Yet, this method faces several drawbacks (LeSage, 2000). First, the

EM algorithm does not provide standard-error for the spatial lag. Second, the method requires an arbitrary parameterization of the heteroskedasticity caused by the introduction of spatial dependence. Third, the approach is highly computation intensive when the number of cross-sections is large. To address the issue of spatial heteroskedasticity, Case (1992) proposes an alternative estimator that groups each cross-section into regions whose errors are assumed to be strictly independent of each other. Instead of expressing the spatial discrete choice model as a maximum likelihood function, Pinkse & Slade (1998), among others, derive the necessary moments conditions and apply a two-step Generalized Method of Moments estimator. Both Case (1992) and Pinkse & Slade (1998) approaches ignore standard cross-section heteroskedasticity making them consistent but not necessarily efficient estimators. More recently, Beron et al. (2003) extend the Recursive-Importance-Sampling (RIS) method to estimate consistently the spatial probit and compute the associated standard-errors necessary for inference. The main disadvantage of this simulation method is its computational burden, which makes it less practical to account for heteroskedasticity in the error term[42]. To address all of these issues, LeSage (2000) extends the Bayesian Markov-Chain-Monte-Carlo (MCMC) method to a spatial discrete choice model by using the Metropolis-Hastings-within-Gibbs sampling approach. The first advantage of the Bayesian strategy is to be able to derive the condition distribution of each parameter, and thus compute different moments of the distribution (e.g. mean, standard-error,...). The second advantage is its flexibility to account for the heteroskedasticity in the error terms (Smith and LeSage, 2004). That is why equation (1) will be estimated using the Bayesian MCMC approach.

3.2 Standard Bayesian MCMC Approach

According to the Bayesian approach, the product of the likelihood function and the prior density, which both depend on certain assumptions, determines the posterior distribution of the parameters that fits the data best[43]. Thus, in order to estimate the set of parameters β, $\mathbf{V}$ and ρ their associated priors ($\pi(\cdot)$) have to be specified independently of each other. First, the explanatory variables are assigned a normal prior, $\pi(\beta) \sim N(c,s)$. Second, in order to account for heteroskedastic variance $\sigma^2 v_i$, the relative variance parameters, $\mathbf{V} = diag(v_1, v_2, \ldots, v_N)$ **,** are assumed to follow an independent $\chi^2(r)/r$ distribution, which depends on the single

[42] In their empirical application, the estimation method proposed by Beron et al. (2003) does not rule out explosive spatial dependence $\hat{\rho} > 0$ (see Table 4 p. 292), which can be problematic.

[43] See Holloway et al. (2002), Thomas (2007) or LeSage & Pace (2009) for a more thorough introduction of Bayesian theory extended to the spatial probit.

parameter *r*, $\pi(r/v_i)\sim iid\ \chi^2(r),\ i=1,...,N$. The constant σ is usually set to 1. Third, the spatial lag is assumed to be distributed according to a uniform distribution, $\pi(\rho)\sim U(\lambda_{\min}^{-1},\lambda_{\max}^{-1})$, where $\lambda_{\min}^{-1}$ and $\lambda_{\max}^{-1}$ represent, respectively, the minimum and maximum eigenvalues of the matrix $\mathbf{W}$. Another alternative is to assume a beta prior for the spatial autoregressive parameter $\pi(\rho)\sim \mathrm{B}(b,b)$ (LeSage & Parent, 2007). Based on these priors, LeSage (2000) extends the work of Albert & Chib (1993) and Geweke (1993) to derive the conditional posterior distributions of the set of parameters:

$$p(\beta \mid \rho,\mathbf{V},\mathbf{Y},\mathbf{Y}^*)\sim \mathrm{N}[\mathbf{C},\ \mathbf{S}] \tag{4}$$

$$\mathbf{C}=\mathbf{S}^{-1}\left[\mathbf{X}'\mathbf{V}^{-1}(I_N-\rho\mathbf{W})\mathbf{Y}^*/\sigma^2+cs^{-1}\right]$$

$$\mathbf{S}=\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}/\sigma^2+s^{-1}$$

$$p(v_i \mid \rho,\beta,\mathbf{V}_{-i},\mathbf{Y},\mathbf{Y}^*)\propto (u_i^2/\sigma^2+r)/v_i \tag{5}$$

$$p(\rho \mid \beta,\mathbf{V},\mathbf{Y},\mathbf{Y}^*)\propto |I_N-\rho\mathbf{W}|\, e^{-(½\sigma^2)(\mathbf{u}'\mathbf{V}^{-1}\mathbf{u})} \tag{6}$$

where $\propto$ means that the expression on the left-hand side is proportional up to a constant to the expression on the right-hand side. $\mathbf{V}_{-i}$ denotes all the elements of matrix $\mathbf{V}$ beside v_i. Note that expression (6), the prior of the spatial lag, cannot be generated from a standard normal distribution, that is why the Metropolis-Hastings algorithm, which is a standard accept-reject algorithm, has to be used. In the case of the alternative beta prior, a univariate numerical integration is applied to construct the conditional posterior distribution of the spatial autoregressive term and then sample it by *inversion* (LeSage & Pace, 2009).

Finally, the posterior distribution of the latent/auxiliary variable, $\mathbf{Y}^*$, conditional on the parameters is specified as a truncated multivariate normal distribution:

$$p(\mathbf{Y}^* \mid \rho,\beta,\mathbf{V},\mathbf{Y})\sim \mathrm{N}[\boldsymbol{\mu},\ \boldsymbol{\Sigma}]\,Ind(\mathbf{Y},\mathbf{Y}^*) \tag{7}$$

$$\boldsymbol{\mu}=(I_N-\rho\mathbf{W})^{-1}\mathbf{X}\beta$$

$$\boldsymbol{\Sigma}=\left[(I_N-\rho\mathbf{W}')\mathbf{V}^{-1}(I_N-\rho\mathbf{W})\right]^{-1}$$

where $Ind(\mathbf{Y},\mathbf{Y}^*)$ represents an indicator function which truncates from the left by zero if $\mathbf{Y}_i=1$ and from the right by zero if $\mathbf{Y}_i=0$. Note that the marginal distribution of the individual elements of $\mathbf{Y}^*$, $p(\mathbf{Y}_i^* \mid \rho,\beta,\sigma,v_i,\mathbf{Y}_i)$, does not correspond to an univariate truncated normal. LeSage's method relies on the Geweke (1991) approach to sample the conditional distribution for $\mathbf{Y}_i^*$ from a truncated multivariate normal distribution subject to independent inequality linear constraints (LeSage & Pace, 2009). Although there are other methods to simulate a truncated multivariate variables subject to inequality linear constraints, including Rodriguez-Yam, Davis & Scharf (2004) efficient approach, the standard method in spatial probit is the Geweke method.

Once the complete conditional distributions of all parameters in the model are specified, the MCMC sampling method can be implemented. While in standard Monte-Carlo simulation, the draws are generated independently based on a specified underlying distribution, in Gibbs sampler, each draw depends on the previous one in such a way that the produced samples display properties identical to those of the joint population. Thus, LeSage (2001) suggests taking iterative random draws from (4), followed by (5) and (6), and then (7). With a sufficient number of draws, the sample statistics can approximate the set of estimates that converges in the limit to the joint posterior distribution of the parameters.

3.3 Joint Updating Bayesian MCMC Approach

However, the main drawback of LeSage's iterative sampling method is the presence of strong posterior correlation between β, $\mathbf{Y}^*$ and ρ. Although a correlated draws chain provides an unbiased picture of the distribution, the number of draws has to be sufficiently large. That is why, in order to tackle the issue of potential slow mixing in the Markov chain, I follow Holmes & Held (2006) and extend the joint updating of β and $\mathbf{Y}^*$ to a spatial framework by using the product rule to decompose the joint probability of β and $\mathbf{Y}^*$ as follows:

$$p(\beta, \mathbf{Y}^? \mid \rho, \mathbf{V}, \mathbf{Y}) = p(\mathbf{Y}^? \mid \rho, \mathbf{V}, \mathbf{Y})\, p(\beta \mid \rho, \mathbf{V}, \mathbf{Y}, \mathbf{Y}^?)$$

The auxiliary variable is now updated according to its marginal distribution once it has been integrated over β (see equation (4)):

$$p(\mathbf{Y}^* \mid \rho, \mathbf{V}, \mathbf{Y}) \propto \mathrm{N}[\boldsymbol{\mu}, \boldsymbol{\Omega}]\, Ind(\mathbf{Y}, \mathbf{Y}^*) \tag{8}$$
$$\boldsymbol{\mu} = (I_N - \rho\mathbf{W})^{-1}\mathbf{X}\beta$$
$$\boldsymbol{\Omega} = (I_N - \rho\mathbf{W})^{-1}[\mathbf{X}\mathrm{s}\mathbf{X}' + \mathbf{V}](I_N - \rho\mathbf{W}')^{-1}$$

In other words, the proposed approach consists of sampling $\mathbf{Y}_i^*$ according to its marginal multivariate truncated normal function $\left(p(\mathbf{Y}_i^* \mid \rho, \mathbf{V}, \mathbf{Y}_{-i}^*, \mathbf{Y}_i)\right)$ and updating β's conditional means

(**C**) after each update to $\mathbf{Y}_i^*$. Once all the individual elements of $\mathbf{Y}^*$ have been sampled, β is generated based on its conditional normal distribution (4). More formally, the new procedure consists of iteratively[44]:

1. updating $\{\beta, \mathbf{Y}^*\}$ jointly according to (8), given ρ and **V**;
2. updating **V** according to (5), given the remaining parameters;
3. updating ρ according to (6), given the remaining parameters.

Thanks to the joint updating approach, the mixing and sampling efficiency in the chain should be improved. In order to compare the performance between the standard iterative sampler and the joint updating sampler, I conduct a small Monte-Carlo simulation study. The latent variable $\mathbf{Y}^*$ is generated according to equation (2) and used to determine the values of $\mathbf{Y}_i$ as follows: $\mathbf{Y}_i = 1$ if $\mathbf{Y}_i^* > 0$ or $\mathbf{Y}_i = 0$ otherwise. The matrix of explanatory variables includes a constant and two standard random normal variables. The coefficient vector is set to the following values: $\beta = (0,1,-1)'$. The spatial weight matrix W is a row-standardized rook-type matrix of order 10 (i.e. the ten nearest neighbors). The spatial autoregressive parameter ρ is set to 0.75. For simplicity, the individual shocks are assumed to follow a standard Gaussian distribution, whose variance is homoskedastic $(\mathbf{u} \sim N(0, I_N))$. Four different sample sizes are considered: $N = \{250, 500, 750, 1000\}$.

For each of theses designs, 10 samples are generated and estimated according to 6 different samplers:

1. iterative update with Geweke approach to simulate the latent variable and Metropolis-Hastings algorithm to draw the spatial lag.
2. iterative update with Geweke approach to simulate the latent variable and numerical integration to draw the spatial lag.
3. iterative update with Rodriguez-Yam, Davis & Scharf approach to simulate the latent variable and Metropolis-Hastings algorithm to draw the spatial lag.
4. iterative update with Rodriguez-Yam, Davis & Scharf approach to simulate the latent variable and numerical integration to draw the spatial lag.
5. joint update with Metropolis-Hastings algorithm to draw the spatial lag.
6. joint update with numerical integration to draw the spatial lag.

[44] The algorithms are written in Matlab and available upon request.

To my knowledge, this the first time that the approach suggested by Rodriguez-Yam, Davis & Scharf (2004) and the joint update have been applied to estimate a spatial probit. In order to measure efficiency, I compute, for each chain, the CPU computation time in seconds, the parameters' average Euclidean update distance between each iteration as well as the Raftery-Lewis (1992, 1995) convergence statistic. A large value of Euclidean distance means good mixing while a high total number of draws necessary to ensure an i.i.d chain implies high autocorrelation in the chain and slow convergence.

Table 1: MCMC Algorithms Performance

	Iterative Update with Geweke and Metropolis-Hastings			Iterative Update with Rodriguez-Yam et al. and Metropolis-Hastings			Joint Update with Metropolis-Hastings		
N	CPU Time	Total Draws	Param. Distance	CPU Time	Total Draws	Param. Distance	CPU Time	Total Draws	Param. Distance
250	51	5268	0.145	37	4032	0.149	45	4164	0.192
500	176	5130	0.103	135	4209	0.106	170	3844	0.137
750	390	5172	0.083	313	4118	0.086	416	3947	0.111
1000	725	5160	0.072	585	3863	0.075	841	3816	0.096

	Iterative Update with Geweke and Numerical Integration			Iterative Update with Rodriguez-Yam et al. and Numerical Integration			Joint Update with Numerical Integration		
N	CPU Time	Total Draws	Param. Distance	CPU Time	Total Draws	Param. Distance	CPU Time	Total Draws	Param. Distance
250	55	5580	0.158	41	4065	0.163	49	3911	0.204
500	194	5164	0.112	153	3947	0.116	187	3916	0.145
750	444	5251	0.092	368	3813	0.095	469	3811	0.118
1000	839	5035	0.078	715	3770	0.081	956	3905	0.101

Note: Performance measures based on the data generating process (2) with $\rho = 0.75$ and $\beta = (0, 1, -1)'$.

Table 1 presents the results averaged over the four parameters (β and ρ) and the 10 runs. First, the method proposed by Rodriguez-Yam et al. (2004) is more efficient than Geweke (1993) with higher mixing and faster convergence. Second, applying numerical integration to draw the spatial lag improves also the mixing and usually reduces autocorrelation in the chain. Third, although it is relatively more time consuming when the sample size is large, the joint updating sampler yields a larger average distance jumped between iterations and usually relies in smaller total draws. In fact, once the performances are standardized by their respective computation time, the joint sampling algorithm relies on a 45% smaller total number of draws to ensure convergence[45]. In addition, in comparison to Geweke method, the joint sampler yields 15 to

[45] Appendix 7.1 reports the relative performance of the iterative samplers with respect to the joint updating sampler.

50% more mixing in the chain. Interestingly, Rodriguez-Yam et al.'s approach can lead to the same level of mixing than the joint updating method when the number of cross-sections is particularly large. Overall, these findings suggest that the spatial probit model should be estimated by the joint update algorithm, especially when the sample size is relatively small (as it is the case in this study).

3.4 Spatial Marginal Effects

Once the model has been estimated, it is crucial to be able to interpret the coefficients (first and second moments of the conditional distribution). Yet, just like in standard discrete choice models, parameter estimates from a spatial probit cannot be interpreted directly. They must be transformed to yield estimates of the marginal effects, i.e. a change in the predicted probability associated with changes in the explanatory variables. However, unlike classical approaches, models including a spatial lag of the dependent variables have to be interpreted in a special way (Beron & Vijverberg, 2004). This comes from the fact that a change in a single country associated with a given explanatory variable will lead to a direct impact on the country itself, but can potentially affect all other countries indirectly. In fact, the spatial probit model allows for complex feedback loops that might take place when a shock in country *i* affects countries *j* and *k* to finally change back country *i*. The derivative of Y_i with respect to a variable *r* in country *j*, (*j* = 1,…,*N*) takes the following form:

$$\hat{\Xi}_{ijr} \equiv \frac{\partial E\left[Y_i \mid X_{jr}\right]}{\partial X_{jr}} = \phi\left(\left[\left(I_N - \hat{\rho}\mathbf{W}\right)^{-1} X\hat{\beta}\right]_i / \hat{s}_i\right)\left[\left(I_N - \hat{\rho}\mathbf{W}\right)^{-1}\right]_{ij} \hat{\beta}_r / \hat{s}_i$$

where ϕ is the density function of a standard normal distribution and $\hat{s}_i = \hat{\sigma}^* \Sigma_i \hat{\varpi}^*_{ij}$ with ω_{ij} being the ij^{th} elements of the matrix $(I - \rho\mathbf{W})^{-1}\mathbf{u}$.

Since it might be difficult to keep track of the $N^2 \cdot K$ spatial effect estimates, when the spatial weight matrix size and the number of explanatory variables are large, LeSage & Pace (2009) suggest some useful summary measures of these effects for each explanatory variable *r*:

- Average Total Effect: $\hat{\Xi}^T_r = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Xi}_{ijr}$.
- Average Direct Effect: $\hat{\Xi}^D_r = \frac{1}{N}\sum_{i=1}^{N}\hat{\Xi}_{iir}$.
- Average Indirect Effect: $\hat{\Xi}^I_r = \hat{\Xi}^T_r - \hat{\Xi}^D_r$.

Going back to the eco-label analysis, these impacts could be interpreted in the following way. The direct effect indicates how a rise in an explanatory variable across the sample of countries would affect the expected probability of the (average) country to adopt an eco-label scheme. The indirect effect measures how a shift in this explanatory variable would affect the (average) neighboring country's eco-label program adoption decision. Obviously, the size of these types of feedback will depend on the position and degree of connectivity of each country with each other (spatial weight matrix W), the strength of spatial dependence (spatial lag ρ) and the importance of the explanatory variables (parameters β).

4 Empirical Results

The empirical analysis of the effect of eco-labels on trade flows suffers from fundamental data deficiencies (OECD, 2004). Because import and export statistics apply the same universal codes for tracking trade flows of eco-labelled and non-eco-labelled products, there is no information available on international trade in eco-labelled products. That is why in this paper the dependent variable is a dummy variable indicating whether a country has implemented a type I multi-sector eco-label program[46]. Type I eco-labelling is owned and operated by third parties, which may be governmental organizations or private non-commercial entities, and is awarded for products and manufacturing processes which meet certain environmental criteria. Unlike one single product category label (e.g. canned tuna caught in a dolphin safe way) or private eco-label (e.g. certified wood), multiproduct eco-labelling covers a wide range of different manufactured products categories (e.g. Germany's Blue Angel). The firm's participation is completely voluntary. Indeed, manufacturers must pay for the right to display the label and demonstrate continued adherence to relevant product guidelines to maintain their certifications. Data on the adoption of a type I eco-label is taken from the Global Eco-labelling Network as well as the New Zealand's Ministry of Economic Development and the website ecolabelling.org. Appendix 7.2 reports the countries which have adopted a multi-sector eco-label before 2009. Due to missing data, Liechtenstein, Malta, Singapore and Taiwan are not considered, although they have implemented an eco-labelling scheme. In addition, Costa Rica, South Africa, Turkey and Zimbabwe have introduced a tourism eco-label. But since this type of eco-labelling focuses on non-traded goods, they are deliberately omitted.

[46] The International Organization for Standardization (ISO) distinguishes two additional categories of ecolabels. Type II is associated with informative environmental self-declaration claims which are not verified by any independent third party. Type III covers voluntary programs that provide quantified environmental data of a product, under pre-set categories of parameters set by a qualified third party and based on life cycle assessment, and checked by that or another qualified third party.

The selection of the explanatory variables is mainly dictated by their availability in order to maximize the number of countries in the sample ($N = 141$). The table in Appendix 7.4 lists the countries considered in this paper. In order to reduce multicollinearity issues, I deliberately restrict the number of explanatory variables. Multicollinearity leads to technical issues and convergence problems. The table in Appendix 7.5 presents the different factors and proxies as well as their sources considered in this paper. Unless specified otherwise, the data are averaged over the 2003-2005 period. The tables in Appendix 7.6 and 7.7 report the descriptive statistics of the variables, Moran's spatial autocorrelation statistic as well as the correlation matrix. As suspected, the eco-label dummy variable displays significant positive spatial autocorrelation. The same is true for the remaining variables considered in this study, which definitively calls for a spatial framework.

In order to avoid any endogeneity issue in the estimation process, the interdependence of the spatial autoregressive parameter (ρ) is based on a geographical distance weighting scheme ($\mathbf{W}$), which is by definition strictly exogenous. The benchmark spatial weight matrix is based on a negative exponential distance measure:

$$w_{ij} = \begin{cases} \exp\left(\frac{-distance_{ij}}{500}\right) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

where $distance_{ij}$ is the bilateral geographical distance between the capital of country *i* and *j* (in kilometers). The advantage of this spatial scheme is to give a positive weight to countries which are close to each other (within a region) and almost zero weight to nations that are geographically remote from each other. Note that the spatial weight is row-standardized so that the sum of each row is equal to one.

Before implementing the joint updating MCMC algorithm, several decisions have to be made. These include the values for the priors' parameters, the number of total iterations, the number of initial burn-in to be discarded and the spacing between iterations to be retained for the inference[47]. In fact, it is important to determine whether the sampling chain has converged to a stationarity distribution. So the question is what is the number of runs until the Markov chain approaches stationarity. According to Raftery and Lewis (1992, 1995) diagnostic statistics and autocorrelation measures, the estimation relies on a Monte Carlo chain which is based on at least 15,000 draws with 1,000 burn in draws and a thinning factor of at least 5. The main reason

[47] Unless specified otherwise, the priors' settings are set as follows: $\pi(\beta) \sim N(0, I_N \cdot 1e^{12})$, $\pi(\rho) \sim \mathrm{B}(1.01, 1.01)$, and $\pi(4/v_i) \sim \chi^2(4)$, $i = 1, \ldots, N$.

to ignore 1,000 draws is to make sure that there is no systematic information left in the random numbers generation process for the remaining draws. In addition, only every 5^{th} draw is saved for inference in order to reduce autocorrelation and avoid unjustified higher standard deviation in the parameters. If the chain for each parameter were to display high autocorrelation, further draws from the chain should be skipped in order to get proper inference on the standard deviation.

Table 2 reports the main results. For each spatial specification, the average coefficients' posterior and its associated standard errors are reported in the first column. The estimated parameters' t-statistics are not computed. The main reason is that normalization by standard errors does no longer leads to a Student distribution, because the simulated draws are themselves approximation to Student distributions (Holloway et al., 2002). Here, a parameter is significant at the 5 percent significance level, if the quantiles at the 2.5 and 97.5 percent have the same sign, i.e. zero does not belong to the 95% interval. In addition, as suggested by LeSage & Pace (2009), the second to fourth columns report the average marginal direct, indirect and total effect, respectively. But first, I estimate the model assuming there is no spatial dependence in the eco-label decision in a homoskedastic framework.

The traditional probit estimator performs relatively poorly. Most variables are not significant at the conventional level, except the number of international environmental treaties adopted, the share of high technology exports and the number of ISO14001 certificates. The presence of large standard errors might be due to the strong assumption of homoskedasticity in the error term and a potential variable omission when spatial dependence is not accounted for. These issues are addressed by estimating a heteroskedastic spatial probit model. According to different convergence checks instruments, the different Monte Carlo chains have reached stationarity[48]. In particular, the dependence statistic *I*, which reports the ratio of the total number of draws required to achieve a 5 percent test accuracy and the minimum number of draws needed to ensure an identically and independently distributed draws, is lower than 5 $(I = 1.93)$. Therefore additional draws are not required and proper inference can thus be performed.

[48] In order to check the convergence of the MCMC samplers, autocorrelation, Raftery-Lewis and Geweke diagnostics have been performed. To save space, they are not reported here but are available upon request.

Once interdependence in the eco-label adoption is taken into account, the results improve significantly. In fact, although not necessarily significant at the conventional level, most parameters display the expected sign. In particular, countries which have reached a high stage of development are more likely to adopt an eco-label. This is confirmed by the fact that GDP per capita affects positively the environmental label decision. As the economy growths, the government has the incentive and the means to introduce an eco-labelling scheme. Interestingly, the pollution pressure captured by emissions of SO_2 decreases the probability of implementing an eco-label scheme. This counterintuitive result might be due to three reasons. First, although several high income countries are large SO_2 emitters despite possessing an eco-labelling scheme (e.g. Switzerland or Sweden), many low income countries without eco-label generate even higher levels of SO_2 emissions (e.g. Mali or Nigeria). Second, most environmental labellings cover products in industries characterized with relatively low SO_2 emissions (e.g. footwear or textile)[49]. Third, since the SO_2 emissions' impact on the environment and health is relatively local, the need of direct regulation can be higher and the government might prefer command-and-control measures instead of self-enforcement instruments. This finding is robust to potential outliers characterized by a large mining sector (e.g. Chile and Peru). However it is no longer significant when other proxies for pollution pressure like urban particles, indoor air pollution and regional ozone are considered.

Going back to the estimation results, the existence of scale effects captured by the manufacture value added share, increases the probability of eco-labelling implementation. In other words, if the costs associated with the eco-label procedure can be compensated by economies of scales, the probability to introduce an eco-label is positively significant. The same does not hold for the presence of a green price premium proxied by the population share below 45 years old. This probably comes from the fact that it is an imperfect proxy, since most high income countries with an eco-label in place have fewer young adult people relative to developing countries. Results show that economies displaying a relative production cost advantage through previous environmental preferences experience, innovation and a high share in high-technology exports are more inclined towards adopting a voluntary environmental program. This result is in line with Grolleau & El Harbi (2008), who find that economies characterized by higher technological innovation capacities use the eco-labelling scheme as a tool to enhance and reinforce their innovation potential. In fact, the diffusion of environment friendly organizational innovations

[49] Non-ferrous metals, petroleum, non-metallic mineral and chemical products are associated with large level of SO_2 emissions (see Emission Data Base for Global Atmospheric Research).

through ISO14001 certificates seems to lead more easily to the adoption of an eco-label program. Although there seems to be a substituability relationship between the adoption of an eco-label and the average manufacturing tariff, it is not statistically significant. Finally, the results confirm the interdependent nature of eco-labelling decisions, because nations which are closer to each other in geographical terms display a higher probability of adopting an eco-label program.

Table 2: Estimation Results

Variable	Standard Probit		Spatial Probit with Exponential Distance			
	Coefficient	Marginal Effect	Coefficient	Direct Effect	Indirect Effect	Total Effect
Constant	-0.043		2.662			
	[4.104]		[7.29]			
Real GDP per Capita	0.099	0.025	0.195***	0.007***	0.001**	0.009***
	[0.06]	[0.014]	[0.087]	[0.003]	[0.001]	[0.004]
SO_2	-0.026	-0.007	-0.087*	-0.003*	-0.001*	-0.004*
	[0.026]	[0.008]	[0.055]	[0.002]	[0.001]	[0.002]
Political and Civil Rights	-0.157	-0.04	-0.216	-0.009	-0.001	-0.01
	[0.219]	[0.048]	[0.336]	[0.014]	[0.003]	[0.016]
Population Below 45	-0.087	-0.022	-0.17	-0.006	-0.001	-0.008
	[0.081]	[0.015]	[0.123]	[0.005]	[0.001]	[0.006]
Environmental Treaties	0.067***	0.017***	0.111***	0.004***	0.001**	0.005***
	[0.024]	[0.01]	[0.039]	[0.001]	[0]	[0.002]
Manufacture Value Added	0.065	0.017	0.194***	0.007***	0.001**	0.009***
	[0.044]	[0.015]	[0.074]	[0.003]	[0.001]	[0.003]
High Technology Exports	0.043*	0.011*	0.094***	0.004***	0.001**	0.004***
	[0.023]	[0.008]	[0.04]	[0.002]	[0]	[0.002]
ISO14001	0.289***	0.074***	0.399***	0.016***	0.003**	0.018***
	[0.094]	[0.03]	[0.138]	[0.006]	[0.001]	[0.006]
Manufacturing Tariff	-0.104	-0.026	-0.09	-0.004	-0.001	-0.004
	[0.075]	[0.022]	[0.105]	[0.004]	[0.001]	[0.005]
Spatial Lag			0.155**			
			[0.06]			
Observations (with eco-label)	141 (47)		141 (47)			
Total Draws / Omitted Draws	-		15000 /1000			
Thinning Factor	-		5			
I-statistic	-		1.927			
Log-likelihood	-17.568		-18.712			
Pseudo R^2	0.804		0.792			

Note: The mean of the posterior distribution is reported as coefficient while the standard deviation is reported in the brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 3: Model's Predictive Power

	Standard Probit Model Predicted Eco-label	Standard Probit Model Predicted No Eco-label	Spatial Probit Model Predicted Eco-label	Spatial Probit Model Predicted No Eco-label	Total
Observed Eco—label	37	10	40	7	47
Observed No Eco-label	3	91	1	93	94
Total	40	101	41	100	141

In comparison with the classical probit model, there are major differences in terms of average coefficients and marginal effects, the spatial direct marginal effects being theoretically equivalent to the standard marginal effects. Beside the fact that the number of explanatory variables significantly different from zero is higher in the spatial probit estimation, accounting for spatial dependence usually yields higher average estimates but lower average direct marginal effects in absolute value. This is mainly due to a large average coefficient of the constant in the spatial probit and a rate of decay relatively faster of the spatial dependence. This also explains why the indirect effects, which can be interpreted as the probabilistic impact of a rise in the neighborhood of a given explanatory variable on the eco-label decision, are always smaller than the direct effects. The largest average total effects are associated with a high number of ISO14001 certificates, a strong level of economic development and potentially large scale effects. These empirical findings can explain why the pattern of eco-labelling adoption between developed and developing countries has been unequal. Low income countries do not possess the economic structure that would allow them to bear the fixed costs of eco-labelling through a large green premium and a high innovation level. These obstacles might be further exacerbated by the lack of perceived or actual credibility of the eco-labelling scheme in those countries. This can explain why the few developing economies with an eco-label program in place are operated by governmental entities.

In order to further investigate the impact of interdependence, it can be of interest to compare the predictive power of the spatial probit model with respect to the standard probit model, because the McFadden pseudo R^2 can be misleading. As in most empirical studies considering a probit framework, I assume that the model is able to predict the eco-label adoption (i.e. $Y_i = 1$) when the predicted probability is equal to or larger than 50 percent. In the benchmark sample, the proportion of countries having introduced an eco-label program is about 33.33 percent. The fact that the sample is *unbalanced* makes any prediction more difficult. Yet, Table 3 suggests that accounting for spatial dependence definitively increases the explanatory power of the model. The standard probit model predicts 90.78 percent of the cases correctly, while its spatial extension displays a higher predictive power with 94.33 percent. Among the eco-label adopters, 85.11 percent are also predicted correctly by the spatial model and only 78.72 percent by the simple probit model. Once again, this highlights the importance of accounting for interdependence in the eco-label adoption. Interestingly, according to the spatial probit model, some less developed countries should not have implemented an eco-label (e.g. Indonesia, Russia, Ukraine,...), while Chile should have adopted one. Actually at the end of 2009 the Chilean government proposed a bill that required producers, distributors and importers to label goods with information on the environmental impacts posed by their products.

5 Robustness Check

In order to investigate the robustness of the previous findings, several sensitivity analysis are performed. First, I check if the estimates are sensitive to a modification of the Markov Chain Monte Carlo settings. Second, different expressions of the spatial weight matrix are considered. Finally, I estimate several econometric specifications, including the spatial error and Durbin version of the benchmark model in order to account for some potential omission variables. Overall, the conclusions based on the benchmark model prevail and confirm the existence of spatial dependence in the eco-labelling decision.

5.1 MCMC's setting

Several Monte Carlo studies showed that the estimation of non-linear probability models can be sensitive to the degree of heteroskedasticity of the disturbances. The benchmark results were obtained under the following variance's prior $\pi(4/v_i) \sim \chi^2(4)$. Therefore, in order to check the robustness of the main results, I re-estimate the model by setting a different value for the hyperparameter *r*, either assuming the prior distribution is more asymmetric and skewed (*r* = 3) or the prior distribution is symmetric (*r* = 10). I also estimate the model under the assumption of no heteroskedasticity in the residuals' variance, i.e. $v_i = 1$, $i = 1,...,N$.

To save space, I only report in Table 4 the results of the heteroskedastic and asymmetric versus the symmetric and homoskedastic case. Overall, the results remain qualitatively similar. Higher or lower skewness parameter of the variance χ^2 distribution does not alter the findings. The decision to introduce an ecological label is still spatially dependent. However when *k* = 3, the

Table 4: MCMC Robustness Check Results

Variable	Spatial Probit with Exponential Distance Heteroskedastic andAsymmetric Case (k = 3) Coefficient	Direct Effect	Indirect Effect	Total Effect	Spatial Probit With Exponential Distance Symmetric and Homoskedastic Case Coefficient	Direct Effect	Indirect Effect	Total Effect
Constant	3.59				1.948			
	[9.174]				[4.858]			
Real GDP per Capita	0.198**	0.007**	0.001**	0.008**	0.15**	0.007**	0.001**	0.009**
	[0.104]	[0.004]	[0.001]	[0.004]	[0.069]	[0.003]	[0.001]	[0.004]
SO_2	-0.116**	-0.004**	-0.001*	-0.005**	-0.059**	-0.003**	-0.00001*	-0.003**
	[0.075]	[0.002]	[0.001]	[0.003]	[0.031]	[0.001]	[0.00001]	[0.002]
Political and Civil Rights	-0.219	-0.008	-0.001	-0.009	-0.267	-0.013	-0.002	-0.015
	[0.38]	[0.014]	[0.003]	[0.016]	[0.234]	[0.012]	[0.002]	[0.014]
Population Below 45	-0.174	-0.006	-0.001	-0.007	-0.144	-0.007	-0.001	-0.008
	[0.161]	[0.006]	[0.001]	[0.007]	[0.099]	[0.005]	[0.001]	[0.006]
Environmental Treaties	0.131***	0.004***	0.001**	0.005***	0.085***	0.004***	0.001**	0.005***
	[0.052]	[0.002]	[0.000]	[0.002]	[0.028]	[0.001]	[0.0001]	[0.001]
Manufacture Value Added	0.234***	0.008***	0.002**	0.009***	0.126***	0.006***	0.001**	0.007***
	[0.098]	[0.003]	[0.001]	[0.004]	[0.051]	[0.002]	[0.001]	[0.003]
High Technology Exports	0.111***	0.004***	0.001**	0.005***	0.073***	0.004***	0.001**	0.004***
	[0.051]	[0.002]	[0.001]	[0.002]	[0.031]	[0.001]	[0.000]	[0.002]
ISO14001	0.448***	0.015***	0.003**	0.018***	0.387***	0.019***	0.003**	0.022***
	[0.173]	[0.007]	[0.001]	[0.008]	[0.106]	[0.004]	[0.001]	[0.004]
Manufacturing Tariff	-0.114	-0.004	-0.001	-0.005	-0.089	-0.004	-0.001	-0.005
	[0.125]	[0.005]	[0.001]	[0.005]	[0.07]	[0.004]	[0.001]	[0.004]
Spatial Lag	0.165**				0.133**			
	[0.073]				[0.044]			
Observations (with eco-label)	141 (47)				141 (47)			
Total Draws / Omitted Draws	15000 /1000				15000 /1000			
Thinning Factor	4				5			
I-statistic	2.403				1.786			
Log-likelihood	-24.998				-14.188			
Pseudo R^2	0.721				0.842			

Note: The mean of the posterior distribution is reported as coefficient while the standard deviation is reported in the brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

results are more sensitive, because the algorithm breaks down when a smaller value for the hyperparameter r is considered (r = 3). In any case, most of the marginal estimates remain in the same range as in the benchmark, suggesting that the results are robust to an alteration of the MCMC's algorithm. This might be surprising, since a simple *t-test* rejects the hypothesis that each average estimated individual variance term ($\hat{v}_i$) is equal to 1. Although not reported, but available upon request, the results are also robust to a change in the initial value of the spatial autoregressive parameter and the covariance matrix of the prior distribution of β.

5.2 Spatial Weight Matrices

In order to shed light on the relative intensity of interdependence in the eco-label decision, I re-estimate the model using six alternative spatial weighting matrix $\mathbf{W}$. First, I consider the simple inverse bilateral distance which allocates a positive weight to all countries, including very remote ones. In fact, strategic dependence can be effective even beyond its own geographical region. The corresponding spatial weight matrix is defined as follows:

$$w_{ij} = \begin{cases} \frac{1}{distance_{ij}} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

A second spatial weight matrix is constructed based on the minimum political distance database built by Gleditsch & Ward (2001). The main reason to consider this measure is that the type of eco-labels considered in this study are in most cases the result of a political decision. In addition, empirical evidence suggests that political interests are a major determinant in the attitude towards preventing environmental damage. This political distance measure is build on the minimum geographical distances for all governments within 950 kilometers of each other. However, this measure is not simply geographical as it account for political ties and influences between governments in the year 2000. For instance, according to Gleditsch & Ward, there is no political influence between Canada and USA:

$$w_{ij} = \begin{cases} \frac{1}{\text{political distance}_{ij}} & \text{if } i \neq j \text{ and } \text{distance}_{ij} \leq 950\,\text{km} \\ 0 & \text{otherwise} \end{cases}$$

Third, I consider a spatial weight matrix based on the average bilateral trade between 1995-2000 to explicitly account for trade intensity between countries. Accordingly, strategic dependence should be higher among countries characterized by high trade intensity standardized by distance:

$$w_{ij} = \begin{cases} \frac{\text{trade}_{ij}}{\text{distance}_{ij}} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

However, trade intensity might be too broad as a measure of strategic interdependence, because despite important trade flows, countries can in reality share small strategic interdependence if each one is specialized in different type of products. That is why, following Cao & Prakash (2009), the last spatial weight matrices are constructed based on countries' export and import structural equivalence. The export (import) profile corresponds to the correlation between manufacturing exports (imports) of country *i* and *j* to the remaining economic partners at both bilateral and sector levels[50]. An export (import) structural equivalence$_{ij}$ close to 1 implies that country *i* and *j* export (import) the same type of goods to (from) the same partner countries. In other words, economy *i* and *j* are competitors (partners) since they export (import) similar products to (from) the same foreign markets. Therefore, one can expect strategic interactions to be stronger between countries in competition in export and import markets. I construct three spatial weight matrices based on exports and/or imports structural equivalence, respectively:

$$w_{ij} = \begin{cases} \text{structural equivalence}_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

Results given in Table 5 are qualitatively similar in terms of direct effect independently of the spatial weight considered. This is additional evidence of the robustness of the benchmark results. The differences in the marginal indirect effects (and thus total effects) are mainly due to the average value taken by the posterior distribution of the spatial autoregressive parameter. For instance, the spatial lag associated with the inverse geographical distance weight matrix is larger than any other spatial scheme. This suggests that spatial interactions in the eco-label decision are not bounded within their own geographical region. For instance, the Chinese government will not only consider the actions of its Asian neighbors, but also pay attention to European nations as well as the United States.

[50] Based on the United Nations' Standard International Trade Classification, four manufacturing sectors are considered: (1) chemical and related products; (2) manufactured goods; (3) machinery and transport equipment; (4) miscellaneous manufactured articles.

Table 5: Spatial Weight Robustness Check Results

Variable	Spatial Probit with Inverse Distance				Spatial Probit With Inverse Political Distance			
	Coefficient	Direct Effect	Indirect Effect	Total Effect	Coefficient	Direct Effect	Indirect Effect	Total Effect
Constant	12.81				1.809			
	[12.52]				[5.253]			
Real GDP per Capita	0.349***	0.007***	0.022***	0.029***	0.137**	0.006**	0.001**	0.007**
	[0.132]	[0.003]	[0.023]	[0.024]	[0.073]	[0.003]	[0.001]	[0.003]
SO_2	-0.209***	-0.004***	-0.015***	-0.02***	-0.06**	-0.003**	-0.001**	-0.003**
	[0.096]	[0.002]	[0.02]	[0.021]	[0.031]	[0.001]	[0.0001]	[0.002]
Political and Civil Rights	-0.491	-0.01	-0.014	-0.023	-0.357	-0.015	-0.003	-0.018
	[0.662]	[0.013]	[0.046]	[0.056]	[0.249]	[0.011]	[0.003]	[0.013]
Population Below 45	-0.431**	-0.008**	-0.028**	-0.036**	-0.168*	-0.007*	-0.001*	-0.009*
	[0.246]	[0.005]	[0.039]	[0.041]	[0.103]	[0.004]	[0.001]	[0.005]
Environmental Treaties	0.179***	0.004***	0.01***	0.014***	0.104***	0.004***	0.001***	0.005***
	[0.072]	[0.001]	[0.009]	[0.009]	[0.029]	[0.001]	[0.0001]	[0.001]
Manufacture Value Added	0.3***	0.006***	0.02***	0.026***	0.086*	0.004*	0.001*	0.004*
	[0.131]	[0.003]	[0.022]	[0.023]	[0.051]	[0.002]	[0.0001]	[0.002]
High Technology Exports	0.175***	0.004***	0.012***	0.016***	0.083***	0.003***	0.001***	0.004***
	[0.078]	[0.002]	[0.014]	[0.015]	[0.031]	[0.001]	[0.0001]	[0.002]
ISO14001	1.035***	0.02***	0.061***	0.081***	0.437***	0.018***	0.004***	0.022***
	[0.397]	[0.007]	[0.054]	[0.054]	[0.09]	[0.003]	[0.001]	[0.004]
Manufacturing Tariff	-0.135	-0.003	-0.003	-0.006	-0.062	-0.003	-0.0001	-0.003
	[0.233]	[0.005]	[0.014]	[0.017]	[0.071]	[0.003]	[0.001]	[0.004]
Spatial Lag	0.703***				0.18***			
	[0.129]				[0.057]			
Observations (with eco-label)	141 (47)				141 (47)			
Total Draws / Omitted Draws	15000 /1000				15000 /1000			
Thinning Factor	10				10			
I-statistic	1.991				1.991			
Log-likelihood	-14.033				-14.109			
Pseudo R^2	0.844				0.843			

Note: The mean of the posterior distribution is reported as coefficient while the standard deviation is reported in the brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 5: Spatial Weight Robustness Check Results (continued)

Variable	Spatial Probit with Trade Intensity				Spatial Probit With Export Structural Equivalence			
	Coefficient	Direct Effect	Indirect Effect	Total Effect	Coefficient	Direct Effect	Indirect Effect	Total Effect
Constant	-1.013				1.657			
	[4.406]				[7.402]			
Real GDP per Capita	0.097	0.004	0.001	0.005	0.23***	0.008***	0.004*	0.012***
	[0.066]	[0.003]	[0.001]	[0.003]	[0.095]	[0.003]	[0.004]	[0.006]
SO_2	-0.046*	-0.002*	-0.001*	-0.003*	-0.054	-0.002	-0.001	-0.003
	[0.029]	[0.001]	[0.001]	[0.002]	[0.044]	[0.001]	[0.001]	[0.002]
Political and Civil Rights	-0.257	-0.012	-0.002	-0.014	-0.408	-0.014	-0.007	-0.021
	[0.267]	[0.013]	[0.004]	[0.015]	[0.332]	[0.012]	[0.009]	[0.019]
Population Below 45	-0.128	-0.006	-0.001	-0.007	-0.246*	-0.009*	-0.004	-0.013*
	[0.098]	[0.005]	[0.002]	[0.006]	[0.144]	[0.005]	[0.005]	[0.009]
Environmental Treaties	0.108***	0.005***	0.001***	0.006***	0.14***	0.005***	0.002*	0.007***
	[0.031]	[0.001]	[0.001]	[0.002]	[0.047]	[0.001]	[0.002]	[0.003]
Manufacture Value Added	0.114**	0.005**	0.001**	0.007**	0.137*	0.005*	0.002	0.007*
	[0.054]	[0.002]	[0.002]	[0.004]	[0.076]	[0.003]	[0.003]	[0.005]
High Technology Exports	0.04	0.002	0.001	0.002	0.092**	0.003**	0.001*	0.005**
	[0.029]	[0.001]	[0.001]	[0.002]	[0.048]	[0.002]	[0.001]	[0.003]
ISO14001	0.253***	0.012***	0.002***	0.014***	0.418***	0.014***	0.006*	0.02***
	[0.108]	[0.005]	[0.001]	[0.004]	[0.135]	[0.005]	[0.004]	[0.005]
Manufacturing Tariff	-0.137**	-0.006**	-0.001**	-0.008**	-0.169	-0.006	-0.003	-0.008
	[0.07]	[0.003]	[0.001]	[0.004]	[0.115]	[0.004]	[0.003]	[0.006]
Spatial Lag	0.186***				0.281*			
	[0.114]				[0.155]			
Observations (with eco-label)	141 (47)				141 (47)			
Total Draws / Omitted Draws	15000 /1000				15000 /1000			
Thinning Factor	7				6			
I-statistic	1.787				1.628			
Log-likelihood	-34.690				-28.111			
Pseudo R^2	0.613				0.687			

Note: The mean of the posterior distribution is reported as coefficient while the standard deviation is reported in the brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 5: Spatial Weight Robustness Check Results (continued)

Variable	Spatial Probit with Import Structural Equivalence Coefficient	Direct Effect	Indirect Effect	Total Effect	Spatial Probit With Export + Import Structural Equivalence Coefficient	Direct Effect	Indirect Effect	Total Effect
Constant	0.674				1.28			
	[7.072]				[5.556]			
Real GDP per Capita	0.208***	0.007***	0.016**	0.023***	0.158***	0.007***	0.012**	0.019***
	[0.103]	[0.003]	[0.027]	[0.028]	[0.077]	[0.003]	[0.018]	[0.019]
SO_2	-0.075*	-0.003*	-0.006	-0.009*	-0.041	-0.002	-0.003	-0.005
	[0.047]	[0.002]	[0.014]	[0.015]	[0.03]	[0.001]	[0.006]	[0.007]
Political and Civil Rights	-0.898**	-0.03**	-0.072*	-0.102**	-0.416	-0.019	-0.029	-0.048
	[0.488]	[0.016]	[0.114]	[0.123]	[0.294]	[0.012]	[0.046]	[0.052]
Population Below 45	-0.301*	-0.01*	-0.025*	-0.036*	-0.203**	-0.009**	-0.016*	-0.025**
	[0.172]	[0.006]	[0.051]	[0.054]	[0.123]	[0.005]	[0.029]	[0.032]
Environmental Treaties	0.149***	0.005***	0.011**	0.016***	0.093***	0.004***	0.006***	0.011***
	[0.048]	[0.002]	[0.017]	[0.017]	[0.033]	[0.001]	[0.009]	[0.009]
Manufacture Value Added	0.167**	0.006**	0.013**	0.019**	0.113**	0.005**	0.008*	0.014**
	[0.082]	[0.003]	[0.022]	[0.023]	[0.063]	[0.003]	[0.014]	[0.015]
High Technology Exports	0.101***	0.003***	0.008**	0.012***	0.065**	0.003**	0.005**	0.008**
	[0.048]	[0.002]	[0.016]	[0.017]	[0.032]	[0.001]	[0.008]	[0.009]
ISO14001	0.563***	0.019***	0.04**	0.059***	0.309***	0.014***	0.019***	0.034***
	[0.163]	[0.006]	[0.057]	[0.057]	[0.11]	[0.004]	[0.023]	[0.022]
Manufacturing Tariff	-0.13	-0.004	-0.005	-0.009	-0.096	-0.004	-0.005	-0.01
	[0.154]	[0.005]	[0.017]	[0.02]	[0.078]	[0.004]	[0.009]	[0.011]
Spatial Lag	0.575**				0.514***			
	[0.219]				[0.188]			
Observations (with eco-label)	141 (47)				141 (47)			
Total Draws / Omitted Draws	15000 /1000				15000 /1000			
Thinning Factor	10				5			
I-statistic	1.669				1.481			
Log-likelihood	-18.531				-14.743			
Pseudo R^2	0.794				0.836			

Note: The mean of the posterior distribution is reported as coefficient while the standard deviation is reported in the brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

This is also confirmed when the spatial lag is based on export and/or import structural equivalence. In fact, strategic dependence goes beyond pure geographical and political distances and captures trade competition among structural equivalent economies. Two countries close geographically and trading intensively but whose manufacturing industries are structurally different will not necessarily share a strong strategic interaction in the decision to implement an eco-labelling scheme (e.g. one country trades food goods, while the other one trades manufacturing goods). This could partially explain why the spatial autoregressive associated with trade intensity is not large. The other reason might be due to the fact that the spatial weight matrix with bilateral trade is no longer strictly exogenous (but potentially predetermined). Yet the main findings remain almost unchanged, except that the substituability relationship between the average manufacturing tariff and the decision to introduce an eco-label is now statistically significant. This is in line with the view that eco-label might work as a non-tariff barriers. But since this finding is not robust, this suggests that the underlying protectionist motivation behind the implementation of an eco-labelling scheme is not as strong as least developed countries might fear.

5.3 Alternative Model Specifications

Due to data availability, it is extremely difficult to consider a large number of different potential explanatory variables without reducing drastically the sample size. By reducing the number of countries, not only is the notion of interdependence altered, but the MCMC algorithm will yield less efficient estimates. That is why, I restrict myself to estimate different versions of the benchmark model, including the spatial error and Durbin models, to account for potential additional (spatial) variables omission.

As suggested by Table 6, the results are robust to the inclusion of additional explanatory variables[51]. In particular, the level of the mortality rate, rule of law and corruption does not have a significant impact on the likelihood of eco-labelling. However, a high share of tertiary enrollment and population density, associated with a high stage of development and the existence of a potential green premium, increases the probability to implement an eco-label program. The same is true for the export orientation of the country proxied by manufacturing net trade. In other words, exporting economies are more likely to introduce an eco-labelling scheme.

[51] Several other specifications were estimated that included GDP per capita squared or a European Union dummy. Although those results lead to the same conclusion, they suffer from high collinearity and lack of convergence in the MCMC estimation.

Table 6: Additional Covariates Results

	Spatial Probit with Exponential Distance					
Variable	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Constant	4.468	5.552	3.909	0.367	-2.102	4.404
	[5.613]	[6.768]	[5.442]	[7.627]	[7.587]	[7.555]
Real GDP per Capita	0.127*	0.29***	0.21**	0.064	0.437***	0.272***
	[0.074]	[0.111]	[0.093]	[0.072]	[0.131]	[0.131]
SO_2	-0.063**	-0.116***	-0.078***	-0.131***	-0.135**	-0.138***
	[0.032]	[0.05]	[0.036]	[0.056]	[0.056]	[0.061]
Political and Civil Rights	-0.35	-0.303	-0.16	-0.385	-1.056***	-0.557
	[0.252]	[0.355]	[0.282]	[0.332]	[0.354]	[0.413]
Population Below 45	-0.18*	-0.242*	-0.165*	-0.067	-0.375***	-0.25
	[0.107]	[0.143]	[0.106]	[0.141]	[0.147]	[0.181]
Environmental Treaties	0.082***	0.138***	0.086***	0.087**	0.278***	0.148***
	[0.028]	[0.045]	[0.029]	[0.041]	[0.064]	[0.055]
Manufacture Value Added	0.108*	0.198***	0.143***	0.118**	0.369***	0.168***
	[0.055]	[0.067]	[0.057]	[0.06]	[0.095]	[0.075]
High Technology Exports	0.086***	0.102***	0.081***	0.168***	0.218***	0.13***
	[0.03]	[0.04]	[0.031]	[0.072]	[0.061]	[0.056]
ISO14001	0.426***	0.594***	0.407***	0.586***	0.707***	0.531***
	[0.095]	[0.184]	[0.108]	[0.2]	[0.141]	[0.215]
Manufacturing Tariff	-0.072	-0.107	-0.08	-0.045	-0.269***	-0.117
	[0.074]	[0.089]	[0.075]	[0.105]	[0.088]	[0.09]
Mortality Rate	-0.025					
	[0.021]					
Rule of Law		-1.093				
		[0.755]				
Corruption Index			-0.576			
			[0.640]			
Tertiary Enrollment				0.069***		
				[0.029]		
Population Density					0.113***	
					[0.032]	
Manufacturing Net Trade						0.084***
						[0.03]
Spatial Lag	0.122**	0.152***	0.144***	0.109**	0.192***	0.148***
	[0.043]	[0.032]	[0.041]	[0.038]	[0.03]	[0.041]
Observations (with eco-label)	141 (47)	141 (47)	141 (47)	141 (47)	141 (47)	141 (47)
Total Draws / Omitted Draws	15000/1000	15000/1000	15000/1000	15000/1000	15000/1000	15000/1000
Thinning Factor	5 / 1.399	10 / 1.577	10 / 1.410	15 / 1.440	15 / 1.962	10 / 1.876
I-statistic	-13.930	-13.630	-13.859	-11.503	-9.851	-10.404
Log-likelihood	0.845	0.848	0.846	0.872	0.890	0.884
Pseudo R^2	4.468	5.552	3.909	0.367	-2.102	4.404

Note: The mean of the posterior distribution is reported as coefficient while the standard deviation is reported in the brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Beside the spatial lag, spatial dependence can also be specified in the error term. The spatial error model whose errors are spatially correlated reads as:

$$\mathbf{Y}^* = \mathbf{X}\beta + \mathbf{e}$$
$$\mathbf{e} = \lambda\mathbf{W}\mathbf{e} + \mathbf{u}$$

Estimation of the spatially autocorrelated error model through the MCMC method is very similar to the spatial lag model, because the prior of each parameter is defined independently of each other. The marginal effects in a spatial error model are computed as follows:

$$\hat{\Xi}_{ijr} = \phi\left([X\hat{\beta}]_i / \hat{t}_i\right)\hat{\beta}_r / \hat{t}_i$$

where $\hat{t}_i = \hat{\sigma}^2 \Sigma_i \hat{\varpi}_{ij}^2$ with ω_{ij} being the ij^{th} elements of the matrix $(I - \lambda\mathbf{W})^{-1}\mathbf{u}$.

Ignoring spatial error dependence yields an omitted variable bias, when the omitted variables are spatially dependent. The issue is that the spatial lag might capture uncorrected spatial dependence related to the error term. Despite its econometric foundation, there is no direct economic interpretation of the expected sign of the spatial error term. As highlighted by Pace and LeSage (2007), the spatial Durbin model has the advantage of reducing spatially dependent omitted variable bias and having a direct economic interpretation. The spatial Durbin model corresponds to the extension of the benchmark model which includes the spatially weighted average of the dependent variable ($\mathbf{WY}^*$) as well as the explanatory variables beside the constant term ($\mathbf{WX}^{nc}$):

$$\mathbf{Y}^* = \rho\mathbf{W}\mathbf{Y}^* + \mathbf{X}\beta + W\mathbf{X}^{nc}\boldsymbol{\gamma} + \mathbf{u}$$

Note that in the spatial Durbin model, the spatial effects are modified to account for the presence of the spatial covariates:

$$\hat{\Xi}_{ijr} = \phi\left([(I_N - \hat{\rho}\mathrm{W})^{-1}(X\hat{\beta} + \mathrm{WX}^{\mathrm{nc}}\hat{\gamma})]_i / \hat{s}_i\right)\left([(I_N - \hat{\rho}\mathrm{W})^{-1}]_{ij}\hat{\beta}_r + [(I_N - \hat{\rho}\mathrm{W})^{-1}\mathrm{W}]_{ij}\hat{\gamma}_r / \hat{s}_i\right)$$

One important drawback of the spatial Durbin model is the introduction of additional collinearity as it is the case here. As a consequence, there are some convergence issues. That is why, some of the MCMC settings are modified. In particular, the number of burn-in and draws are set to 20,000 and 100,000, respectively, because the posterior distribution of the spatial lag is not converging with a smaller number of draws.

Table 7: Spatial Error and Durbin Results

Variable	Spatial Error Probit		Spatial Durbin Probit with Exponential Distance				
	Coefficient	Marginal Effect	Coefficient	Spatial Coefficient	Direct Effect	Indirect Effect	Total Effect
Constant	7.254		-0.477				
	[6.972]		[9.358]				
Real GDP per Capita	0.396***	1.077***	0.761***	-0.409	0.004***	-0.003	0.001
	[0.16]	[0.58]	[0.295]	[0.555]	[0.003]	[0.003]	[0.003]
SO_2	-0.159**	-0.406**	-0.449***	-1.326***	-0.002**	-0.004***	-0.006***
	[0.079]	[0.205]	[0.177]	[0.328]	[0.001]	[0.003]	[0.004]
Political and Civil Rights	0.854	2.218	4.311***	-15.61***	0.024***	-0.061***	-0.037***
	[0.592]	[1.605]	[1.713]	[4.803]	[0.018]	[0.044]	[0.03]
Population Below 45	-0.333*	-0.898*	-0.245	-0.551	-0.001	-0.001	-0.003
	[0.188]	[0.617]	[0.471]	[1.019]	[0.003]	[0.004]	[0.005]
Environmental Treaties	0.236***	0.622***	0.515***	1.434***	0.002**	0.004***	0.006***
	[0.072]	[0.223]	[0.182]	[0.439]	[0.002]	[0.003]	[0.005]
Manufacture Value Added	0.274***	0.725***	0.989***	2.202**	0.004**	0.006**	0.011***
	[0.112]	[0.332]	[0.335]	[0.824]	[0.003]	[0.005]	[0.008]
High Technology Exports	0.139**	0.359**	0.417***	1.387**	0.002**	0.004*	0.006***
	[0.064]	[0.173]	[0.141]	[0.629]	[0.001]	[0.004]	[0.005]
ISO14001	0.482***	1.275***	1.415***	0.071	0.007***	-0.002***	0.005***
	[0.136]	[0.452]	[0.265]	[0.451]	[0.005]	[0.002]	[0.004]
Manufacturing Tariff	0.091	0.249	1.269***	-2.431	0.007***	-0.01*	-0.004
	[0.126]	[0.363]	[0.338]	[1.478]	[0.005]	[0.009]	[0.006]
Spatial Error	0.958***						
	[0.023]						
Spatial Lag			-0.412**				
			[0.173]				
Observations (with eco-label)	141 (47)		141 (47)				
Total Draws / Omitted Draws	15000 /1000		100000 /20000				
Thinning Factor	5		5				
I-statistic	1.251		1.126				
Log-likelihood	-22.571		-0.313				
Pseudo R^2	0.749		0.997				

Note: The mean of the posterior distribution is reported as coefficient while the standard deviation is reported in the brackets. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 7 reports the results of the spatial error and Durbin probit models with the negative exponential distance weighting scheme. In comparison with the spatial autoregressive model, the spatially autocorrelated error model yields higher posterior mean for each parameter, including the spatial error variable $\hat{\lambda} = 0.95$. As a consequence, the marginal effects are larger than the direct effects in the benchmark model. Since the errors seems to be spatially correlated, it might be justified to consider the spatial Durbin model to control for the potential omitted variable bias in the spatial error. As mentioned previously, the spatial Durbin model suffers from multicollinearity, which explains why the posterior mean of most variables is higher (in particular the political and civil liberties index). Therefore these results should be analyzed cautiously. Evidence suggests that the average government will not only look if its partners have implemented an eco-labelling scheme, but also consider their economic characteristics. That is why the indirect marginal effects are larger than direct ones in absolute value for most explanatory variables. Finally, accounting for spatial dependence in the error term yields a negative and significant spatial autoregressive coefficient. This surprising finding is in line with the view that eco-labelling might be subject to free-riding.

6 Conclusion

The decision of a government to introduce an eco-labelling scheme program depends on many factors. One potential determinant, usually omitted in empirical literature, is the existence of interdependence in the adoption's decision. To address this issue, a limited dependent variable spatial probit model is estimated through a Markov Chain Monte Carlo method using a new joint update sampler algorithm. In comparison with an aspatial specification, accounting for the spatial dependence leads to a higher explanatory power of the model, with type-I and type-II errors decreasing from 9.2% to 5.7% of sample observations. The main findings indicate that the probability for a government to introduce an eco-labelling scheme is positively related to the economy's stage of development decision, the existence of potential scale effects as well as a relative production cost advantage through innovation. In addition, there is robust evidence of positive strategic dependence in the eco-labelling decision. In other words, a country has a higher probability of adopting an eco-labelling scheme if its immediate neighbours do so.

These empirical findings help to understand the reasons behind the asymmetric diffusion of eco-labelling schemes between developed and developing economies. As the latter confront harsher socio-economic problems, they lack incentive to implement eco-label programs and tend to behave as "standard-taker" rather than "standard-setter". This is further exacerbated by the positive strategic interaction prevailing among partner countries. This leads to a world where eco-labels tend to cluster into high-income regions, distorting trade flows and reducing world welfare.

In order to promote eco-labelling among low income countries, eco-label programs already in place should be revised to be as transparent and unbiased as possible and thus avoid any trade barrier effects. Harmonization in the standards, greater transparency in the certification awarding and mutual recognition in eco-labelling schemes are also necessary in order to reduce the fixed costs associated with eco-labelling adoption. Such changes will not be achieved easily, as they involve not only national governments but also the implicated firms and agencies in both developed and developing countries.

Although the above-mentioned results have been proved to be reasonably robust, there is ample scope for further improvements. In particular, future research should account for the time dynamic in the eco-label program implementation and thus consider a dynamic extension of the spatial probit model for several reasons. First, the starting phase of most eco-labelling schemes is associated with procedural and methodological uncertainties that could be even more important and lasting in developing countries. Second, most current eco-labels in developed countries are able to exist because of subsidies. A slowed economy or a change in environmental policy could limit the sources of subsidies (e.g. individual supports, foundations, governments) affecting the functioning of eco-label programs. Third, the important consolidation and vertical integration within and between most segments of the market chains, that have been taken place during the past 30 years, can also alter the industry market structure and affect the eco-labelling schemes. Last but not least, the potential oversupply of eco-labels may prevent consumption of environmentally friendly products because of information congestion. This could ultimately lead to the elimination of some eco-label programs. Hopefully, with the development of better disaggregated data on eco-labelled trade flows, additional investigation will help to disentangle the different effects that are at play between eco-labelling and international trade.

7 Appendices

7.1 Relative Algorithms Performance

Performance relative to Joint Update with Metropolis-Hastings

	Iterative Update with Geweke and Metropolis-Hastings		Iterative Update with Rodriguez-Yam et al. and Metropolis-Hastings	
N	Dependence Factor	Parameter Distance	Dependence Factor	Parameter Distance
250	1.102	0.659	1.102	0.659
500	1.287	0.724	1.287	0.724
750	1.397	0.797	1.397	0.797
1000	1.569	0.869	1.569	0.869

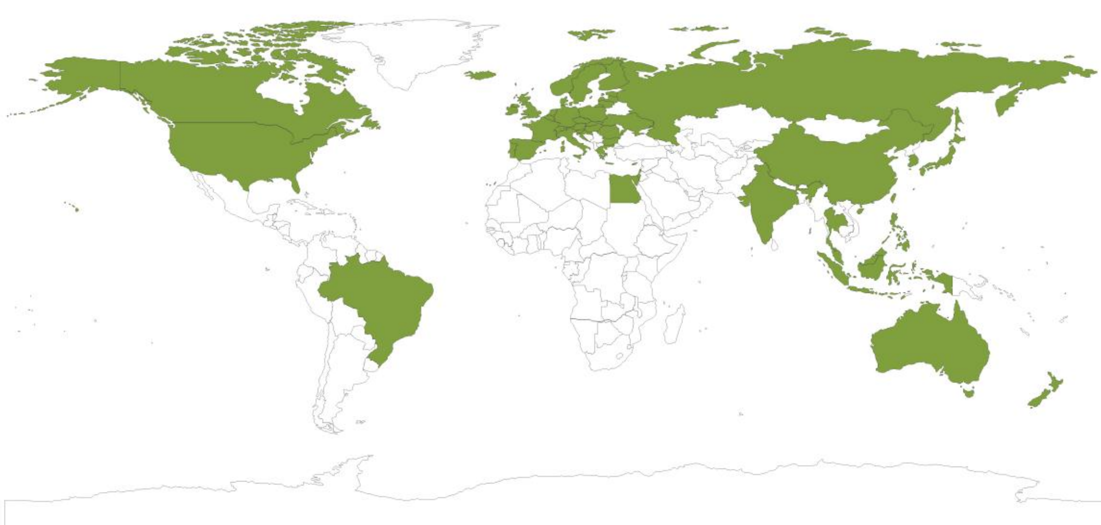
Note: Performance standardized with respect to CPU.

Performance relative to Joint Update with Metropolis-Hastings

	Iterative Update with Geweke and Numerical Integration		Iterative Update with Rodriguez-Yam et al. and Numerical Integration	
N	Dependence Factor	Parameter Distance	Dependence Factor	Parameter Distance
250	1.256	0.684	1.234	0.950
500	1.274	0.745	1.233	0.976
750	1.455	0.818	1.276	1.022
1000	1.469	0.879	1.290	1.064

Note: Performance standardized with respect to CPU.

7.2 Manufactured Multi-products Eco-Label Map of the World



7.3 Manufactured Multi-products Eco-Label List

Country	National Eco-label	Supranational Eco-label	Country	National Eco-label	Supranational Eco-label
Australia	1991		Korea, Rep.	1992	
Austria	1990	1992*	Latvia		2004*
Belgium		1992*	Liechtenstein		1992*
Brazil	1992		Lithuania	2001	2004*
Bulgaria		2007*	Luxembourg	1992	1992*
Canada	1988		Malaysia	1996	
China	1993		Malta		2004*
Croatia	1993		Netherlands	1992	1992*
Cyprus		2004*	New Zealand	1990	
Czech Republic	1993	2004*	Norway		1989◊, 1992*
Denmark		1989◊, 1992*	Philippines	2001	
Egypt, Arab Rep.	1999		Poland	2004	2004*
Estonia		2004*	Portugal		1992*
Finland		1989◊, 1992*	Romania		2007*
France	1992	1992*	Russian Federation	2007	
Germany	1977	1992*	Singapore	1992	
Greece		1992*	Slovak Republic	1996	1996*
Hong Kong, China	2000		Slovenia		2004*
Hungary	1994	1994*	Spain	1994	1992*
Iceland		1989◊, 1992*	Sweden	1989	1989◊, 1992*
India	1991		Switzerland	2000	
Indonesia	1994		Taiwan	1992	
Ireland		1992*	Thailand	1994	
Israel	1993		Ukraine	2002	
Italy		1992*	United Kingdom		1992*
Japan	1989		United States	1988	

Notes: ◊ Nordic Swan; * EU Eco-labelling. Sources: Global Eco-labelling Network, New Zealand's Ministry of Economic Development

7.4 Country List

Albania; Algeria; Angola; Argentina; Armenia; Australia; Austria; Azerbaijan; Bangladesh; Belarus; Belgium-Luxembourg; Belize; Benin; Bolivia; Bosnia and Herzegovina; Botswana; Brazil; Bulgaria; Burkina Faso; Burundi; Cambodia; Cameroon; Canada; Central African Republic; Chad; Chile; China; Colombia; Congo, Dem. Rep.; Congo, Rep.; Costa Rica; Croatia; Cyprus; Czech Republic; Côte d'Ivoire; Denmark; Djibouti; Dominican Republic; Ecuador; Egypt, Arab Rep.; El Salvador; Estonia; Ethiopia; Finland; France; Gabon; Georgia; Germany; Ghana; Greece; Guatemala; Guinea; Guinea-Bissau; Guyana; Honduras; Hong Kong, China; Hungary; Iceland; India; Indonesia; Iran, Islamic Rep.; Ireland; Israel; Italy; Jamaica; Japan; Jordan; Kazakhstan; Kenya; Korea, Rep.; Kuwait; Kyrgyz Republic; Lao PDR; Latvia; Lebanon; Lithuania; Macedonia, FYR; Madagascar; Malawi; Malaysia; Mali; Mauritania; Mauritius; Mexico; Moldova;

Mongolia; Morocco; Mozambique; Namibia; Nepal; Netherlands; New Zealand; Nicaragua; Niger; Nigeria; Norway; Oman; Pakistan; Panama; Papua New Guinea; Paraguay; Peru; Philippines; Poland; Portugal; Romania; Russian Federation; Rwanda; Saudi Arabia; Senegal; Sierra Leone; Slovak Republic; Slovenia; South Africa; Spain; Sri Lanka; Sudan; Swaziland; Sweden; Switzerland; Syrian Arab Republic; Tajikistan; Tanzania; Thailand; Togo; Trinidad and Tobago; Tunisia; Turkey; Turkmenistan; Uganda; Ukraine; United Arab Emirates; United Kingdom; United States; Uruguay; Uzbekistan; Venezuela, RB; Vietnam; Yemen, Rep.; Zambia; Zimbabwe.

7.5 Data Sources

Factors	Variable	Expected sign	Source
Stage of development	Real GDP Per Capita	+	WDI 2007
Economic Efficiency	Civil And Political Liberties Index	+	Freedom House
Pollution Pressure	SO_2 Emission level	+	EDGAR
Scale Effect	Manufacture Value Added	+	WDI 2007
Price Premium	Population Below 45 Years Old	+	U.S. Census Bureau
Environmental Experience	International Environmental Treaties	+	ENTRI
Innovation	High Technology Exports	+	WDI 2007
Innovation	ISO14001 Certificates	+	ISO
Non-Tariff Barriers	Manufacturing Tariff	-/+	WDI 2007
Spatial Dependence	Geographical Distance	-/+	CEPII
Spatial Dependence	Political Distance	-/+	Gleditsch Ward (2001)
Spatial Dependence	Bilateral Trade Flows/Structural Equivalence	-/+	UN Comtrade

Note: Non spatial data is averaged over 2003-2005, except SO_2 emission data which is only available for 2000.

7.6 Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max	Moran I
Eco-label	141	0.33	0.47	0	1	0.57***
Real GDP per capita	141	8967.93	13800.01	38.76	62812.54	0.54***
SO_2 emissions	141	91.78	13.29	0	99.8	0.2***
Political and civil liberties	141	-3.37	1.85	-7	-1	0.48***
Population share below 45 years old	141	41.09	4.88	31.17	65.05	0.48***
Environmental treaties	141	41.35	20.45	4	104	0.68***
Manufacture value-added	141	15.19	7.57	2.3	42	0.34***
High-technology exports	141	0.09	0.12	0	0.71	0.16***
ISO14001 certificates	141	775.67	2565.73	0	21881	0.26***
Manufacturing tariff	141	8.63	5.57	0	31.85	0.44***

Notes: The Moran statistics tests the absence of spatial autocorrelation. *, ** and *** denotes significance at 10%, 5% and 1%, respectively.

7.7 Correlation Matrix

	Eco-label	real GDP per capita	SO_2 emission	Political liberties	Population below 45	Env.Treaties	Manuf. value added	Technology exports	ISO certificates	Manuf. tariff
Eco-label	1									
Real GDP per capita	0.65***	1								
SO_2 emission	-0.43***	-0.42***	1							
Political and civil liberties	0.58***	0.55***	-0.32***	1						
Pop. below 45	0.42***	0.4***	-0.46***	0.27***	1					
Environmental treaties	0.71***	0.64***	-0.37***	0.62***	0.37***	1				
Manufacture value-added	0.44***	0.13	-0.16*	0.21**	0.39***	0.35***	1			
High-technology exports	0.48***	0.33***	-0.22***	0.31***	0.23***	0.25***	0.28***	1		
ISO14001 certificates	0.4***	0.34***	-0.21**	0.19**	0.18**	0.35***	0.26***	0.26***	1	
Manufacturing tariff	-0.53***	-0.56***	0.33***	-0.57***	-0.45***	-0.51***	-0.3***	-0.3***	-0.21**	1

Note: *, ** and *** denotes significance at 10%, 5% and 1%, respectively.

Conclusion

This final chapter brings the thesis to a conclusion by reporting the main econometric and empirical findings as well as policy implications from the previous chapters. In addition, potential extensions of this work for further research are discussed.

Findings Overview

This PhD dissertation, composed of four chapters, provides an attempt to investigate and disentangle the relationship between foreign direct investment and environmental regulation in an integrated world. From a methodological point of view, this work proposes several innovations as new and more efficient estimators are implemented in various spatial frameworks. The general conclusion drawn from this thesis lies in the importance of accounting and correcting for spatial dependence properly in order to assess the relationship between international trade, pollution and environmental policies. In particular, I found robust evidence of spatial dependence in the (strategic) allocation of FDI as well as environmental policies. Overall, these findings suggest to further enhance the collaboration between governments in order to mitigate negative and capture positive spatial spillovers.

The first chapter, entitled "Spatial Dynamic Panel and System GMM: a Monte-Carlo Investigation", investigates the finite sample properties of estimators for spatial dynamic panel models in the presence of several endogenous variables. The simulation results suggest that in order to account for the endogeneity of several covariates, one should apply the system generalized moments method estimator instead of spatial MLE, spatial dynamic MLE or spatial dynamic QMLE. In addition, system GMM is computationally less demanding and more easier to adapt to unbalanced panel data. Yet, when the sample size is relatively small and the spatial lag is the main variable of interest, spatial dynamic models should be estimated with traditional spatial estimators.

In the second chapter, "Pollution Havens: a Spatial Vector Autoregression Approach", I investigate the mutual relationship between FDI, pollution and trade openness for 124 countries during the nineties. After applying second generation panel unit root tests, I estimate a trivariate spatial VAR model where a shock occurring in a given country can potentially affect the economic conditions in the neighboring countries. Based on generalized empirical likelihood estimations, the results highlight a reverse causality between FDI and pollution emission, suggesting the existence of pollution haven and pollution halo effects in low income countries in particular. In addition, spatial spillovers play a key role in the linkage between FDI, trade openness and SO_2 emissions and highlight the strategic nature of these variables.

In line with the previous chapter, the third study, entitled "Complex FDI and Environmental Regulation", investigates if differences in environmental regulations did influence OECD's FDI allocation during 1981-2000 taking into account third-country effects in a multi-country setting. The findings, based on system GMM estimates of a spatial dynamic gravity model, confirm the existence of a negative relationship between FDI and environmental stringency, once endogeneity and spatial dependence are corrected. In addition, the evidence of positive third-country effects for FDI suggests the prevalence of highly complex vertically integrated FDI from OECD countries to developing economies. Multinationals tend to allocate each part of the production process, including the most polluting one, to host countries, which ensure cost minimization.

In the last chapter called "Unequal Spatial Diffusion of Eco-labels", the decision to introduce an eco-labelling scheme is analyzed through a heteroskedastic Bayesian spatial probit. This framework allows the government's decision of introducing an eco-label program to be influenced by the decision of the neighbouring countries. I propose to estimate this model by extending the joint updating approach developed by Holmes & Held (2006) to a spatial framework. Based on a sample, including 141 developing and developed economies, I show that a country’s probability to introduce an eco-labelling scheme depends among other things on the eco-label programs adopted by countries which are spatially close or sharing a strong trade intensity relationship with each other. These findings explain why developing countries, usually standards takers, have been naturally put at a disadvantage in terms of eco-labelling adoption.

Policy Implications

This thesis offers a global framework supporting the view that pollution and environmental policies might be affected by strategic trade and investment considerations. However, simply stating that environmental policies have a negative impact on a country’s competitiveness would be an oversimplification. From a policy point of view, environmental regulation should be designed with enough flexibility (e.g. market based instruments), whenever possible (e.g. local pollutant), to minimize compliance costs and enhance environmental preservation. In order to make the shift as smooth as possible, adjustment programs between high income, emerging and low income countries seem necessary. The global or regional dimension of some environmental issues (e.g. transboundary pollution) should also rely on the joint active cooperation between developed and developing economies. In particular, efforts should be made at harmonizing environmental policies among countries, not on a one-to-one mechanism, but by acknowledging any relevant differences among economies such as the level of development and the pollution absorption capacity. Overall, it is important to create conditions capable of generating additional resources targeted towards environmental quality.

Further Research

Although the different empirical findings of this thesis provide additional insights on the reasons behind inconclusive pollution haven effect evidence in the empirical literature, further analysis should be done in order to enhance the formulation of optimal policy and direct the priority of policy makers.

Most of the empirical analysis in this thesis has been carried out using aggregate level data. Although panel data approach can account for cross-country heterogeneity, it might sometimes suffer from measurement errors. An analysis at a more disaggregated level would allow to elaborate more relevant policy recommendations, since some sectors are expected to be more sensitive to environmental regulation. Extending the period of analysis would also be important as the 2000's have seen the emergence of new poles of growth (e.g. Brazil, China, India,...). In particular, it would be interesting to investigate the relationship between environmental regulation and capital allocation among these emerging economies and developing countries (e.g. Africa). But most importantly, a better dynamic measure of environmental stringency at the cross-country level would be able to alleviate some of the ambiguity associated with the use of output pollution measures. Coupled with a more theoretical foundations of the spatial effects, this would allow for a better rationale to interpret the results from a political economy perspective for instance.

Overall, additional theoretical and empirical research on the impact of trade and capital liberalization on the level of environmental stringency is required, before a satisfactory conclusion on this topic is reached.

References

Chapter 1: Spatial Dynamic Panel and System GMM

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

Anselin, L., 2001. "Spatial Econometrics," in *A Companion to Theoretical Econometrics*. B.H. Baltagi (Ed.), Blackwell Publishers Lte., Massachusetts.

Anselin, L., 2003. "Spatial Externalities, Spatial Multipliers and Spatial Econometrics," *International Regional Science Review*, 26(2), pp. 153--166.

Anselin, L., 2006. "Spatial Econometrics," in *Palgrave Handbook of Econometrics*: *Volume* 1, *Econometric Theory*. T.C. Mills, K. Patterson (Eds.), Palgrave MacMillan, Basingstoke.

Arellano, M., and S. Bond, 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58(2), pp. 277-297.

Azomahou, T. 2008. "Minimum Distance Estimation of the Spatial Panel Autoregressive Model," Cliometrica, *Journal of Historical Economics and Econometric History*, 2(1), pp. 49-83.

Badinger, H., W. Müller and G. Tondl, 2004. "Regional Convergence in the European Union, 1985- 1999: A Spatial Dynamic Panel Analysis," *Regional Studies*, 38(3), pp. 241-253.

Bhargava, A., and J. D. Sargan, 1983. "Estimating Dynamics Random Effects Models from Panel Data Covering Short Time Periods," *Econometrica*, 51(6), pp. 1635-1659.

Beck, T., and R. Levine, 2004. "Stock Markets, Banks, and Growth: Panel Evidence," *Journal of Banking and Finance*, 28(3), pp. 423-442.

Beenstock M. and D. Felsenstein, 2007. "Spatial Vector Autoregressions," *Spatial Economic Analysis*, 2(2), pp. 167-196.

Binder, M., C. Hsiao and M H. Pesaran, 2005. "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," *Econometric Theory*, 21(4), pp. 795-837.

Blundell, R., and S. Bond, 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87(1), pp. 115-143.

Blundell, R., S. Bond, and F. Windmeijer, 2000. "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator," in *Advances in Econometrics* (*Vol*. 15) *Nonstationary Panels, Panel Cointegration and Dynamic Panels*, B.H. Baltagi (Ed.), Elsevier Science.

Bun, M. J. G., and M. A . Carree, 2006. "Bias-Corrected Estimation in Dynamic Panel Data Models with Heteroskedasticity," *Economics Letters*, 92(2), pp. 220-227.

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

Dall'erba S. and J. Le Gallo, 2008. "Regional Convergence and the Impact of European Structural Funds Over 1989--1999: A Spatial Econometric Analysis," *Papers in Regional Science*, 87(2), pp. 219--244.

Das, D., H. H. Kelejian and I. R. Prucha, 2003. "Finite Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances," *Journal of Economics*, 82(1), pp. 1-26.

Driscoll, J. and A. Kraay, 1998. "Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data," *The Review of Economics and Statistics*, 80(4), pp. 549-560

Elhorst, J. P., 2003. "Specification and Estimation of Spatial Panel Data Models," *International Regional Science Review*, 26(3), pp. 244-268.

Elhorst, J. P., 2005. "Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels," *Geographical Analysis*, 37(1), pp. 85-106.

Elhorst J.P., 2008. "Estimation of Dynamic Panels with Endogenous Interaction Effects," presented at the second World Conference of the Spatial Econometrics Association.

Elhorst J.P., 2009, "Spatial Panel Data Models," in *Handbook of Applied Spatial Analysis*. M.M. Fischer, A. Getis (Eds.). Springer Berlin Heidelberg New York.

Fingleton, B. and J. Le Gallo, 2008. "Estimating Spatial Models with Endogenous Variables, a Spatial Lag and Spatially Dependent Disturbances: Finite Sample Properties," *Papers in Regional Science*, 87(8), pp. 319-339.

Foucault, M., T. Madies and S. Paty, 2008. "Public Spending Interactions and Local Politics. Empirical Evidence From French Municipalities," *Public Choice*, 137(1), pp. 57-80.

Franzese, R. and J. C. Hays, 2007. "Spatial-Econometric Models of Cross-Sectional Interdependence in Political-Science Panel and Time-Series-Cross-Section Data," Political Analysis, 15(2), pp. 140--164.

Haining, R.,1978. "Estimating Spatial-Interaction Models," *Environment and Planning A*, 10(3), pp. 305--320.

Hansen, L.P., J. Heaton and A. Yaron, 1996. "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), pp. 262-280.

Hayakawa, K., 2007. "Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models," *Economics Letters*, 95(1), pp. 32-38.

Hong, E., L. Sun, and T. Li, 2008. "Location of Foreign Direct Investment in China: A Spatial Dynamic Panel Data Analysis by Country of Origin," Discussion Paper 86, Department of Financial and Management Studies, University of London.

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 2002. "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109(1), pp. 107-150.

Kapoor M., H.H. Kelejian and I.R. Prucha, 2007. "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics*, 140(1), pp. 97--130.

Kelejian, H.H. and I.R. Prucha, 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17(1), pp. 99--121.

Kelejian, H.H. and I.R. Prucha, 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40(2), pp. 509--533.

Kelejian, H., and I. R. Prucha, 2007. "HAC Estimation in a Spatial Framework," *Journal of Econometrics*, 140(1), pp. 131-154.

Korniotis, G. M., 2010. "Estimating Panel Models with Internal and External Habit Formation," *Journal of Business and Economic Statistics*, 28(1), pp. 145-158.

Kukenova, M. and J.-A. Monteiro, 2009. "Does Lax Environmental Regulation Attract FDI when accounting for "third-country" effects?."

Lee, L., 2003. "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances," Econometric Reviews, 22(4), pp. 307-335.

Lee, L. and J. Yu, 2007. "A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects," University of Rochester.

Lee L. and J. Ju, 2009. "Some Recent Developments in Spatial Panel Data Models," University of Kentucky.

Madariaga, N. and S. Poncet, 2007. "FDI in Chinese Cities: Spillovers and Impact on Growth," *The World Economy*, 30(5), pp. 837-862.

Okui, R., 2009. "The Optimal Choice of Moments in Dynamic Panel Data Models," *Journal of Econometrics*, 151(1), pp. 1-16.

Phillips, P. and H. Moon, 2000. "Nonstationary Panel Data Analysis: an Overview of some Recent Developments," *Econometric Reviews*, 19(3), pp. 263-286.

Roodman, D., 2009. "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics*, 71(1), pp. 135-158.

Yu, J. and L. Lee, 2009. "Some Recent Developments in Spatial Panel Data Models," University of Kentucky.

Yu, J., R. de Jong and L. 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.

Yu, J., R. de Jong and L. Lee, 2007. "Quasi-Maximum Likelihood Estimators For Spatial Dynamic Panel Data With Fixed Effects When Both N and T Are Large: A Nonstationary Case," University of Rochester.

Chapter 2: Pollution Havens: a Spatial Vector Autoregression Approach

Alessi, L., M. Barigozzi, and M. Capasso, 2008. "A robust criterion for determining the number of static factors in approximate factor models," Working Paper Series 903, European Central Bank.

Arellano, M., and S. Bond, 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58(2), pp. 277-97.

Arellano, M., and O. Bover, 1995. "Another Look at the Instrumental Variables Estimation of Error Components Models," *Journal of Econometrics*, 68(1), pp. 29-51.

Bai J., and S. Ng, 2002. "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), pp. 191-221.

Bai J., and S. Ng, 2004. "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), pp. 1127-1178.

Beenstock, M. and D. Felsenstein, 2007. "Spatial Vector Autoregressions," *Spatial Economic Analysis*, 2, pp. 167-196.

Binder, M., C. Hsiao and M. H. Pesaran, 2005. "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," *Econometric Theory*, 21(4), pp. 795-837.

Blonigen, B. A., R. B. Davies, H. T. Naughton, and G. R. Waddell, 2007. "FDI in space: Spatial Autoregressive Relationships in Foreign Direct Investment," *European Economic Review*, 51(5), pp. 1303-1325.

Blundell, R., and S. Bond, 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87(1), pp. 115-143.

Breitung, J., and S. Das. 2005. "Panel Unit Root Tests Under Cross-sectional Dependence," *Statistica Neerlandica,* 59, pp. 414-433.

Chang, Y., 2002. "Nonlinear IV Unit Root Tests in Panels with Cross--Sectional Dependency," *Journal of Econometrics*, 110(2), pp. 261-292.

Chang, Y., 2004. "Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency," *Journal of Econometrics*, 120(2), pp. 263-293.

Chen, X., and T.G. Conley, 2001. "A New Semiparametric Spatial Model for Panel Time Series," *Journal of Econometrics*, 105(1), pp. 59--83.

Choi I., 2001. "Unit Root Tests for Panel Data," *Journal of International Money and Finance*, 20(2), pp. 249-272.

Choi I., 2002. "Combination Unit Root Tests for Cross-Sectionally Correlated Panels", Mimeo, Hong Kong University of Science and Technology.

Cole, M.A., R.J.R. Elliott and P.G. Fredriksson, 2006. "Endogenous Pollution Havens: Does FDI Influence Environmental Regulations?," *Scandinavian Journal of Economics*, 108(1), pp. 157-78.

Cole, M.A., and P.G. Fredriksson, 2009. "Institutionalized Pollution Havens," *Ecological Economics*, 68(4), pp. 1239-1256.

Copeland, B. R., and M. S. Taylor, 2003. "Trade and the Environment: Theory and Evidence," Princeton University Press.

Di Giacinto, V., 2003. "Differential Regional Effects of Monetary Policy: A Geographical SVAR Approach", *International Regional Science Review*, 26 (3), pp. 313-341.

Dickey, D. A. and W. A. Fuller, 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49(4), pp. 1057-1072.

Demetrescu, M., U. Hassler and A.I. Tarcolea, 2006. "Combining Significance of Correlated Statistics with Application to Panel Data," *Oxford Bulletin of Economics and Statistics*, 68 (5), pp. 647-663.

Ederington, J., and J. Minier, 2003. "Is Environmental Policy a Secondary Trade Barrier? An Empirical Analysis," *Canadian Journal of Economics*, 36(1), pp. 137-154.

Egger, P., and M. Larch , 2008. "Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis," *Journal of International Economics*, 76 (2), pp. 384--399.

Eliste, P., and P. G. Fredriksson, 2004. "Does Trade Liberalization Cause a Race to the Bottom in Environmental Policies? A Spatial Econometric Analysis," Ch. 18 in *Advances in Spatial Econometrics: Methodology, Tools, and Applications*, edited by L. Anselin, R.J.G.M. Florax and S.J. Rey.

Fredriksson, P. G., 1999. "The Political Economy of Trade Liberalization and Environmental Policy," *Southern Economic Journal*, 65(3), pp. 513-535.

Hadri, K. and R. Larsson, 2005. "Testing for Stationarity in Heterogeneous Panel Data Where the Time Dimension Is Finite," *Econometrics Journal*, 8(1), pp. 55-69.

Hoffmann R., C.-G. Lee, B. Ramasamy and M. Yeung, 2005. "FDI and Pollution: a Granger Causality Test using Panel Data," *Journal of International Development*, 17(3), pp. 311-317.

Hurlin, C. and V. Mignon, 2007. "Second Generation Panel Unit Root Tests", Mimeo, University of Paris X.

Im, K.S., M.H. Pesaran and Y. Shin, 2003. "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115(1), pp. 53-74.

Kellenberg, D., 2009. "An Empirical Investigation of the Pollution Haven Effect with Strategic Environment and Trade Policy," *Journal of International Economics*, 78(2), pp. 242-255.

LeSage, J. P., and Z. Pan, 1995. "Using Spatial Contiguity as Bayesian Prior Information in Regional Forecasting Models," *International Regional Science Review*, 18(1), pp. 33-53.

Levin, A., C.F. Lin and C.S.J. Chu, 2002. "Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties," *Journal of Econometrics*, 108(1), pp. 1-24.

Love, I., and L. Zicchino, 2006. "Financial Development and Dynamic Investment Behavior: Evidence from Panel VAR," *The Quarterly Review of Economics and Finance*, 46(2), pp. 190-210.

Maddala, G.S., and S. Wu, 1999. "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, special issue, pp. 631-652.

Moon, H. R. and B. Perron, 2004. "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122(1), pp. 81-126.

Moon, H. R., B. Perron and P.C.B. Phillips, 2007. "Incidental Trends and the Power of Panel Unit Root Tests," *Journal of Econometrics*, 141(2), pp. 416-459.

Mutl, J, 2009. "Panel VAR Models with Spatial Dependence", Economics Series 237, Institute for Advanced Studies.

Mur, J. and J. H. P. Paelinck, 2009. "Some issues on the concept of Causality in Spatial Econometric models," presented at the third World Conference of the Spatial Econometrics Association.

Newey, W. K., and R. J. Smith, 2004. "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72(1), pp. 219-255.

Pesaran, M. H., 2007. "A Simple Panel Unit Root Test in the Presence of Cross-section Dependence," *Journal of Applied Econometrics*, 22(2), pp. 265-312.

Sims, C., A., 1980. "Macroeconomics and Reality," *Econometrica*, 48(1), pp 1-48.

Smarzynska, B. J., and S.-J. Wei, 2005. "Pollution Havens and Foreign Direct Investment: Dirty Secret or Popular Myth?," *Contributions to Economic Analysis & Policy*, 3, 1244-1244.

Spatareanu, M. 2007. "Searching for Pollution Havens: The Impact of Environmental Regulations on Foreign Direct Investment", *The Journal of Environment Development*, 16, 161-182.

Stern, D. I., 2005. "Global Sulfur Emissions from 1850 to 2000," Chemosphere, 58 (2), pp. 163-175.

Zedlewski, J., 2008. "Practical Empirical Likelihood Estimation with matElike", Harvard Mimeo.

Chapter 3: Complex FDI and Environmental Regulation

Baltagi, B. H., P. Egger, and M. Pfaffermayr, 2005. "Estimating Models of Complex FDI: Are There Third-Country Effects?.", *Journal of Econometrics*, 127(1), pp. 260-281.

Bergstrand, J. H. and P. Egger, 2007. "A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises," *Journal of International Economics*, 73(2), pp. 278-308.

Blonigen, B. A., R. B. Davies, H. T. Naughton, and G. R. Waddell, 2008. "Spacey Parents: Spatial Autoregressive Patterns in Inbound FDI." in S. Brakman and H. Garretsen (eds.), *Foreign Direct Investment and the Multinational Enterprise*, MIT Press

Blonigen, B. A., R. B. Davies, H. T. Naughton, and G. R. Waddell, 2007. "FDI in space: Spatial Autoregressive Relationships in Foreign Direct Investment," *European Economic Review*, 51(5), pp. 1303-1325.

Blundell, R., and S. Bond, 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87(1), pp. 115-143.

Busse M., J., Königer and P. Nunnenkamp, 2010. "FDI Promotion through Bilateral Investment Treaties: More Than a Bit?," *Review of World Economics*, 146(1), pp. 147-177.

Chinn, M. D., and H. Ito, 2007. "A New Measure of Financial Openness," *Journal of Comparative Policy Analysis*, 10(3), pp. 309-322.

Cole, M, A., and R. J. R. Elliott, 2005. "FDI and the Capital Intensity of Dirty Sectors: A Missing Piece of the Pollution Haven Puzzle," *Reviews of Development Economics*, 9(4), pp. 530-548.

Cole, M. A., R. J. R. Elliott, and P. G. Fredriksson, 2009. "Institutionalized Pollution Havens," *Ecological Economics*, 68(4), pp. 1239-1256.

Cole, M. A., R. J. R. Elliott, and K. Shimamoto, 2005. "Why the Grass Is Not Always Greener: the Competing Effects of Environmental Regulations and Factor Intensities on US Specialization," *Ecological Economics*, 54(1), pp. 95-109.

Coughlin, C. C., and E. Segev, 2000. "Foreign Direct Investment in China: A Spatial Econometric Study," *The World Economy*, 23(1), pp. 1-23.

Davies, R. B., and H. T. Naughton, 2006. "Cooperation in Environmental Policy: A Spatial Approach," .

Dean, J. M., M. E. Lovely, and H. Wang, 2008. "Are Foreign Investors Attracted to Weak Environmental Regulations? Evaluating the Evidence from China," *Journal of Development Economics*, 90(1), pp. 1-13.

Drukker, D., and D. Millimet, 2007. "Assessing the Pollution Haven Hypothesis in an Interdependent World," .

Ederington, J., and J. Minier, 2003. "Is Environmental Policy a Secondary Trade Barrier? An Empirical Analysis," *Canadian Journal of Economics*, 36(1), pp. 137-154.

Ederington, J., Levinson, A. and J. Minier, 2005. "Footloose and Pollution-Free," *The Review of Economics and Statistics*, 87(1), pp. 92-99.

Eliste, P. and P. G. Fredriksson, 2001. "Does Trade Liberalization Cause a Race to the Bottom in Environmental Policies? A Spatial Econometric Analysis," in L. Anselin and R.Florax (eds*) New Advances in Spatial Econometrics*, Berlin: Springer-Verlag.

Eskeland, G. S., and A. E. Harrison, 2003. "Moving to Greener Pastures? Multinationals and the Pollution Haven Hypothesis," *Journal of Development Economics* 70(1), pp. 1-23.

Evenett, S J., and W. Keller, 2002. "On Theories Explaining the Success of the Gravity Equation," *Journal of Political Economy*, 110(2), pp. 281-316.

Franzese, R. J. and J. C. Hayes, 2007. Spatial-Econometric Models of Cross-Sectional Interdependence in Political-Science Panel and Time-Series-Cross-Section Data, *Political Analysis*, 15(2), pp. 140--164.

Fredriksson, P. G., J. A. List, and D. L. Millimet, 2003. "Bureaucratic Corruption, Environmental Policy and Inbound US FDI: Theory and Evidence," *Journal of Public Economics*, 87(7-8), pp. 1407-1430.

Garretsen, H., and J. Peeters, 2009. "FDI and the Relevance of Spatial Linkages: Do Third Country Effects Matter for Dutch FDI?," *Review of World Economics*, 145(2), pp. 319-320.

Head, K., and T. Mayer, 2004. "Market Potential and the Location of Japanese Investment in the European Union," *The Review of Economics and Statistics*, 86 (4), pp. 959-972.

Head, K. and J. Ries, 2008. "FDI as an outcome of the market for corporate control: Theory and evidence," *Journal of International Economics*, 74(1), pp. 2-20.

Hoffmann, R., L. C. Ging, B. Ramasamy, and M. Yeung, 2005. "FDI and Pollution: A Granger Causality Test using Panel Data," *Journal of International Development*, 17(3), pp. 311-317.

Javorcik, S. B. J., and S.-J. Wei, 2005. "Pollution Havens and Foreign Direct Investment: Dirty Secret or Popular Myth?," *Contributions to the Journal of Economic Analysis & Policy*, 3, pp. 1244-1244.

Kellenberg, D. K., 2009. "An Empirical Test of the Pollution Haven Hypothesis with Strategic Environment and Trade Policy," *Journal of International Economics*, 78(2), pp. 242-255.

Kukenova, M. and J.-A. Monteiro, 2009. "Spatial Dynamic Panel Model and System GMM: A Monte Carlo Investigation," .

Levinson, A. and M. S. Taylor, 2008. "Trade and the Environment: Unmasking the Pollution Haven Hypothesis," *International Economic Review*, 49(1), pp. 223-254.

List, J. A., and C. Y. Co, 2000. "The Effects of Environmental Regulations on Foreign Direct Investment," *Journal of Environmental Economics and Management*, 40(1), pp. 1-20.

Markusen, J.R. and K.E. Maskus, 2002. "Discriminating Among Alternative Theories of the Multinational Enterprise," *Review of International Economics*, 10(4), pp. 694-707.

Milner, H., Botcheva-Andonova, L. and E. D. Mansfield, 2007. "International Trade and Environmental Policy in the Post-Communist World," *Comparative Political Studies*, 40(7), pp. 782-807.

Roodman, D., 2007. "A Short Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics*, 71(1), pp. 135-1580.

Stern, D. I., 2005. "Global Sulfur Emissions from 1850 to 2000," *Chemosphere*, 58 (2), pp. 163-175.

Wagner, U., and C. Timmins, 2008. "Agglomeration Effects in Foreign Direct Investment and the Pollution Havens Hypothesis," *Environmental and Resource Economics*, 43(2), pp. 231-256.

Xing, Y., and C. Kolstad, 2002. "Do Lax Environmental Regulations Attract Foreign Investment?," *Environmental & Resource Economics* 21(1), pp. 1-22.

Yeaple, S. R., 2003. "The Complex Integration Strategies of Multinationals and Cross Country Dependencies in the Structure of Foreign Direct Investment," *Journal of International Economics*, 60(2), pp. 293-314.

Chapter 4: Unequal Spatial Diffusion of Eco-labels

Albert, J. H., and S. Chib, 1993. "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, pp. 669-679.

Basu, A. K., Chau, N. H. and U. Grote, 2004. "On Export Rivalry and the Greening of Agriculture -- the Role of Eco-labels," *Agricultural Economics*, 31(2-3), pp. 135-147.

Beaulieu, E. and J. Gaisford, 2002. "Labour and Environmental Standards: The 'Lemons Problem' in International Trade Policy," *The World Economy*, 25(1), pp. 59-78.

Beron, K. J., Murdoch, J. C. and W. P. M. Vijverberg, 2003. "Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol," *The Review of Economics and Statistics*, 85(2), pp. 286-297.

Bonsi, R. Hammett, A.L., and B. Smith, 2008. "Eco-labels and International Trade: Problems and Solutions," *Journal of World Trade*, 42(3), pp. 407-432.

Bruce, C., and A. Laroiya, 2007. "The Production of Eco-Labels," *Environmental and Resource Economics*, 36(3), pp. 275-293.

Cao, X. and A. Prakash, 2009. "Trade Competition and Domestic Pollution: A Panel Study, 1980-2003," *International Organization*, forthcoming.

Fleming, M., 2002. "Techniques for Estimating Spatially Dependent Discrete Choice Models," in Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics*. Springer-Verlag, Heidelberg.

Franzese, R. J., and J. C. Hays, 2008. "The Spatial Probit Model of Interdependent Binary Outcomes: Estimation, Interpretation and Presentation."

Geweke, J, 1993. "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics*, 8, pp. 19-40.

Gleditsch, K.S., and M.D. Ward, 2001. "Measuring Space: a Minimum Distance Database." *Journal of Peace Research*, 38, pp. 749-768.

Greaker, M., 2006. "Eco-labels, Trade and Protectionism," *Environmental and Resource Economics*,33, pp. 1-37.

Grolleau, G. and S. El Harbi, 2008. "Why some Countries Adopt Ecolabeling Schemes in their Regulatory Arsenal and Others Do Not?," *Economics Bulletin*, 17(3), pp. 1-11.

Grolleau, G., Ibanez, L., and N. Mzoughi, 2007. "Industrialists hand in hand with environmentalists: how eco-labeling schemes can help firms to raise rivals' costs," *European Journal of Law and Economics*, Springer, 24(3), pp. 215-236.

Holloway, G., Bhavani, S., and R. Sanzidur, 2002. "Bayesian spatial probit estimation: a primer and an application to HYV rice adoption," *Agricultural Economics*, 27(3), pp383-402.

Holmes C. C. and L. Held, 2006. "Bayesian Auxiliary Variable Models for Binary and Multinomial Regression," Bayesian Analysis, 1(1), pp. 145-168.

LeSage, J.P., 2000. "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models," *Geographical Analysis*, 32, pp. 19-35.

LeSage, J.P. and O. Parent, 2007. "Bayesian Model Averaging for Spatial Econometric Models," *Geographical Analysis*, 39(3), pp. 241-267.

LeSage, J. and R. K. Pace, 2009. *Introduction to Spatial Econometrics*, CRC Press Inc.

Magnani, E., 2000. "The Environmental Kuznets Curve, Environmental Protection Policy and Income Distribution," *Ecological Economics*, 32(3), pp. 431-443.

McMillen, D.P., 1992. "Probit with Spatial Autocorrelation," Journal of Regional Science, 3, pp. 335-348.

Mason, C., 2006. "An Economic Model of Ecolabelling," *Environmental Modelling and Assessment*, 11(2), pp. 131-143.

Melser, D., and P. E. Robertson, 2005. "Eco-labelling and the Trade-Environment Debate," *The World Economy*, 28(1), pp. 49-62.

Nadai, A., 1999. "Conditions for the Development of a Product Ecolabel," European Environment, 9, pp. 202-211.

OECD, 2004. "Effects of Eco-labelling Schemes: Compilation of Recent Studies," Joint Working Party on Trade and Environment.

Piotrowski, R. and S. Kratz, 1999. "Eco-labelling in the Globalised Economy, " *International Politics and Society*, 4, pp. 430-443.

Peattie, K., 2001. "Golden Goose or Wild Goose? The Hunt for the Green Consumer," B*usiness Strategy and the Environment*, 10, pp. 187-199.

Pedersen, E.R. and P. Neergaard, 2006. "Caveat Emptor -- Let the Buyer Beware! Environmental Labelling and the Limitations of 'Green' Consumerism," *Business Strategy and the Enviro*nment, 15(1), pp. 15-29.

Pinkse, J. and M. Slade, 1998. "Contracting in Space: An Application of Spatial Statistics to Discrete Choice Models," *Journal of Econometrics*, 85, pp. 125-154.

Porter, M. E and C. van der Linde, 1995. "Toward a New Conception of the Environment-Competitiveness Relationship," *Journal of Economic Perspectives*, 9(4), pp. 97-118.

Raftery, A.E. and S.M. Lewis, 1992. "One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo," *Statistical Science*, 7, pp. 493-497.

Raftery, A.E. and S.M. Lewis, 1995. "The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms," in *Practical Markov Chain Monte Carlo*, W.R. Gilks, D.J. Spiegelhalter and S. Richardson, eds.

Robertson, P. E., 2007. "Global Resources and Eco-labels: a Neutrality Result," *Review of International Economics*, 15(4), pp. 753-743.

Rodriguez-Yam, G., Davis, R. A., and L. Scharf, 2004. "Efficient Gibbs Sampling of Truncated Multivariate Normal with Application to Constrained Linear Regression."

Sammer, K., and R. Wüstenhagen, 2006. "The Influence of Eco-Labelling on Consumer Behaviour - Results of a Discrete Choice Analysis for Washing Machines," *Business Strategy and the Environment*, 15, pp. 185-199.

Sedjo, R. A., and S. K. Swallow, 2002. "Voluntary Eco-Labeling and the Price Premium," Land Economics, 78(2), pp. 272-284.

Smith, T. E. and J. P. LeSage, 2004. "A Bayesian Probit Model with Spatial Dependencies," in *Advances in Econometrics*: *Volume* 18: *Spatial and Spatiotemporal Econometrics*, J. P. LeSage and R. K. Pace (eds.), Oxford, pp. 127-160.

Swallow, S.K. and R.A. Sedjo, 2000. "Eco-labeling Consequences in General Equilibrium: A Graphical Assessment," *Land Economics* 76(2), pp. 28-36.

Teisl, M. F., Roe, B., and A. S. Levy, 1999. "Ecocertification: Why It May Not Be a Field of Dreams," *American Journal of Agricultural Economics*, 81(5), pp. 1066-1071.

Thomas, T.S., 2007. "A Primer for Bayesian Spatial Probits with an Application to Deforestation in Madagascar."

UNEP, 2005. "The Trade and Environmental Effects of Ecolabels: Assessment and Response."