ABSTRACT

Title of dissertation: DISCRETE CHOICE UNDER SPATIAL DEPENDENCE AND A MODEL OF INTERDEPENDENT PATENT RENEWALS

Diana Alexandra Iercosan,
Doctor of Philosophy, 2011

Dissertation directed by: Professor Ingmar Prucha
Department of Economics

In my thesis I develop a theoretical model of interdependent choices and an estimation strategy which I apply to model patent renewal. The model and the estimation are not confined to my application, but rather can have other applications in which firms or people are making strategic and simultaneous decisions. Chapter 1 is the introduction which contains a brief description of the structure of the thesis.

Chapter 2 provides a literature review of studies that have focused on spatial dependence with discrete choice dependent variables; recent contributions include Pinkse and Slade (1998), LeSage (2000), Kelejian and Prucha (2001), Beron and Vijverberg (2004), and Wang et al. (2009). A major difficulty in the estimation of spatially dependent discrete choice models is computational intensity.

Chapter 3 is a Monte Carlo study that investigates the small sample properties of an estimator for spatially dependent discrete choice models which is computationally simple. The analogue of a linear probability can be formulated as a spatial autoregressive Cliff and Ord (1973, 1981)-type model. The sets of Monte

Carlo experiments show that the parameters of the model can be estimated without bias using a spatial 2SLS estimator.

Chapter 4 is a study is on the determinants of patent renewal, using US patents for Computer Hardware and Software granted between 1994 and 1997. Patent protection is important in that it encourages innovation by allowing firms to rely on patents to appropriate the returns to their R&D efforts. Returns to patents are modeled to depend on the firm's willingness to pay the patent renewal fees by, e.g., Harhoff et al. (2003), Serrano (2006), and Bessen (2008, 2009), and typically ignored potential interdependences in the decision making. Liu et al. (2008) showed that patent renewal was more likely if the patent was part of a firm's sequence of citing patents. I elaborate on their result and formulate a model in which the decision to renew a patent is dependent on the decisions of other firms to renew technologically similar patents. The theoretical model implies for the probability to renew a patent to depend on the probabilities to renew other patents, where the extent of interdependence is modeled based on a measure of similarity for patents. By making use of the estimation strategy from Chapter 3, I find that indeed the decision to renew a patent is dependent on the decision to renew related patents. Results in the literature which ignored this interdependence may hence suffer from specification biases. One plausible explanation for the interdependence I find is defensive patenting in the form of patent fencing, patent blocking and patent thickets. In the latter case, litigation and negotiation can impose high costs to society and their anticipation can lead to a hold up problem, which could deter investment in R&D.

DISCRETE CHOICE UNDER SPATIAL DEPENDENCE AND A MODEL OF INTERDEPENDENT PATENT RENEWALS

by

Diana Alexandra Iercosan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Ingmar Prucha, Chair
Professor John Chao
Professor Harry Kelejian
Professor Albert Kyle
Professor John Rust

Acknowledgments

I would like to thank my thesis committee members Professors Ingmar Prucha, Harry Kelejian, John Rust, John Chao and John Ham for their guidance, my colleagues Kahwa Douoguih, Gabriel Lara-Ibarra, and David McArthur for their comments, and my family for all their support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2SLS	Two Stage Least Squares
GMM	General Methods of Moments
HAC	Heteroskedasticity and Autocorrelation Consistent
OLS	Ordinary Least Squares
SAE	Spatial Autoregressive Error
SAL	Spatial Autoregressive Lag
SLPM	Spatial Linear Probability
USPTO	United States Patent and Trademark Office

Chapter 1

Introduction

My dissertation links a spatial discrete choice estimation model and a structural interdependent choice model to analyze patent renewal. The analysis uses a new estimation methodology that of a spatial linear probability model (SLPM). I show the small sample properties of this estimator and then I use this econometric framework to investigate strategic behavior behind patent renewal.

Chapter 2 is a review of the discrete choice models in the spatial literature. The review includes parametric, simulation and Bayesian methods that were considered for estimation of the spatial models. As opposed to the linear model I investigate in the Chapter 3, most of the existing methods in the literature have a high computational burden or only include estimation for models with spatially correlated errors, but do not allow for models with spatially lagged dependent variables. The chapter extends into the peer effects literature and presents the identification issues that concern linear in means models, as well as how nonlinearities can be used to solve these issues.

Chapter 3 examines the small sample properties of the SLPM. This model is specified with an endogenous spatial lag in the dependent variable. The SLPM model can be estimated in two steps. The first step is a two stage least squares procedure as in Kelejian and Prucha (1998, 2010), with instruments that are standard for the spatial literature. In the second step I used an IV estimator, where the optimal instrument is constructed based on first step estimates. The Monte Carlo experiments show that the parameter estimates for the model are unbiased both under uncorrelated and under correlated choice variables.

Chapter 4 includes an application of the estimator to patent renewal. Patent protection is important in that it encourages innovation by allowing firms to rely on patents to appropriate the returns to their R&D efforts. Formally, I consider a simultaneous choice single equilibrium discrete model. The model involves individuals playing a game in which they maximize their individual utilities while taking into account the synergies coming from strategic interactions. The model is based on an assumption made after Heckman and Snyder (1997) that the random utility components have a uniform distribution in order to linearize the choice probabilities. It turns out that this model can express the probability of making a choice in term of the probabilities of one's neighbor of making the same choice and can be estimated by means of spatial linear probability model. I confirm that there is strategic behavior and positive interdependence in renewal for patents on similar technologies. Results in the literature which ignored this interdependence may hence suffer from specification biases. One plausible explanation for the interdependence I find is defensive patenting in the form of patent fencing, patent blocking and patent thickets. In the latter case, litigation and negotiation can impose high costs to society and their anticipation can lead to a hold up problem, which could deter investment in R&D.

Chapter 2

Spatial Discrete Choice Models Literature Review

2.1 Introduction

The purpose of this chapter is to review some the existing literature on spatial discrete choice models, and occasionally draws parallels with non-spatial literature, mostly from time series. More specifically it mainly examines properties and potential problems of one of the binary outcome models, the spatial probit model. Further, it will analyze cases of departure from the normal error structure and extensions to multinomial models. Afterwards, this chapter will include models of social interaction which examine different specifications, estimations and identifications strategies.

2.2 Some Spatial Binary Choice Models

There are two specification for the binary choice spatial model that define the underlying latent variable. One model is the spatial autoregressive error, while the other is the spatial autoregressive lagged dependent variable model. One might encounter the need to estimate either model, or a combination of the two.

The *spatial autoregressive error (SAE)* model is:

$$y_i^* = x_i\beta + u_i \tag{2.1}$$

where y_i^* is an unobserved latent variable, X is an $n \times k$ matrix of regressors, with rows of the form x_i, and β is the corresponding parameter vector. As opposed to the standard probit model the errors u_i do not have an independent and identically distributed (i.i.d.) structure. Instead the errors have a correlated structure of the form:

$$u_i = \rho \sum_{j=1}^{n} w_{ij} u_j + \epsilon_i \tag{2.2}$$

where ϵ_i are zero mean i.i.d. error terms, ρ is the spatial autoregressive parameter that encompasses the degree of correlation in the errors, while w_{ij} is the typical element in row i and column j of the proposed spatial weight matrix W. In spatial models the weight matrix W is namely some type of exogenous measure of inverse distance, while in social interaction models group membership offers another possible structure for the weight matrix.

The *spatial autoregressive lagged dependent variable (SAR)* model has the structure:

$$y_i^* = \lambda \sum_{j=1}^{n} w_{ij} y_j^* + x_i \beta + \epsilon_i \tag{2.3}$$

where ϵ_i are i.i.d. error terms, λ is the spatial autoregressive parameter that encompasses the interdependence, while w_{ij} are again spatial weights.

For both models the underlying latent variable is not observed, but a truncation of it is observed such that

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

By rearraging the error structure $u = (I - \rho W)^{-1}\epsilon$ the SAE model becomes

$$y^* = X\beta + (I - \rho W)^{-1}\epsilon. \tag{2.5}$$

Analogously, the SAL model can be rewritten

$$y^* = (I - \lambda W)^{-1}X\beta + (I - \lambda W)^{-1}\epsilon. \tag{2.6}$$

If a multivariate normal distribution is assumed for the error terms, ideally, according to Fleming (2004) the models would be estimated parametrically by maximizing the likelihood function:

$$L(\theta) = P(Y_1 = y_1, ..., Y_n = y_n) = \int \int_A \phi(\epsilon) d\epsilon \tag{2.7}$$

The limits of integration are over the set $A = \times_{i=1}^{n} A_{y_i}$ where

$$A_{y_i} = \begin{cases} (-\infty, 0] & \text{if } y_i = 0 \\ (0, \infty) & \text{if } y_i = 1. \end{cases}$$

The multivariate normal pdf is of the form $\phi(\epsilon) = (2\pi)^{-n/2} |\Omega|^{-1/2} \exp(-\frac{1}{2}\epsilon'\Omega^{-1}\epsilon)$ where for the SAE we have $\Omega = \sigma_\epsilon^2 (I - \rho W)^{-1}(I - \rho W)^{-1}$ while for the SAL we have $\Omega = \sigma_\epsilon^2 (I - \lambda W)^{-1}(I - \lambda W)^{-1}$.

These spatial models differ substantially from the non-spatial specifications because the spatially correlated covariance structure does not allow the simplification of the multivariate distribution into the product of univariate distributions.

In terms of parametric estimation Lee (2004) investigates the asymptotic properties of the maximum likelihood estimator (MLE) and the quasi-maximum likelihood estimator (QMLE) for the SAR model under the normal distributional specification. The author shows that the estimator consistency and asymptotic normality under some under some regularity conditions on the spatial weights matrix. However, the theory is developed for continuous dependent variables, which excludes the probit maximum likelihood.

To obtain consistent and efficient estimates for the βs and the spatial parameters, we need to solve for these parameters from the first order conditions of the full likelihood. However, these first order conditions are quite complicated and we need to solve them numerically. In practice this might not be attainable due to computational limitations.

If we attempted to estimate the spatial model, we would need to deal with the structure of the variance covariance matrix Ω, which implies a high degree of correlation and heteroskedasticity. Ignoring heteroskedasticity in a linear representation causes only a decrease in efficiency, while in the context of parametric non-spatial discrete choice models Yatchew and Grilitches (1985) warn that it leads to inconsistency of the estimates.

2.3 Maximum Likelihood Estimates Adjusted for Heteroskedasticity

2.3.1 Explicit Adjustment for Heteroskedasticity in the Probit Model

Since the standard probit model under heteroskedasticity and non normality produces inconsistent estimates. Case (1992) normalizes the variance, while assuming independence, of her SAR model specification

$$y^* = \lambda W y^* + X\beta + \epsilon \tag{2.8}$$

where y^* is the unobserved expected profits, and W is an adjacency neighborhood matrix, and the errors ϵ are i.i.d. This model is used to explain the adoption of an agricultural technology. The technology is adopted, i.e. $d_i = 1$, if it passes a certain threshold

$$d_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

After the transformation the model in (2.8) becomes

$$y^* = (I - \lambda W)^{-1} X\beta + v \tag{2.9}$$

where $v = (I - \lambda W)^{-1}\epsilon$.

For any district with n observations we have $\mathrm{I}_n - \lambda W_n = [1 + \lambda/(n-1)]I_n - \lambda/(n-1)ee' = \theta_1 I_n - \theta_2 ee'$ and by inverting it

$$(I_n - \lambda W_n)^{-1} = (1/\theta_1)[I_n + \theta_2/(\theta_1 - n\theta_2)ee']. \tag{2.10}$$

Making use of the expansion in (2.10) the model becomes $\mathrm{y}^* = (1/\theta_1)[X\beta + \theta_2/(\theta_1 - n\theta_2)\bar{X}\beta] + v$ where $\bar{X}$ is the matrix of mean household characteristics.

The covariance matrix structure $\Omega = (I - \lambda W)^{-1}(I - \lambda W')^{-1}\sigma_\epsilon^2$ introduces heteroskedasticity and correlation into the model. Consequently, Case (1992) normalizes the variance of the model in (2.9) to control for heteroskedasticity and transforms the model:

$$y^{**} = (D^{*-1})y^* = (D^{*-1})(I - \lambda W)^{-1} X\beta + (D^{*-1})(I - \lambda W)^{-1}\epsilon \tag{2.11}$$

$$= (D^{*-1})(I-\lambda W)^{-1}X\beta + (D^{*-1})(I-\lambda W)^{-1}\epsilon \quad (2.12)$$

where

$$\begin{aligned} D^* &= diag(\Omega)^{1/2} = diag(E(vv'))^{1/2} \\ &= (1/\theta_1)[1+2\theta_2/(\theta_1-n\theta_2)+n\theta_2^2/(\theta_1-n\theta_2)^2]^{1/2}I \end{aligned}$$

Since $\Pr(d_i=1)=\Pr(y_i^*>0)$ after normalization in (2.12) through premultiplcation by D^{*-1}, is equivalent to $\Pr(y^{**}>0)$, leading to an ordinary probit (OP) estimation.

This correction for heteroskedasticity method for the SARAR(1,1) model can be obtained similarly:

$$y^* = (I-\lambda W)^{-1}X\beta + v \quad (2.13)$$

with

$$v = (I-\lambda W)^{-1}(I-\rho W)^{-1}\epsilon \quad (2.14)$$

based on the covariance matrix

$$E[vv'] = \sigma_\epsilon^2(I-\lambda W)^{-1}(I-\rho W)^{-1}(I-\rho W')^{-1}(I-\lambda W')^{-1} \quad (2.15)$$

$$= \sigma_\epsilon^2\left[(I-\lambda W')(I-\rho W')(I-\rho W)(I-\lambda W)\right]^{-1} \quad (2.16)$$

The standard deviation is not identified, so one would normalize it, i.e. $\sigma_\epsilon^2 = 1$. The heteroskedasticity correction is done by pre multiplying the model in (2.13) by $\mathrm{D}^{*-1} = diag\,(E(vv'))^{-1/2}$ and estimating the transformed model by maximum likelihood assuming incorrectly observational independence:

$$\ln L = \sum_i y_i \ln \Phi\left(X_i^*\beta\right) + (1-y_i)\ln\left[1-\Phi\left(X_i^*\beta\right)\right] \quad (2.17)$$

where $X^{*} = D^{*-1}(I - \lambda W)^{-1} X$ and $\Phi(.)$ is the standard normal cumulative distribution function. Case (1992) offers no proof on why this pseudo likelihood function in (2.17) produces consistent estimates.

2.3.2 Maximum Likelihood Equivalence to Nonlinear Weighted Least Squares

Another correction for heteroskedasticity was done by McMillen (1992), who looks at the case of heteroskedasticity and no autocorrelation, in a framework related to the spatial literature. The SAE model based on (2.1), (2.2) and (2.4) could be expressed as

$$y_i = x_i\beta + u_i \tag{2.18}$$

where the heteroskedasticity in the spatially correlated error term u_i can be modeled using the spatial expansion method. This expansion is similar to the one in Case (1992), which replaces the parameters of the model with functional forms of the weights, e.g. $\sigma_{u_i} = g(Z_i\gamma)$. The observation specific standard deviation is not identified, so one would normalize it for a certain part of the subpopulation and estimate the rest of the terms in the variance matrix based on the normalization. The objective function that is being maximized is of the form

$$\ln L = \sum_i y_i \ln \Phi\left(X_i\beta/g(Z_i\gamma)\right) + (1 - y_i)\ln\left[1 - \Phi\left(X_i\beta/g(Z_i\gamma)\right)\right] \tag{2.19}$$

If the heteroskedasticity is specified correctly, McMillen (1992) state without proof that the maximum likelihood (ML) estimates, β_{ML} and γ_{ML}, are consistent, even if the underlying errors are spatially autocorrelated. Moreover, the estimators are efficient if the errors are not autocorrelated. McMillen (1992) show that the ML estimator for $\theta = (\beta', \gamma')'$ is equivalent to a nonlinear weighted least squares estimator (NLWLS) such that

$$\theta = (R'H'HR)^{-1}R'H'Hy \tag{2.20}$$

where $\mathrm{R}_{i.} = \left[\frac{\phi_i}{g_i}X_i, -g_i'\phi_i g_i^{-2}X_i\beta Z_i\right]$ with $\phi_i = \phi(X_i\beta/g(Z_i\gamma))$ and $g_i = g(Z_i\gamma)$, also H is a diagonal weight matrix which has entries of the form $[\Phi_i(1-\Phi_i)]^{-1/2}$ such that $\Phi_i = \Phi(X_i\beta/g(Z_i\gamma))$.

McMillen (1992) suggests to use an iterated WLS procedure to estimate the SAE model. To compare this iterated WLS estimator to the OP estimator McMillen (1995) performs a Monte Carlo study. McMillen (1995) estimates the heteroskedasticity corrected model by maximizing the likelihood specified in (2.19). In the specification of the experiment the author introduces heteroskedasticity of the form $\sigma_i = \exp(\gamma_0 z_i + \gamma_1 z_i^2)$. Based on Dubin (1992), the author defines spatial autocorrelation of the form $\Omega_{ij} = \exp\left(-\frac{|x_i - x_j|}{\lambda}\right)$ in the model defined by (2.1) and (2.4). The homoskedastic probit model holds when $\gamma_0 = \gamma_1 = \lambda = 0$. In this model the variance increases in γ_0 penalizing the heteroskedasticity adjusted probit, by fitting the data poorly. The value of λ changes between 0 to 0.01, while γ_0 varies between 0 and 1. γ_1 changes to maintain $E\sigma_i = 1$. The number of replications is 400, which is low, in order to compromises the large computational costs of estimating a nonlinear model. Additionally, in the MC experiments the sample sizes reported are 200, 500 and 800. McMillen (1995) study shows that the OP estimator is preferable, due to smaller variance, to the more complicated heteroskedasticity adjusted estimator, when sample sizes and the degree of heteroskedasticity are small. Once the sample increases to 500 there is little gain for the OP estimator, holding the variances constant. The experiment also show that when the errors are truly homoskedasticity, the MSE of the standard probit is lower for any sample size. Autocorrelation has little effect on the results in small sample, when the errors are assumed incorrectly to be independent.

The author claims without proof that the standard probit model provides inconsistent estimates when $\gamma_0 \neq 0$ and $\gamma_1 \neq 0$. Similarly, McMillen (1995) claims in this case the heteroskedastic probit model provides consistent estimates of β, γ_0, and γ_1. Moreover he states without proof that imposing the restriction of no spatial autocorrelation, i.e. $\lambda = 0$, incorrectly does not result in inconsistent estimates in small samples. McMillen (1995) also show that the Lagrange Multiplier (LM) test has low power in small samples, and that one does better by using the Likelihood Ratio (LR) to test for heteroskedasticity.

2.3.3 Partial MLE Spatial Bivariate Probit: Theory and Simulation Results

Wang et al. (2009) are concerned with estimation of the SAE model with discrete choice dependent variable and define a partial maximum likelihood estimator for their model. They focus on a probit model, but their approach generalizes to other discrete choice models (e.g. logit). The idea in the paper is to divide spatial dependent observations into many small groups in which adjacent observations belong to a group. They motivate this approach by claiming that adjacent observations account for the most important spatial correlations between observations. They specify conditional joint distributions within groups, which utilizes more information of spatial correlation. By estimating their model by Partial Maximum Likelihood (PML) they obtain estimates that are consistent and more efficient than the GMM estimators discussed before.

To define the bivariate distribution, they partition the sample pairwise, and define n groups for $2n$ observations. The authors let the latent variables for each group g be $\{Y_g^* : g = 1, ..., n\}$, and the observed binary response variables for pair g be: $Y_g = (Y_{g1}, Y_{g2})$.

The authors define the partial likelihood function for the SAE model in the bivariate case as

$$\begin{aligned} L(\theta) &= \sum_{g=1}^{n} Y_{g1} Y_{g2} \log P_g^{11} + Y_{g1} (1 - Y_{g2}) \log P_g^{10} \\ &\quad + (1 - Y_{g1}) Y_{g2} \log P_g^{01} + (1 - Y_{g1}) (1 - Y_{g2}) \log P_g^{00} \end{aligned}$$

where $\mathrm{P}_g^{ij} = P_g(Y_{g1} = i, Y_{g2} = j | X_g)$, with $i, j \in \{0, 1\}$ is the conditional bivariate probability. They prove consistency and asymptotic normality of the estimator. Rather then using "in fill" asymptotics, they impose the restriction that the sampling area is increasing uniformly at a rate $\sqrt{n}$, essentially the process needs to be mixing in a spatial sense.

Since a full likelihood is not specified they use only information across groups. Rather than assuming group independence, they limit the degree correlation between groups. The key assumption of $\alpha-$mixing they employ is that the dependence among groups decays sufficiently quick, as the distance between groups increases. Further, Wang et al. (2009) proposed an estimation method for the variance covariance matrix.

They simulate the partial maximum likelihood estimator (PMLE) and show it is more efficient than the GMM estimator of Pinkse and Slade (1998), and less computationally demanding relative to the full information methods. The moment conditions constructed based on the generalized residuals of the heteroskedastic probit only use information on the diagonal elements, while the PMLE uses additional on the off diagonal information between two closest neighbors.

2.3.4 Maximum Likelihood Estimator: Spatial Bivariate Probit

In their paper Pinkse and Slade (2007) look at an estimator for a binary choice spatial model where the spatial endogeneity is defined in terms of the probabilities of other expected choices. The estimator can be used for games with a rich set of choices that are correlated across decision makers. However, the bivariate probit model they suggest applies only to models with two rival players. For a larger number of players the likelihood function becomes intractable. The model I will discuss in Chapter 4 assumes a uniform distribution rather than a normal distribution for the errors so that it allows for a game with a multitude of players. Pinkse and Slade (2007) model the binary decisions as functions of covariates and rival's probabilities of making certain choices, not of their actual choices, removing the possibility of having multiple equilibria of each period game. In Pinkse and Slade (2007) the decisions are made simultaneously, while in Pinkse et al. (2006) the decision was based on past observed choices. Pinkse and Slade (2007) apply the estimator to a dynamic game of price competition, decision to advertise and choice of aisle display.

The latent variable model in Pinkse and Slade (2007)

$$Y^*_{bst} = \alpha \sum_{b' \neq b} P_{b'st}(v_{st}, \theta) + x'_{bst}\beta + \gamma v_{st} - u_{bst} \tag{2.21}$$

where Y^*_{bst} represents the benefits from advertising brand b, in store s, and at time t. $P_{b'st}(v_{st}, \theta)$ is the conditional probability of advertising of rival brand b' given known information to brands in the store s, v_{st}, with $\theta = (\alpha, \beta', \gamma)'$. Moreover the private information to brand b is u_{bst}. The stochastic components v_{st} and u_{bst} are normally distributed, with u_{bst} independent across b, and v_{st} creating dependence between brands in the same store.

The probability of advertising based on known v is $\mathrm{P}_{bst}(v, \theta) = P(Y^*_{bst} \geq 0 | v_{st}, X_{st}; \theta) = \Phi\left(\alpha \sum_{b' \neq b} P_{b'st}(v_{st}, \theta) + x'_{bst}\beta + \gamma v_{st}\right)$.

In the case of a duopoly, since the decisions are made simultaneously, four joint probabilities need to be specified of the form

$$P(Y_{bst} = 1, Y_{b'st} = 1 | X_{st}) = P^{11}_{bb'st}(\theta) = \int P_{bst}(v, \theta) P_{b'st}(v, \theta) \phi(v) dv$$

Pinkse and Slade (2007) specify the likelihood function

$$\hat{L}(\theta) = \sum_{bst} l_{bb'st}(\theta) \tag{2.22}$$

$$\begin{aligned} l_{bb'st}(\theta) \quad = \quad & Y_{bst} Y_{b'st} \log P^{11}_{bb'st}(\theta) + Y_{bst}\left(1 - Y_{b'st}\right) \log P^{10}_{bb'st}(\theta) \\ & + \left(1 - Y_{bst}\right) Y_{b'st} \log P^{01}_{bb'st}(\theta) + \left(1 - Y_{bst}\right)\left(1 - Y_{b'st}\right) \log P^{00}_{bb'st}(\theta) \end{aligned}$$

under the assumption of independence across time and store, and they correct for the spatial correlation by adjusting the standard errors in the next step. The asymptotics of the estimator that maximizes (2.22), based on previous work of Pinkse et al. (2006), are of the form

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, \Gamma^{-1}(\theta)\Omega(\theta)\Gamma^{-1}(\theta))$$

where $\Gamma(\theta) = \lim_{n\to\infty} \frac{1}{n}\sum_{bb'st}\left(E\frac{\partial^2 l_{bb'st}}{\partial\theta\partial\theta'}(\theta)\right)$ and, the variance covariance $\Omega(\theta)$ is estimate based on the Newey-West type spatial weights, which corrects for the bias caused by the spatial correlation:

$$\Omega(\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{bb'st\tilde{b}\tilde{b}'\tilde{s}\tilde{t}} E\left(\frac{\partial l_{bb'st}}{\partial\theta}(\theta)\frac{\partial l_{\tilde{b}\tilde{b}'\tilde{s}\tilde{t}}}{\partial\theta'}(\theta)\right).$$

2.3.5 Spatial Panel Probit Model: with Assumed Non-Endogeneity.

Egger and Larch (2008) define one latent variable model for preferential trade agreements

$$\begin{aligned} y_t^* &= \rho W_{t-5}y_{t-5} + X_{t-5}\beta + \bar{\rho}\bar{W}\bar{y} + \bar{X}\bar{\beta} + \epsilon_t \\ y_t &= I(y_t^* > 0) \end{aligned}$$

where bars indicate time averages. Under their assumptions $W_{t-5}y_{t-5}$ is sequentially exogenous, i.e. the past outcome cannot be determined by a future outcome, they estimate the model as a probit with heteroskedastic disturbances.

2.4 Generalized Method of Moments (GMM) Estimators

2.4.1 Probit Model in Time Series

In the time series literature Gourieux et al. (1984) consider the case of homoskedastic and weakly autocorrelated disturbances and show that the ordinary probit (OP) estimates, i.e. established under assumption of independence, are consistent and asymptotically normal, but inefficient. These results come as an extension, for the probit model, from the ordinary Tobit results in Robinson (1982), which hold under certain regularity conditions. Poirier and Ruud (1988) develop a computationally simple estimator that improves upon the OP estimator in the presence of serial correlation. They show that a consistent estimator in the homoskedastic case is a solution to the moment conditions of the form $\mathrm{X}' \{E [u|y]\} = 0$.

The moments result from the first order conditions of the maximization of the likelihood function, and their solution constitutes the ML estimator. The moment conditions are based on the expectation with respect to all y's, while in the OP case u_i is defined conditional on y_i Consequently, the ML estimates are more efficient than the OP ones. This estimator and others such as Avery et al. (1983) are part of what are known as orthogonality condition (OC) estimators for the probit model. Avery et al. (1983) show that the quasi ML estimator under the incorrect assumption of no serial correlation is a special case of their OC estimator. Given this, it is not surprising that when spatial correlation is limited in the SAE model, the pooled probit that accounts for the heteroskedasticity in the marginal distribution is generally consistent for spatially correlated data.

2.4.2 GMM based on Generalized Residuals

Similarly to McMillen (1992), Pinkse and Slade (1998) analyze a broader class of heteroskedastic discrete choice models, which includes the SAE model specified by (2.1), (2.2) and (2.4). Pinkse and Slade (1998) build a GMM estimator based on the generalized residuals spatially corrected for heteroskedasticity

$$\begin{aligned}\tilde{u}_i\left(\beta_0, \rho_0\right) &= E\left(u_i | y_i; \beta_0, \rho_0\right) \\ &= \frac{y_i - \Phi\left(X_i\beta_0/\sigma_i(\rho_0)\right)}{\Phi\left(X_i\beta_0/\sigma_i(\rho_0)\right)\left[1-\Phi\left(X_i\beta_0/\sigma_i(\rho_0)\right)\right]}\phi\left(X_i\beta_0/\sigma_i(\rho_0)\right)\end{aligned}$$

Pinkse and Slade (1998) propose using the ML score vector for the discrete choice model as a set of moment conditions for the GMM model. Further, they prove consistency and asymptotic normality of the M-estimators of β and ρ that minimize the objective function

$$Q(\beta_0, \rho_0) = \left[\tilde{u}\left(\beta_0, \rho_0\right)' Z\right] M \left[Z'\tilde{u}\left(\beta_0, \rho_0\right)\right] \tag{2.23}$$

where Z is a matrix of instruments and M is a positive semidefinite matrix. The regularity conditions required for the proof to hold are that the variances are finitely bounded. However, Lahiri (1996) shows that the GMM estimators based on spatial data are inconsistent under infill asymptotics, since they converge to non-degenerate limiting random vectors. When ρ is unknown the GMM model has to estimate β and ρ jointly, requiring the evaluation of the variance covariance matrix Ω for any candidate of ρ. This can be computationally difficult since the nonlinear optimization involves inverses of $n \times n$ matrices. The authors do not report the covariance estimates, since the asymptotic results might not hold in small sample.

2.4.3 GMM Framework

Further, Pinkse et al. (2006) look at a binary choice model allowing for generic spatial and time series dependence in the errors and prove that the GMM Continuous Updating Estimator (CUE) is consistent and asymptotically normal. They apply their technique to estimate the impact that price, operating costs, reserves, capacity, price volatility, and prior state have on Canadian mines yearly decision whether to operate or not. The decisions are not interdependent in their model.

In a dynamic space-time framework Pinkse et al. (2006) consider a discrete-choice with fixed effects model:

$$\mathrm{y}_{it} = I(x'_{it1}\theta_0 - \epsilon_{it1} \geq 0)y_{i,t-1} + I(x'_{it0}\theta_0 - \epsilon_{it0} \geq 0)(1 - y_{i,t-1}) + \eta_i + u^*_{it}$$

where y_{it} represents the binary choice for firm i at time t, u^*_{it}'s are errors, x'_{its} are regressors vectors, and η_i are the fixed individual effects that enter linearly the decision of the firm. The interpretation of the fixed effects will have to be according to this specification.

The error terms ϵ_{its} are distributed normally, independent of $y_{i,t-1}$ and of past and current x'_{its}. There is assumed to be a vector of instruments z_{it} independent of ϵ_{its} such that

$$\begin{aligned} E(y_{it}|z_{it}) &= E(I(x'_{it1}\theta_0 - \epsilon_{it1} \geq 0)y_{i,t-1}|z_{it}) \\ &+ E(I(x'_{it0}\theta_0 - \epsilon_{it0} \geq 0)(1 - y_{i,t-1})|z_{it}) + E(\eta_i|z_{it}) \\ &= E(\Phi(x'_{it1}\theta_0)y_{i,t-1}|z_{it}) + E(\Phi(x'_{it0}\theta_0)(1 - y_{i,t-1})|z_{it}) + E(\eta_i|z_{it}) \end{aligned}$$

To cancel the fixed effects the model is differenced $\mathrm{E}[\mathrm{y}_{it} - y_{i,t-1} - \Phi(x'_{it1}\theta_0)y_{i,t-1} + \Phi(x'_{i,t-1,1}\theta_0)y_{i,t-2} - \Phi(x'_{it0}\theta_0)(1 - y_{i,t-1}) + \Phi(x'_{i,t-1,0}\theta_0)(1 - y_{i,t-2})|z_{it}] = 0$ and the moment condition is derived

$$\begin{aligned} 0 &= g_{it}(\theta) = z_{it}(y_{it} - y_{i,t-1} - \Phi(x'_{it1}\theta_0)y_{i,t-1} + \Phi(x'_{i,t-1,1}\theta_0)y_{i,t-2} \\ & \quad -\Phi(x'_{it0}\theta_0)(1 - y_{i,t-1}) + \Phi(x'_{i,t-1,0}\theta_0)(1 - y_{i,t-2})) \end{aligned}$$

such that $Eg_{it}(\theta_0) = 0,$ for any i, t.

The CUE of Pinkse et al. (2006) is $\hat{\theta} = \arg\min_{\theta\in\Theta} \hat{\Omega}_n(\theta)$ based on the objective function $\Omega = \bar{g}'_n(\theta)\hat{W}_n(\theta)\bar{g}_n(\theta)$ where the sample moments are $\bar{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_{ni}(\theta)$, where g_{ni} is some vector-valued function, while the weight matrix is of the form: $\hat{W}_n(\theta) = \Psi_n\hat{V}_n^{-1}(\theta)$. The CUE is similar to the two-step GMM estimator, with the weight matrix $\hat{W}_n(\theta)$ parametrized immediately. In small sample the weight matrix $\hat{W}_n(\theta)$ is based on: $\hat{V}_n(\theta) = \frac{1}{n}\sum_{i,j=1}^{n} \lambda_{nij}(g_{ni}(\theta) - \bar{g}_n(\theta))(g_{ni}(\theta) - \bar{g}_n(\theta))'$. The Ψ_ns in $\hat{W}_n(\theta)$ are scalars and do not affect the estimates, and are taken to be equal to the maximum row sum of the location matrix Λ_n. The general element in column i and row j of Λ_n is λ_{nij} and is taken to be close to one when observation i's location is close to j's. Convergence is shown in n, where n is N or NT depending if T is fixed or not.

While the two step GMM estimator and the GMM CUE have the same asymptotic distribution. The CUE estimator is a GEL (Generalized Empirical Likelihood) and unlike in the two step GMM a second order, i.e. n^{-1}, bias correction is possible. The bias correction improves performance in moderate sample sizes. The bias correction in Pinkse et al. (2006) is based on the one provided by Iglesias and Phillips (2008).

2.4.4 Expectation Maximization (EM) Algorithm

The SAE model was developed to account for the autocorrelation in the error terms, but it implicitly introduced heteroskedasticity. McMillen (1992) offer without proof of consistency an estimator that takes into account both heteroskedasticity and spatial autocorrelation resulting from the SAE, as well as the SAL case.

In the SAE case:

$$y_i^* = X_i\beta + v_{1i} \tag{2.24}$$

with $v_{1i} = \sum_j \delta_{1ij}\epsilon_j$ with δ_{1ij} a typical element of $(I - \rho W)^{-1}$, while in the SAL case

$$y_i^* = \sum_j \delta_{2ij} X_j \beta + v_{2i} \tag{2.25}$$

with $v_{2i} = \sum_j \delta_{2ij}\epsilon_j$ with δ_{2ij} a typical element of $(I - \lambda W)^{-1}$. Based on the assumptions v_{1i} and v_{2i} are distributed normally with heteroskedastic variances of the form $\sigma^2_{v1i} = \sigma^2_\epsilon \sum_j \delta^2_{1ij}$, and respectively $\sigma^2_{v2i} = \sigma^2_\epsilon \sum_j \delta^2_{2ij}$. A standard assumption needed for identification is $\sigma^2_\epsilon = 1$.

McMillen (1992) suggest using the EM algorithm to estimate the models in (2.24) and (2.25). The author shows no proof for the consistency or asymptotic properties of the resulting estimators. Dempster et al. (1977) introduce an approach of iterative computation of maximum-likelihood estimates under incomplete data. Their algorithm is called EM, since each iteration of the algorithm consists of an expectation step followed by a maximization step. Dempster et al. (1977) show that convergence of the algorithm implies a stationary point of the likelihood. Convergence is defined in practice as a small change in the parameters, Levine and Casella (2001), or in the log-likelihood function, Weeks and Lange (1989) and Aitkin and Aitkin (1996). Other criteria that involve the gradient and the Hessian, such as in Ruud (1991), are computationally harder to implement.

In the probit case the method involves replacing the binary dependent variable by the expectation of the continuous underlying latent variable. The formulas used by McMillen (1992) to construct the two continuous variables for the SAE model are

$$\begin{aligned} E(y_i^*|y_i &= 1) = X_i\beta + \sigma_{v1i}\frac{\phi\left(X_i\beta/\sigma_{v1i}\right)}{\Phi\left(X_i\beta/\sigma_{v1i}\right)} \\ E(y_i^*|y_i &= 0) = X_i\beta - \sigma_{v1i}\frac{\phi\left(X_i\beta/\sigma_{v1i}\right)}{1-\Phi\left(X_i\beta/\sigma_{v1i}\right)}, \end{aligned}$$

while for the SAL are

$$\begin{aligned} E(y_i^*|y_i &= 1) = X_i^*\beta + \sigma_{v2i}\frac{\phi\left(X_i^*\beta/\sigma_{v2i}\right)}{\Phi\left(X_i^*\beta/\sigma_{v2i}\right)} \\ E(y_i^*|y_i &= 0) = X_i^*\beta - \sigma_{v2i}\frac{\phi\left(X_i^*\beta/\sigma_{v2i}\right)}{1-\Phi\left(X_i^*\beta/\sigma_{v2i}\right)}, \end{aligned}$$

based on a prior estimate of β, ρ, and λ. Using these expectations as continuous dependent variables one can construct the likelihood function that needs to be maximized. The new parameter estimates are used to calculate a new expectation, and the process is repeated until convergence. Since the likelihood involves n integrals, it makes the calculation of the information matrix to obtain the VC matrix intractable. Consequently, for the empirical part in order to get the standard errors of the estimates McMillen (1992) interprets the probit model as NLWLS. The covariance is estimated conditional on knowing ρ and λ. For the SAE model

$$\begin{aligned} VC(\beta) &= \left(F_1'\Omega_1^{-1}F_1\right)^{-1} \\ F_{1i} &= \left(\phi_i X_i\beta/\sigma_{v1i}\right)X_i \end{aligned}$$

and the VC matrix Ω_1 has a typical element ij of the form $\Phi_2\left(X_i\beta/\sigma_{v1i}, X_j\beta/\sigma_{v1j}, \tau_{1ij}\right) -$ $\Phi\left(X_i\beta/\sigma_{v1i}\right)\Phi\left(X_j\beta/\sigma_{v1j}\right)$ where τ_{1ij} is an element of $[(I-\rho W)'(I-\rho W)]^{-1}$.

For the SAL model

$$\begin{aligned} VC(\beta) &= \left(F_2'\Omega_2^{-1}F_2\right)^{-1} \\ F_{2i} &= \left(\phi_i X_i\beta/\sigma_{v2i}\right)X_i \end{aligned}$$

and the VC matrix Ω_2 has a typical element ij of the form $\Phi_2\left(X_i^*\beta/\sigma_{v2i}, X_j^*\beta/\sigma_{v2j}, \tau_{2ij}\right) -$ $\Phi\left(X_i^*\beta/\sigma_{v2i}\right)\Phi\left(X_j^*\beta/\sigma_{v2j}\right)$ where τ_{2ij} is an element of $[(I-\lambda W)'(I-\lambda W)]^{-1}$.

The author claims without proof that the EM estimates are consistent. Nevertheless, he states that the standard errors may be too small, and the unconditional estimates of ρ and λ are biased.

2.5 Simulation

In the spatial probit case there is no closed-form solution for the likelihood estimates and the n-dimensional integration makes the estimation unfeasible for a large sample. Simulation methods have been proposed to solve this computational burden. One of the methods is the maximum simulated likelihood (MSL) procedure. This method of estimation is the same as the ML but the choice probabilities are being replace with smooth simulated probabilities. Properties for simulations methods outside the spatial literature were derived by Gourieroux and Monfort (1993), Lee (1995), and Hajivassiliou and Ruud (1994).

Another method is that of maximum simulated moments (MSM) covered in a non spatial context by McFadden (1989) and Pakes and Pollard (1989), which is the typical generalized method of moments. In this method, for the error-instrument orthogonality condition, the errors are replaced by the difference between the dependent variable and the simulated probabilities, rather than the exact probabilities. Lee (1992) shows that in the non spatial multinomial model, with a large number alternatives, MSL has computational advantages to its counterpart the MSM. Since in the MSM case each alternative needs to be simulated, while in the MSL case only the chosen alternative needs to be simulated.

2.5.1 Maximum Simulated Likelihood with Spatial Dependence

Hautsch and Klotz (2003) look at a limited dependent variables in the context of spatial dependence on the past decisions of others. There is no simultaneous interdependence in the decision process. The binary decision for each one of the N individuals depends on their own characteristics, on his own decision with respect to the same problem in a prior period and on the decisions of the other decision makers in that prior period.

Individual's ith decision y_i depends on the latent variable y_i^* which is modeled as:

$$y_i^* = \alpha + x_i'\beta + y_i^- \gamma + \kappa_i^- + u_i,\ i = 1, ..., N$$

where y_i^- represents the observed binary choice of individual i in last period and κ_i^- reflects the neighborhood impact of all other individuals on decision maker i:

$$\kappa_i^- = \sum_{j=1 j\neq i}^{N} \left\{ \left[a_1 \exp(-\delta_{ij}^* b_1) y_j^- \right] + \left[a_0 \exp(-\delta_{ij}^* b_0) \left(1 - y_j^- \right) \right] \right\}$$

where a_1, a_0 are the scale rates and b_1, b_0 are the decay rates. δ_{ij}^* depends on the distance between the individuals. In their paper the social distance, δ^*_{ij}, is based on the Euclidean distance among each of the D social characteristics of the individuals i and j. The error terms u_i are i.i.d. $N(0, \sigma^2)$ and $E[\kappa_i^- u_i] = 0$.

When the estimates for the parameters a, b are found to be jointly significant, there is proof of spatial dependence on past observed choices. The model is simulated for sample sizes of 50 to 400. They notice no convergence problem even for small samples. The variance decreases as the sample size increases. There is a small sample bias, in the form of overestimation in absolute value. The estimates with the largest bias are for γ and the spatial scaling parameter a_1, while the estimates of the decay parameter b_1 has the lowest bias. The size of the bias and the variance are too large for samples under 100. Their model is also generalized to a multiple choice with S alternatives and errors with a multivariate normal distribution.

2.5.2 Geweke-Hajivassiliou-Keane Simulator

The Recursive Importance Sampling (RIS) simulator constitutes a method to evaluate an n-dimensional normal probability. In the multinomial case Bolduc et al. (1997) combine the MSL method with the Geweke-Hajivassiliou-Keane (GHK) simulator to obtain estimates in a model with spatial correlation among the choices of an individual. The GHK simulator is a particular case of a RIS simulator, in Vijverberg (1997) and Beron et al. (1997), with a standard normal kernel.

In the multinomial choice model of Bolduc et al. (1997), spatial correlation is in the unobserved utility term and relates the choices an individual faces, while the choices of different individuals are not spatially correlated. The model includes N individuals who each face J choices. The utility function for an individual n for each available option j is expressed in terms of the exogenous characteristics for each individual n of each of the choices j.

Bolduc et al. (1997) construct the empirical expectation of the probability that individual n chooses alternative j, by averaging over the R simulated probabilities: $\mathrm{f}_n(j) = \frac{1}{R}\sum_{r=1}^{R}\prod_{l=1}^{J-1}\Phi\left(a_{nrl}\right)$, where under the GHK simulator Bolduc (1999) derive the exact form of a_{nrl}.

The estimation method is based on the maximization of the logarithm of the simulated likelihood function, over some coefficients on exogenous characteristics of the choices of each individual that help parameterize the utility function:

$$L = \sum_{n=1}^{N} \ln f_n(j) = \sum_{n=1}^{N} \ln \frac{1}{R} \sum_{r=1}^{R} \prod_{l=1}^{J-1} \Phi\left(a_{nrl}\right). \tag{2.26}$$

Another model that is more closely related to the model that I will be discussing in Chapter 3 is that of Beron and Vijverberg (2004). The authors simulate the SAE and SAL RIS estimators for the discrete dependent variable. The SAE model is specified as in (2.1), (2.2) and (2.4), and the SAL model is specified as in (2.3) and (2.4). They perform Monte Carlo simulation for the sample sizes of 50, 100 and 200 and they use a real state-distance matrix, as well as a random one.

Under no spatial correlation, i.e. $\rho = 0$ and $\lambda = 0$, the estimates of ρ and λ have a positive bias. The authors show that the standard probit gives biased estimate of β depending on the nature of the data, with a negative bias for a spatial lag dependence. The estimates for β are more biased under a spatial error autocorrelation structure, even when the correct model is used. The RMSE is lowest for the standard probit, and largest for the SAE model, regardless of the data structure. Under true SAE structure, SAL is a mis-specification, but when attempting to estimate a SAL model one would get positive estimates of λ.

Beron and Vijverberg (2004) results show that comparing models based on the likelihood ratio test is quite reliable, when spatial dependence is not weak. The LR test has low power in small sample. They conclude that it is harder to observe spatial error autocorrelation than spatial lag structures.

2.5.3 Spatial Linear Probability Model: Monte Carlo Experiment

Beron and Vijverberg (2004) compare the linear probability model (LPM) estimator with the RIS estimator for the SAE and SAL models. In their specification the linear SAE model, based on the observed choices, is

$$\begin{aligned} y^* &= X\beta + u \\ u &= \rho W u + \epsilon, \end{aligned}$$

while the linear SAL model

$$y^* = \lambda W y^* + X\beta + \epsilon$$

where y^* is the latent variable and $y = I(y^* > 0)$ is the observed binary variable. The SAL model of Beron and Vijverberg (2004) has a different data generating process and it imposes more structure on the errors terms then the model I will be introducing in Chapter 3.

In a model where there is no spatial autocorrelation of any kind and the weights are based on the contiguity of US states, the likelihood ratio test based on the linear estimator picks this up only 74%, while their spatial probit picks up the lack of a spatial component 90% of the time. In the presence of a spatial component the linear model favors a decision suggesting a spatially correlated error alternative over a lag model. By comparison for larger values of λ and ρ, the spatial probit model distinguishes between the lag and the error model alternative, while the linear probability model favors a decision suggesting an SAE model over a SAL model. Under simulated weights the spatial linear model is able to separate the spatial error and spatial lag model as well as the spatial probit model would. The randomization offsets some of the otherwise possible extreme values that might occur in Wy and Wu. Moreover, when correlation exists and they use a real state based weights matrix, the linear spatial model outperforms the spatial probit model by rejecting correctly the null of no spatial component more times.

2.6 Bayesian: Gibbs Sampler

Bayesian statistical technique are used to avoid the computation of high dimension integrals. Two of these techniques are the Gibbs Sampler and the Metropolis Hastings, also known as Markov Chain Monte Carlo techniques. In the case of multinomial distributed variables draws come from a conditional distribution instead of the joint density, this sampling method is known as Gibbs. Similarly to the EM algorithm the likelihood is formulated as if the dependent variable were continuous.

According to Bolduc et al. (1997) who apply the Gibbs sampler to the SAE multinomial probit model with spatially correlated choices. Bolduc et al. (1997) want to estimate a joint unconditional posterior distribution of the parameters that enter the utility and the coeficient for the spatial correlation in alternatives.

The process is based on drawing alternatively from conditional densities using the Metropolis technique. Once the sequence of random draws converged, the mean of this chain is a consistent estimator of the posterior mean of the parameters of interest. Bolduc et al. (1997) claim that numerical errors are on average less than 1% of the estimated posterior means.

Further, LeSage (2000) incorporates heteroskedastic errors terms independent of spatial error dependence, extending the SAL and SAE model. While Egger and Larch (2008) apply Gibbs sampling based on the conditions derived by LeSage (2000) for the cross-sectional SAL binary choice model: $y^* = \rho W y^* + X\beta + \epsilon$ where $y = I(y^* > 0)$ is the existence of a preferential tax agreement.

The advantage over the EM algorithm of the Gibbs Spatial Sampler of Bolduc et al. (1997) and LeSage (2000) is that consistent standard errors are derived from the posterior distribution. The Gibbs algorithm is computationally and conceptually preferred over the RIS simulator, Bolduc et al. (1997).

2.7 Spatial Logit Adjusted For Dependence

Another discrete choice model is that of Smirnov (2008) who looks at a spatial random utility model for n agents selecting over J alternatives. The author defines the multinomial discrete choice y_{ij} to be equal to one for individual i if his utility from alternative j is highest. The utility vector of the n individuals from alternative j is

$$u_j = \rho W u_j + v_j(\beta) + \epsilon_j$$

where $v_j(\beta)$ is the vector of private deterministic component, W is the individual interaction weight matrix, and ϵ_j is the vector of private stochastic component assumed to come from an extreme value type I distribution.

The reduced form model is of the form

$$u_j = Z v_j(\beta) + Z\epsilon_j$$

such that $Z = (I - \rho W)^{-1}$. Notice that the spatial correlation is again in the unobserved component and among the alternatives as in the similar model of Bolduc et al. (1997). Further Smirnov (2008) decomposes the multiplier matrix Z into private effects of shocks (i.e. the effect of a shock in the individual utility on the utility of the same individual), D, and social effects of shocks (i.e. the effect of a shock in the individual utility on the utilities of other individuals), $Z - D$ such that

$$u_j = Z v_j(\beta) + D\epsilon_j + (Z - D)\epsilon_j.$$

This transformation facilitates the estimation of the model by a pseudo maximum likelihood, which is equivalent to estimating the likelihood of the alternative utility

$$\tilde{u}_j = Z v_j(\beta) + D(\rho)\epsilon_j,$$

which ignores the term $(Z-D)\epsilon_j$. The latter model is easier to estimate by maximum likelihood since the errors are independently distributed. Further Smirnov (2008) simulates the model for 8 alternatives and uses a linear deterministic utility $v_j(\beta)$ in two explanatory exogenous variables. The results show the smallest bias and RMSE for the estimates of ρ. Larger values of ρ cause larger biases and RMSE for the estimates of the β's.

2.8 Models of Social Interaction

2.8.1 Identification in Spatial Models

The seminal paper that discusses identification in linear in means models is that of Manski (1993). Manski (1993)'s original model of social interactions is:

$$y = \alpha + \beta E(y|x) + E(z|x)'\gamma + z'\eta + u$$

$$E(u|x, z) = x'\delta$$

where y is a scalar outcome (e.g. a youth's achievement in high school), x is vector of attributes characterizing an individual's reference group (e.g. a youth's school or ethnic group), and z, u are vectors of attributes that directly affect y (e.g. socioeconomic status and ability). A researcher observes a random sample of realizations of (y, x, z), but does not observe u. Also, $(\alpha, \beta, \gamma, \delta, \eta)$ is a parameter vector.

The author labels $E(y|x)$ to be the endogenous effect, $E(z|x)$ the contextual effect, and $x'\delta$ the correlated effect. In proposition one Manski states the conditions that insure that a social effect exists (e.g. endogenous or contextual), and is identified from the unobserved correlated effects. While in proposition two he lists the conditions for identification of a pure endogenous effect.

In one of the sections Manski (1993) looks at spatial correlation models to discuss models of endogenous effects, with no contextual or correlated effects, of the form

$$y_i = \beta W_{iN} Y + z_i' \eta + u_i$$

where $Y = (y_i, i = 1, ...N)$ is the $N \times 1$ vector of sample realizations of y and W_{iN} is a specified $1 \times N$ weighting vector whose elements add up to one. The disturbances are normally distributed and independent of the x's.

The spatial model above assumes that an endogenous effect is present within the researcher's sample rather than within the population from which the sample was drawn. Manski (1993) says that this makes sense in the case of small group interaction, when all the members know each other. However, this model does not make sense in cases of large group social effect, where the samples are randomly chosen individuals. The theory can be applied for large groups if in the first step they estimate $\mathrm{E}_N(y|x_i) = W_{iN}Y$ non parametrically, and then using z and $E_N(y|x_i)$ as independent variables in the second stage of the regression. While $E_N(y|x_i)$ is in most cases independent of $[1, z]$, the model is unidentified since $E(y|x_i)$ is a function of $[1, z]$. This produces a point estimate of β even when this parameter is unidentified. In contrast to the linear in means model, if the probit model is specified the nonlinearity introduced facilitates identification.

Similarly to Manski (1993), in Moffitt (2001)'s model peer effects are also not identified. As opposed to Manski (1993)'s model which includes the individual in the mean of the group, Moffitt (2001) does not include the individual in the mean of the group and considers groups of the same size.

In contrast, the model of Lee (2007) considers interactions in groups with different sizes, and the individual is also excluded from the mean. He finds that variations in group sizes can yield identification. Bramoulle et al. (2009) generalizes the result of Lee (2007) by considering an extended version of the linear-in-means model where interactions are structured through a social network. They assume that correlated unobservables are either absent, or treated as network fixed effects. They provide easy-to-check necessary and sufficient conditions for identification.

Moreover, Graham and Hahn (2005) focus on identification of endogenous social effects from unobserved group characteristics under the assumption that exogenous social effects are not present. They reinterpret the linear-in-means model as a quasi-panel data model, where the cross sectional dimension equals the number of observed social groups and the time series dimension equals the number of sampled individuals within each group. Using the quasi-panel reinterpretation, it is straightforward to see that Manskis first non-identification result for the linear-in-means model is analogous to the inability of a standard fixed effects regression to identify coefficients on group-invariant regressors. The authors identify the between-group variation that contains information on the social multiplier (Manski (1993); Glaeser et al. (2002)). The identification relies on the existence of instruments generating exogenous between-group variation.

Brock and Durlauf (2001a) characterize the equilibrium of a theoretical model for social global interaction. The authors specify a model for which the people's equilibrium choice is mathematically equivalent to a logit regression. They use a binary logit framework, where each agent i makes observed choices $\omega_i = \{-1, 1\}$ that gives him the utility

$$V(\omega_i) = u(\omega_i) + S(\omega_i, \mu_i^e(\omega_{-i})) + \epsilon(\omega_i) \tag{2.27}$$

The first term represents the private deterministic part of the utility, while the second term represents the social part of the utility, with $\mu_i^e(\omega_{-i})$ being the conditional probability measure agent i places on the choices of others at the time of making his own decision. The last term represents the random payoff term. The $\mu_i^e(\omega_{-i})$ measure is replaced by the average of the subjective expected value of individual $j's$ choice as seen by individual i $\bar{m}_i^e = \frac{1}{I-1}\sum_{j\neq i} m_{i,j}^e$.

Brock and Durlauf (2001a) pick two functional forms for the social component of the utility $\mathrm{S}(\omega_i, \bar{m}_i^e) = J\omega_i\bar{m}_i^e$ for a proportional spillover case, and $\mathrm{S}(\omega_i, \bar{m}_i^e) = -\frac{J}{2}(\omega_i - \bar{m}_i^e)^2$ that penalizes deviations far from the mean more strongly. Also, they linearize the private utility component to $u(\omega_i) = h\omega_i + k$.

The random part of the utility in (2.27) are independent error terms distributed extreme value, such that $\Pr(\epsilon(-1) - \epsilon(1) \leq x) = \frac{1}{1+\exp(-\beta x)}$

Consequently the joint probability measure over all choices is

$$\Pr(\omega_i) = \frac{\exp(\beta(h\omega_i + J\omega_i\bar{m}_i^e))}{\sum_{v_i\in\{-1,1\}}\exp(\beta(hv_i + k + Jv_i\bar{m}_i^e))}. \tag{2.28}$$

When $J = 0$ is the standard logit case, when $J \neq 0$ is the standard form augmented by social interactions, which characterizes the effects of the interactions on community behavior. The mean expectation is $\mathrm{E}(\omega_i) = \tanh(\beta h\omega_i + \beta J\frac{1}{I-1}\sum_{j\neq i} m_{i,j}^e)$.

By imposing rational expectations $m^e_{i,j} = E(\omega_j)$ and by the fixed point theorem they prove that there is a self consistent equilibrium that solves

$$m^* = \tanh(\beta h + \beta J m^*). \tag{2.29}$$

Brock and Durlauf (2001a) show that if $\beta J > 1$ and $h = 0$, there exist a positive, zero and a negative equilibrium value. If $\beta J > 1$ and $h \neq 0$, there exists a threshold H (depending on βJ) such that for $|\beta h| < H$ there exists three roots to (2.29) and one has the sign of h, while the others have opposite signs. On the other hand if $|\beta h| > H$, there exists a unique root with the sign of h.

In the econometric model identification holds because of the nonlinearity the binary choice model introduces. Brock and Durlauf (2001a) parameterize the utility such that $\mathrm{h}_i = k + c'X_i + d'Y_{n(i)}$ where X_i are individual observables and $Y_{n(i)}$ are neighborhood exogenous observables, since $n(i)$ implies neighborhood of individual i. Assuming that the errors follow a logit distribution the model can be solved by maximizing

$$\begin{aligned} L(\omega_I | X_i, Y_{n(i)}, m^e_{n(i)}) &= \prod_i \Pr(\omega_i = 1 | X_i, Y_{n(i)}, m^e_{n(i)})^{(1+\omega_i)/2} \\ &\quad \Pr(\omega_i = -1 | X_i, Y_{n(i)}, m^e_{n(i)})^{(1-\omega_i)/2} \\ &\sim \prod_i \exp(\beta k + \beta c' X_i + \beta d' Y_{n(i)} + \beta J m^e_{n(i)})^{(1+\omega_i)/2} \\ &\quad \prod_i \exp(-\beta k - \beta c' X_i - \beta d' Y_{n(i)} - \beta J m^e_{n(i)})^{(1-\omega_i)/2} \end{aligned}$$

where β is normalized to 1 for identification. The rationality condition, of how people form their global neighborhood expectations, is

$$m^e_{n(i)} = \int \tanh(k + c'X_i + d'Y_{n(i)} + Jm_{n(i)}) dF_{X|Y_{n(i)}}$$

$$= \int \omega_i dF(\omega_i | k + c' X_i + d' Y_{n(i)} + J m_{n(i)}) dF_{X|Y_{n(i)}}.$$

Parallel to Manski's Reflection Problem, Brock and Durlauf (2001a) show that identification of k, c, d, J is possible, while in the linear-in-means model of Manski (1993) the model is not identified since $\mathrm{m}_{n(i)} = \frac{c' E(X_i | Y_{n(i)}) + d Y_{n(i)}}{1-J}$ and $E(X_i | Y_{n(i)})$ is linear in $Y_{n(i)}$. The difference from the linear-in-means case is that the binary choice imposes a nonlinearity between group characteristics and group behaviors. Manski (1993) mentions that further difficulties in empirical work arise from having to infer the social structure as well as the strength of the interaction within neighborhoods.

Further, papers that discuss identification in models of social interaction under other various assumptions are Brock and Durlauf (2001b), Brock and Durlauf (2007) Brock and Durlauf (2002) and Blume and Durlauf (2002).

2.8.2 Simulated Likelihood with Randomized Choice Over Multiple Equilibria

Soetevent and Kooreman (2007) estimate a discrete choice model with social interactions by using simulation methods. Agents maximize a utility functions of the form $\mathrm{V}(\mathrm{y}_i, x_i, y_{-i}, \epsilon_i(y_i)) = u(y_{i,}, x_i) + S(y_i, x_i, y_{-i}) + \epsilon_i(y_i)$, with the vector of binary choices for the n individuals being $y = (y_i, y_{-i})$ where $y_{-i} = (y_1, ..., y_{i-1}, y_{i+1}, ..., y_n)'$. The deterministic private part is $u(1,, x_i) - u(-1, x_i) = \beta x_i$, the social part is $S(y_i, x_i, y_{-1}) = \frac{\gamma}{2N-2} y_i \sum_{j=1, j \neq i}^{N} y_j$, and the random private utility $\epsilon_i(y_i)$.

The derived latent variable model in this case is

$$y_i^* = \beta' x_i + s_i + \epsilon_i \tag{2.30}$$

where y_i^* is defined as the difference in utilities from choosing $y_i = 1$ as opposed to $y_i = -1$, such that the observed choice variable is

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & y_i^* \leq 0 \end{cases} \tag{2.31}$$

and the endogenous social effect is

$$s_i = \frac{\gamma}{N-1} \sum_{j=1, j \neq i}^{N} y_j \tag{2.32}$$

The model assumes no contextual effects.

To deal with multiplicity of equilibria the authors calculate the number of possible Nash Equilibria denoted by:

$$Q(\beta, \gamma, x, \epsilon, N) =$$
$$= \sum_{t=1}^{2^N} \left[I\left(\epsilon_i > -\beta' x_i - \frac{\gamma}{N-1} \sum_{j=1, j\neq i}^{N} y_j\right)^{\frac{1+y_i}{2}} I\left(\epsilon_i \leq -\beta' x_i - \frac{\gamma}{N-1} \sum_{j=1, j\neq i}^{N} y_j\right)^{\frac{1-y_i}{2}} \right]$$

When $\gamma = 0$ the model becomes the standard binary choice formulation without externalities and thus with a unique equilibrium, i.e. $Q(\beta, 0, x, \epsilon, N) = 1$.

In the application the authors make a randomization assumption: whenever the model generates multiple equilibria they assume that one of them will occur with probability equal to one over the number of equilibria. They simulate R times the number of equilibria E_r that exists in the region $W(y, \theta)$, which is the support space of y in ϵ space defined by

$$\epsilon_i > -\beta' x_i - s_i \text{ if } y_i = 1$$

$$\epsilon_i \leq -\beta' x_i - s_i \text{ if } y_i = -1$$

for $i = 1, ..., N$ with s_i defined in (2.32).

In Soetevent and Kooreman (2007) the simulated likelihood function is based on $P(y)$ the simulated probability that we observe the choice pattern y defined as

$$P(y) = P(\epsilon \in W(y, \theta)) \frac{1}{R} \sum_{r=1}^{R} \frac{1}{E_r}.$$

2.8.3 Subjective Expectation

Li and Lee (2006) modify the models by Manski-Brock-Durlauf by replacing the rational expectation with the observed subjective expectation and apply it to model voting decisions between two candidates.

The latent variable model

$$\begin{aligned} y_i^*(1) &= x_i'\beta_1 + z_{(i)}'\delta_1 + \varphi_1 u_{(i)} - \frac{1}{2n_i}\theta \sum_{j=1}^{n_i}(1 - \mu_{ij})^2 + \epsilon_i(1) \\ y_i^*(-1) &= x_i'\beta_2 + z_{(i)}'\delta_2 + \varphi_2 u_{(i)} - \frac{1}{2n_i}\theta \sum_{j=1}^{n_i}(-1 - \mu_{ij})^2 + \epsilon_i(-1) \end{aligned}$$

with the contextual effects linear in individual i's group characteristics $z_{(i)}'$; the unobserved group effects $u_{(i)}$ is distributed standard normal; the deterministic private utility depending linearly on i's known characteristics, x_i; the deterministic social utility measured by the distance between i's choice (1 or -1) and his expectation on the social choices, denoted by $\mu_i = \{\mu_{ij}, j = 1, ..., n_i\}$, where μ_{ij} is individual i's expectation on his jth discussants choice, and n_i is the number of his discussants; and $\epsilon_i(1) - \epsilon_i(-1)$ follows i.i.d logistic distribution.

Assuming rational expectations the probabilities of making one choice or the other are:

$$\begin{aligned}
\Pr(y_i &= 1|x_i, z_{(i)}) = \int \frac{1}{1+\exp(-2x_i\beta - 2z'_{(i)}\delta - 2\varphi u_{(i)} - 2\theta m_{(i)})} dF(u_{(i)}) \\
\Pr(y_i &= -1|x_i, z_{(i)}) = \int \frac{1}{1+\exp(2x_i\beta + 2z'_{(i)}\delta + 2\varphi u_{(i)} + 2\theta m_{(i)})} dF(u_{(i)})
\end{aligned}$$

where $\beta = \frac{1}{2}(\beta_1 - \beta_2)$, $\delta = \frac{1}{2}(\delta_1 - \delta_2)$ and $\varphi = \frac{1}{2}(\varphi_1 - \varphi_2)$etc.

The endogenous effects for individual i is defined as: $m_i = \frac{1}{n_i+1}\sum \tanh(x'_i\beta + z'_{(i)}\delta + \theta m_i)$ where μ_{ij} is the expectation of person's j decision viewed by i, typically different from +1 or -1.

In their empirical voting model respondent i's objective expectation on his/her jth, discussant vote intention, i.e. μ_{ij}, is replaced by subjective expectation , i.e. e_{ij}. At the same time the group mean becomes $p_i = \frac{1}{\tilde{n}_i}\sum_{j=1}^{\tilde{n}_i} e_{ij}$, where $\tilde{n}_i$ is the number of i's discussants whose expected choices are non zero. In their model the standard contextual effects are insignificant, instead Li and Lee (2006) build two contextual variables, a group average partisanship and a frequency of disagreement in discussion. In the model using subjective expectations one cannot identify the unobserved group variable $u_{(i)}$, even though it could be done in the rational expectation model. They also develop a framework for testing rational expectations, and show that expectation on average social choice is rational in their voting model application.

2.9 Conclusion

In the reviewed literature there is little theoretical evidence for the consistency of estimators for spatial discrete choice models. While for the SAE model consistency is proved for some estimators, there is no theory to show consistency in the SAL model case. Several estimators have been proposed and their small sample properties were shown through Monte Carlo Simulations.

In the spatial literature, the hardest part is implementing some of these estimators for the probit case. Unlike the logit case, the normal distribution lacks a closed-form mathematical expression.

The decision to implement one method or another to estimate the spatial probit model depends on the trade-off between the computational burden and accuracy. Out of the estimation methods reviewed in this chapter the RIS simulator has the highest computational costs, while the nonlinear least squares methods are the least computationally intensive. The estimates of the standard errors for the spatial parameter are more accurate in larger samples for all models, except for the EM models and nonlinear SAE model which give biased estimates.

The social interaction models appear as a separate part of the literature, even though it is strongly related to the spatial models. Extensive theory is available on identification of the parameters, the multiplicity of equilibria and their characterization. The results are based on nonstandard law of large numbers, and central limit theorems. In these models the endogenous social effect is computed as an average of rational or subjective expectation of others decision. Some Monte Carlo studies were done to show properties of the ML estimators of the social interaction models. In future work, the spatial and social interaction areas of research should be unified under a common framework.

There is a gap in the spatial literature is gap for consistent and computationally tractable estimation techniques for the SAL model. Consequently, in the next chapter I will discuss a spatial linear probability model and the estimation strategy for this model.

2.10 Discrete Choice Games of Incomplete Information

2.10.1 Estimation based on Constrained Optimization

Su and Judd (2010) propose an approached called Mathematical Program with Equilibrium Constraints (MPEC) for structural estimation. The approach consists of choosing structural parameters and endogenous economic variables so as to maximize the likelihood (or minimize the moment conditions) of the data subject to the constraints that the endogenous economic variables are consistent with an equilibrium for the structural parameters. This approach is computationally faster than others since it only needs to solve exactly for the equilibrium associated with the final estimate of parameters.

The authors introduce a two player discrete choice game with both observed and unobserved heterogeneity. The players are a and b with their respective binary choices to stay active or not being d_a and d_b. The ex-post utility functions for player a and b are:

$$u_a(d_a, d_b, x_a, \epsilon_a) = \theta^a_{d_a,d_b} x_a + \epsilon_a(d_a)$$

$$u_b(d_a, d_b, x_b, \epsilon_b) = \theta^b_{d_a,d_b} x_b + \epsilon_b(d_b),$$

where x_a is the observed type of player a and ϵ_a is the unobserved type of player a, while $\theta^a_{d_a,d_b}$ measures the effect of the type x_a on the utility of player a and is based on the joint decision of both players (d_a, d_b). The parameters and other terms are similarly defined for player b. ϵ_a and ϵ_b are considered independent.

Let p_a be player's b belief of the probability of player a being active. Given his belief p_a, player's b expected utility from taking action d_b is

$$\begin{aligned} U_b(d_b, x_b, \epsilon_b) &= p_a u_b(1, d_b, x_b, \epsilon_b) + (1 - p_a) u_b(0, d_b, x_b, \epsilon_b) \\ &= p_a \theta^b_{1,d_b} x_b + (1 - p_a) \theta^b_{0,d_b} x_b + \epsilon_b(d_b) \end{aligned}$$

The probability that player b stays active is

$$\begin{aligned} p_b &= Prob\{d_b = 1\} \\ &= Prob\{\epsilon_b | U_b(1, x_b, \epsilon_b(1)) > U_b(0, x_b, \epsilon_b(0))\} \\ &\equiv \Psi(p_a, p_b, x_b; \theta^b) \end{aligned}$$

where ϵ_b has an extreme value distribution.

The Bayesian Nash equilibrium equations can be rewritten as:

$$p = \Psi(p, x; \theta)$$

where $x = (x_a, x_b)$, $p = (p_a, p_b)$ and $\Psi = (\Psi_a, \Psi_b)$.

The model is then extended to M markets where the players' decisions in one market do not depend on their decision in other markets. Then the Bayesian Nash equilibrium can be defined for each market m:

$$p^m = \Psi(p^m, x^m; \theta)$$

with$x^m = (x^m_a, x^m_b)$ and $m = 1, ..., M$.

The researcher observes the players' types $x^m = (x_a^m, x_b^m)$ and the players' decisions $d^{mt} = (d_a^{mt}, d_b^{mt})_{t=1}^T$ over T periods. Let X^m denote the data observd in market m:

$$X^m = \{x^m = (x_a^m, x_b^m), d^{mt} = (d_a^{mt}, d_b^{mt}), fort = 1, ..., T\}.$$

The MPEC approach for estimation involves the augmented log likelihood function in market m

$$\begin{aligned} L(X^m, p^m, \theta) &= \sum_{t=1}^{T} d_a^{mt} log(p_a^m) + (1 - d_a^{mt}) log(1 - p_a^m) \\ &+ \sum_{t=1}^{T} d_b^{mt} log(p_b^m) + (1 - d_b^{mt}) log(1 - p_b^m) \end{aligned}$$

and for all markets

$$L(\theta, P) = \sum_{m=1}^{M} L(X^m, p^m, \theta)$$

where P is the collection of the probabilities p^m for all M markets.

The MPEC estimator is the solution to

$$\max_{\theta,P} = \frac{1}{M} L(\theta, P)$$
$$\text{subject to } P = \Psi(P, x, \theta).$$

2.10.2 Computing Equilibria using All-Solution Homotopy

Bajari et al. (2010) propose a numerical algorithm to compute the entire set of equilibrium for private information games. An equilibrium to a private information game is a fixed point that can be represented in closed form. The system of equations characterizing the equilibrium choice probabilities is generated by a logit model. The authors adapt the all solution homotopy to find as many roots at possible to private information games. The authors find that the majority of games with incomplete information have only one equilibrium, and that the equilibrium set of outcomes depends on the incomplete information assumption.

Bajari et al. (2010) present a game of entry where player i's decision to enter a particular market is $a_i = 1$ and $a_i = 0$ denotes the decision to not enter the market. The payoff for player i is:

$$\pi_i(a_i, a_{-1}, s; \theta) = \begin{cases} s'\beta + \delta \sum_{j \neq i} 1\{a_j = 1\} & \text{if } a_i = 1 \\ 0 & \text{if } a_i = 0. \end{cases},$$

where s is a vector of state variables.

The errors $\epsilon_i(a_i)$ capture shocks on the profitability of entry that are private information to firm i and are assumed to be distributed extreme value. Then the profit maximization by firm i implies the choice probability of entry:

$$\sigma_i(a_i = 1|s) = \frac{exp(s'\beta + \delta \sum_{j \neq i} \sigma_j(a_j = 1|s))}{1 + exp(s'\beta + \delta \sum_{j \neq i} \sigma_j(a_j = 1|s))} \tag{2.33}$$

A fixed point equilibrium existing for this model a finite s.

The authors propose getting a first set of estimates $\hat{\sigma}_i(a_i = 1|s)$ assuming that markets always play the same equilibrium through the observed time period. This set of estimates could be done using a strategy such as a linear probability model. The second stage would consist of estimating the logit model. This estimation method works when there is a unique solution to the system in (2.33) or when only one fixed point of this system gets realized in the observed data.

Further, the authors propose an homotopy method for estimation when the multinomial logit choice model has multiple equilibria.

Chapter 3

A Spatial Linear Probability Model

3.1 Introduction

The aim of this paper is to provide some evidence on the finite sample properties for a spatial linear probability model corresponding to a model where probabilities are determined as the solution to an analogue of a spatial autoregressive model. As shown in Chapter 4 this model arises in many empirical applications. The case of interest here is a Cliff and Ord (1973, 1981) type model that contains a binary dependent variable, a spatial lag in the dependent variable, exogenous variables, and disturbance terms with known or unknown correlation structure.

The linear model I consider is a spatial autoregressive lag (SAR) model. However, a large part of the spatial discrete choice literature focuses only on the spatial autoregressive error (SAE) model, in which only the errors are spatially correlated. The model I consider can be estimated by an instrumental variable technique and a heteroskedastic and autocorrelation consistent (HAC) estimator for the variance covariance matrix. Thus the estimation approach does not lead to inconsistent estimates, as may be the case with other estimation approaches that ignore heteroskedasticity in binary parametric models, and that lead to inconsistent estimates as Yatchew and Griliches (1985) point out. In the spatial literature, Case (1992) and McMillen (1992) consider the model with spatially correlated errors (SAE) and propose an explicit adjustment for heteroskedasticity and estimate the transform model by maximum likelihood and nonlinear least squares, respectively. Similarly, Pinkse and Slade (1998) analyze a broader class of heteroskedastic discrete choice models with spatially correlated errors. The authors are the first to prove the consistency and asymptotic normality of GMM estimates based on generalized residuals spatially corrected for heteroskedasticity. Moreover, Wang et al. (2009) suggest using a maximum likelihood estimator under spatially correlated errors and show it is consistent. They specify the partial maximum likelihood function for a bivariate spatial probit and use the information within groups to gain in efficiency. Rather than assuming group independence, they limit the degree of correlation between groups. The assumption is that the dependence across groups decays sufficiently quick as the distance between groups increases. They simulate the partial maximum likelihood estimator (PMLE) and show it is more efficient than the GMM of Pinkse and Slade (1998), and less computationally demanding relative to the full information methods. The moment conditions constructed based on the generalized residuals of the heteroskedastic probit only uses information only on the diagonal elements, while the PMLE uses additional off diagonal information between two closest neighbors. All these papers ignore the interesting case of spatial interdependence in the decisions, which I consider in my model.

Some work has been done for a model with a spatial lag in the discrete dependent variable using simulation and Bayesian techniques. An alternative to integrating the likelihood function is the Recursive Importance Sampling (RIS) simulator. The method simulates the maximum likelihood function and is used to estimate discrete choice models where there is spatial correlation and/or spatial dependence, in Bolduc et al. (1997), Beron and Vijverberg (2004). The latter is a Monte Carlo study that compares the SAE and SAL RIS models to the spatial linear probability model. In case of correlation the linear spatial model outperforms the spatial probit model by rejecting correctly the null of no spatial component many times.

Bayesian statistical techniques are also used to avoid the computation of high dimension integrals. Two of the techniques implemented in spatial models are the Gibbs Sampler in Bolduc et al. (1997) and LeSage (2000) and the Metropolis Hastings, also known as the Markov Chain Monte Carlo. Bolduc et al. (1997) notes that the Gibbs sampler algorithm is computationally and conceptually preferred over the RIS simulator. Even though these techniques to estimate discrete choice SAL models exist, the Bayesian and simulation methods are computationally extremely demanding, and the statistical properties of those estimators have not yet been fully explored.

I consider a theoretical model where the probabilities of making a decision are interdependent. That is the probability of an individual making a decision is expressed as a function of the probability of the decisions of its neighbors and its own exogenous characteristics. As a result, the decision making process is modeled of the form :

$$P = \lambda W P + X\beta,$$

where P denotes the $N \times 1$ vector $P = [P_1, \ldots, P_N]' = [P(y_1 = 1), \ldots, P(y_N = 1)]'$, X denotes an $N \times k$ matrix of non-stochastic regressors, W is the $N \times N$ spatial weight matrix, λ is the spatial autoregressive coefficient, and β is the $k \times 1$ vector of coefficients.

As shown below, this model is then converted into a spatial linear probability model

$$y = \lambda Wy + X\beta + u$$

where the vector of choices is $y = [y_1, y_2, ..., y_N]'$ and where the error terms u's are spatially correlated and heteroskedastic.

In terms of the estimation, I will later show that Kelejian and Prucha (1998, 2010) have an estimation method for such a model, and that I use it to estimate the spatial linear probability model. Kelejian and Prucha (1998, 2010) propose instrumental variable (IV) consistent estimators for the regression parameters of the model with a spatial lag and with autoregressive homoskedastic or heteroskedastic innovations. Further, to estimate the variance covariance matrix consistently one can use the spatial heteroskedasticity and autocorrelation consistent (HAC) estimation of the variance covariance matrix developed by Kelejian and Prucha (2007). Their estimator is a weighted sum of sample covariances with weights depending on the relative distances, for some bandwidth parameter. Kim and Sun (2010) derive the optimal bandwidth parameter, based on the asymptotic truncated MSE criterion, and suggest its data dependent estimation procedure using a parametric plug-in method.

The advantages of the model I investigate with respect to the ones that already exist in the spatial discrete choice literature are that the estimation technique is easy to implement and that the computational burden of the estimation is very small even for large data sets.

3.2 Model Specification and Estimator

Let y_i be a choice variable taking values of 1 and 0. Now suppose that spatial discrete choice model is given by:

$$P(y_i = 1) = \lambda \sum_j w_{ij} P(y_j = 1) + x_i \beta, \tag{3.1}$$

or in matrix notation

$$P = \lambda W P + X\beta.$$

Since $P = Ey$ model (3.1) can also be written equivalently as

$$Ey = \lambda W Ey + X\beta.$$

Now define

$$\epsilon = y - Ey,$$

so that the spatial linear probability model becomes:

$$\begin{aligned} y &= \lambda W y + X\beta + u & (3.2)\\ &= Z\delta + u & (3.3)\\ u &= (I - \lambda W)\epsilon, \end{aligned}$$

with $Z = [X, Wy]$ and $\delta = [\beta', \lambda]'$.

Given that by construction $\epsilon = y - Ey$, we have $E\epsilon = 0$, thus also $Eu = 0$. By letting Ω_ϵ and Ω_y denote the variance covariance matrix of ϵ and y, respectively, then clearly $\Omega_\epsilon = \Omega_y$. The model in (3.1) does not restrict the correlation between ϵ_i and ϵ_j or equivalently y_i and y_j, except that $\Omega_\epsilon = \Omega_y$ has to be positive semidefinite.

Solving for P from the data yields:

$$P = (I - \lambda W)^{-1} X\beta,$$

with the resulting variances of ϵ_i and y_i given by $P_i(1 - P_i)$.

Kelejian and Prucha (2011) discuss some sufficient conditions which will insure that the model is correctly specified and that it has a solution such that $0 \le P_i \le 1$. In particular, Kelejian and Prucha (2011) show that since $\max_i |P_i| \le \frac{1}{1-|\lambda|} \max_i \mu_i$ with $\mu_i = \sum_k \beta_k x_{ik} \ge 0$ a sufficient condition for $0 \le P_i \le 1$ in the case when $w_{ij} > 0$ and $\lambda \in [0, 1)$ is:

$$\min_i \sum_k x_{ik}\beta_k \ge 0$$

and

$$\max_i \sum_k x_{ik}\beta_k \le 1 - \lambda,$$

or even more restrictive

$$x_{ik} \ge 0, \beta_k \ge 0$$

and

$$\max_i x_{ik} \sum_k \beta_k \le 1 - \lambda.$$

The estimation procedure for the cross-section spatial autoregressive model in (3.2) is done in two steps.

The right hand side Wy is clearly endogenous $E(Wy\epsilon') = W(I-\lambda W)^{-1}\Omega_\epsilon \neq 0$. The ideal instrument for Wy is $E(Wy)$. For our model it holds that $E(Wy) = WP$, and alternatively by expansion:

$$
\begin{aligned}
E(Wy) &= E(W(I-\lambda W)^{-1}X\beta) \\
&= W[I+\lambda W+\lambda^2 W^2+\ldots]X\beta \\
&= WX\beta + W^2X(\lambda\beta) + W^3X(\lambda^2\beta)+\ldots
\end{aligned}
$$

The estimator in the first step is the two-stage least square estimator (2SLS) as suggested by Kelejian and Prucha (1998). H is the feasible non-stochastic $N \times p$ matrix of instruments used in the 2SLS procedure for $Z = [X, Wy]$. Since $E(Wy)$ is a linear combination of WX, W^2X, etc, H contains the linearly independent columns of $[X, WX, W^2X]$.

The first step estimator is defined as:

$$
\hat{\delta}_{2SLS} = (\hat{Z}'Z)^{-1}\hat{Z}'y \tag{3.4}
$$

where $\hat{Z} = H(H'H)^{-1}H'Z$.

The second step estimator is based on the constructed ideal instruments $\tilde{Z} = [X, W\hat{P}]$ where $\hat{P} = (I-\hat{\lambda}_{2SLS}W)^{-1}X\hat{\beta}_{2SLS}$[1] such that:

$$
\hat{\delta}_{IV} = (\tilde{Z}'Z)^{-1}\tilde{Z}'y. \tag{3.5}
$$

[1] $\hat{\lambda}$ and $\hat{\beta}$ come from the first estimation step

3.3 Small Sample Properties: A Monte Carlo Experiment

3.3.1 Monte Carlo Design

For each Monte Carlo experiment I perform a total of $M = 1000$ iterations using M generated sets of Y's and one set of X's. I let $X = [e, x]$ consist of a constant, e, and a regressor, x that is generated by drawing from the Uniform(0,1) distribution. Given X and λ, $\beta = (\beta_0, \beta_1)$ are selected such:

$$\max_i x_{i,k} * \sum_k \beta_k < 1 - \lambda,$$

and

$$x_{ik} \geq 0, \beta_k \geq 0$$

where $k = 0, 1$ and $i = 1, ..., N$. These are sufficient condition so that $0 \leq P_i \leq 1$. The true parameter values for these experiments are $\lambda = 0, 0.3, 0.5, 0.8$, while the intercept, $\beta_0 = 0.04$, will be set to be one fourth of the slope, $\beta_1 = 0.16$.

The spatial units are assumed to be located on a regular square grid with coordinates $\{(r, s) : r, s = 1, 2, ..., m\}$ such that there are a total of $N = m^2$ observations. The sample sizes are taken to be $N = 400, 1024$. The distance d_{ij} between observations i and j is given by the Euclidean distance: $d_{ij} = \sqrt{(r_i - r_j)^2 + (s_i - s_j)^2}$ As in Baltagi et al. (2003) the weights matrix W is a $N \times N$ rook matrix, which considers that two units i and j are neighbors if the Euclidean distance is less than or equal to one. On a regular grid, consider an arbitrary interior unit (r^*, s^*), then the neighbors of this unit are $(r^* + 1, s^*)$, $(r^* - 1, s^*)$, $(r^*, s^* + 1)$ and $(r^*, s^* - 1)$. Units on the borders have fewer neighbors that can be defined analogously. The contiguity matrix W has zeros on the main diagonal and is normalized so that the weights in each row sum to one. Define the i,j-th element of W as: $\mathrm{w}_{ij} = w^*_{ij} / \sum_{j=1}^{n} w^*_{ij}$ where, based on the distance measure,

$$w^*_{ij} = \begin{cases} 1 & \text{if } 0 < d_{ij} \leq 1 \\ 0 & \text{else} \end{cases}$$

Since the model in (3.1) allows for independent and correlated outcomes, it makes sense to study the two cases separately.

1. Independent Outcomes

Generate M vectors $y = (y_1, ..., y_N)'$ with independent y_i distributed $Bernoulli(P_i)$, where P_i is the ith element of the vector:

$$P = (I - \lambda W)^{-1} X\beta,$$

The method involves getting $(\zeta_1, ..., \zeta_N)$, an $N \times 1$ vector containing pseudorandom values drawn from the standard uniform distribution on the open interval $(0, 1)$, and letting:

$$y_i = \begin{cases} 1 & \text{if } \zeta_i \leq P_i \\ 0 & \text{else} \end{cases}$$

Use the M sets of y's and one set of X's to estimate the model in (3.2).

2. Correlated Outcomes

The correlation between y_i and y_j is a function of the Euclidean distance:

$$corr(y_i, y_j) = exp(-\sqrt{(r_i - r_j)^2 + (c_i - c_j)^2}/\gamma)$$

where y_i is located on the grid in column c_i and row r_i. Similar to Dubin(1992) and McMillen(1995) who define the spatial autocorrelation:

$$corr(y_i, y_j) = exp(-||z_i - z_j||_p/\gamma)$$

where $||\cdot||_p$ is the Euclidean norm in $\Re_p$. Notice that the correlation in outcomes increases in the parameter γ and decreases in the Euclidean distance between the units.

A valid positive definite covariance matrix Ω_y, with the correlation imposed above, has elements of the form:

$$Cov(y_i, y_j) = corr(y_i, y_j)\sqrt{var(y_i)}\sqrt{var(y_j)}$$

I generate draws from the multivariate Bernoulli with variance covariance matrix Ω_y using Krummenauer (1998) algorithm as described next.

The first step is performed before the simulation. In step 1) I solve the equations

$$\Phi(u_{P_i}, u_{P_j}; \theta_{ij}) = P_{ij}, 1 \leq i \leq j \leq N$$

to obtain the parameters $\theta_{ij}, 1 \leq i \leq j \leq N$, where $\Phi(x, y; \rho)$ stands for the cumulative distribution function of the standard binormal distribution with mean vector (x, y) and correlation ρ. Also, P_{ij}s are derived from the covariance matrix such that $P_{ij} = Cov(y_i, y_j) + P_iP_j$. And u_{P_i} is the P_i-th quantile of the univariate normal distribution. Since θ_{ii} is a correlation it will be equal to 1. Note, that a unique collection of solution parameters θ_{ij} will exist for any consistent choice of covariances due to the strict monotonicity of $\Phi(x, y; .)$.

The next two steps are repeated M times in the simulation to get M sets of outcomes $y = (y_1, ...y_N)'$.

In step 2) generate a N-normal random vector $(\zeta_1, ..., \zeta_N)$ with mean vector $(0, ..., 0)$ and correlation matrix $(\theta_{ij})_{1 \leq i \leq j \leq N}$.

In step 3) obtain the N-variate Bernoulli vector $y = (y_1, ..., y_N)'$ with marginals y_i distributed $Bernoulli(P_i)$ and covariances $P_{ij} - P_i P_j$ for $1 \leq i \leq j \leq N$ by setting:

$$y_i = \begin{cases} 1 & \text{if } \zeta_i \leq u_{P_i} \\ 0 & \text{else} \end{cases}$$

for any $1 \leq i \leq N$.

Then use the M sets of y's and one set of X's to estimate the model in (3.2).

3.3.2 Monte Carlo Results

The results of the simulations for the independent y's case are in Table 3.1 and Table 3.2. When the sample size is 400, see Table 3.1, the bias of $\hat{\lambda}_{2SLS}$, the first step estimator of the spatial autoregressive parameter, is quiet large and it increases in absolute value with λ. For positive λ and large first step bias, the bias of the second step estimators, i.e. $\hat{\lambda}_{IV}$ and $\hat{\beta}_{IV}$, decrease substantially. However, the Root-Mean-Square Error (RMSE) increases in the second step suggesting that the major component of the RMSE is the variance of the estimator, not the bias.

When the sample size is increased from 400 to 1024, the bias for $\hat{\lambda}_{2SLS}$, see Table 3.2, decreases relative to the bias for $\hat{\lambda}_{2SLS}$ in Table 3.1. In Table 3.2, there is an even further decrease in the bias of the estimates in the second step of the estimation, while the RMSE stays almost unchanged between the two steps. A comparison between Table 3.1 and Table 3.2 shows that RMSE declines with an increase in the sample size, providing simulation evidence for the consistency of the procedure.

Further, I show results of the simulations with various degrees of dependence in the y's in Table 3.3, Table 3.4, Table 3.5 and Table 3.6. By comparing the results for both sample size 400 and 1024, the first step bias the estimator $\hat{\lambda}_{2SLS}$ is noticeably larger when $\gamma = 0.1$, i.e. the dependence is smaller, and when λ is positive. However, when there is no spatial interdependence, i.e. $\lambda = 0$, the bias of the first step estimator $\hat{\lambda}_{2SLS}$ is larger for higher dependence in y's , i.e. $\gamma = 0.5$. In most cases there is significant improvement in the bias while RMSE moderately increases, for all the estimates in the second step.

Again, by comparing Table 3.3 with Table 3.4 and Table 3.5 with Table 3.6, the bias in the coefficients decreases systematically as the size of the sample increases from 400 to 1024.

Table 3.1: Independent y's: N=400

True Parameter Values				
λ	0	0.3	0.5	0.8
β_0	0.04	0.04	0.04	0.04
β_1	0.16	0.16	0.16	0.16
	Bias and RMSE	Bias and RMSE	Bias and RMSE	Bias and RMSE
λ				
First Step	-0.0344 0.6105	-0.0988 0.6712	-0.1670 0.7018	-0.1857 0.5097
Second Step	-0.0528 0.7216	-0.0418 0.8139	-0.0366 0.7677	-0.0426 0.4895
β_0				
First Step	-0.0029 0.0784	0.0110 0.1120	0.0317 0.1607	0.0946 0.2799
Second Step	-0.0003 0.0933	-0.0022 0.1318	-0.0088 0.1612	0.0202 0.2441
β_1				
First Step	-0.0093 0.0538	-0.0073 0.0669	-0.0038 0.0767	0.0100 0.1009
Second Step	-0.0122 0.0611	-0.0102 0.0727	-0.0076 0.0833	0.0026 0.1090

Table 3.2: Independent y's: N=1024

True Parameter Values				
λ	0	0.3	0.5	0.8
β_0	0.04	0.04	0.04	0.04
β_1	0.16	0.16	0.16	0.16
	Bias and RMSE	Bias and RMSE	Bias and RMSE	Bias and RMSE
λ				
First Step	0.0122 0.4070	-0.0493 0.4506	-0.0888 0.4499	-0.0895 0.3184
Second Step	0.0294 0.4690	-0.0094 0.5064	-0.0082 0.4880	0.0003 0.3104
β_0				
First Step	-0.0034 0.0478	0.0066 0.0699	0.0162 0.0978	0.0460 0.1731
Second Step	-0.0057 0.0518	-0.0027 0.0785	-0.0052 0.1037	-0.0043 0.1517
β_1				
First Step	-0.0046 0.0360	-0.0027 0.0446	-0.0006 0.0510	0.0005 0.0668
Second Step	-0.0053 0.0369	-0.0050 0.0491	-0.0030 0.0567	-0.0021 0.0717

Table 3.3: Dependent y's: N=400, γ=0.1

True Parameter Values				
λ	0	0.3	0.5	0.8
β_0	0.04	0.04	0.04	0.04
β_1	0.16	0.16	0.16	0.16
	Bias and RMSE	Bias and RMSE	Bias and RMSE	Bias and RMSE
λ				
First Step	-0.0185 0.6093	-0.1223 0.6667	-0.1896 0.6717	-0.2187 0.5430
Second Step	-0.0124 0.7644	-0.0189 0.8536	-0.1004 0.8038	-0.0539 0.5682
β_0				
First Step	-0.0020 0.0781	0.0139 0.1136	0.0336 0.1540	0.1190 0.3020
Second Step	-0.0079 0.0962	-0.0051 0.1334	0.0103 0.1709	0.0262 0.2796
β_1				
First Step	-0.0131 0.0587	-0.0115 0.0667	-0.0049 0.0747	0.0024 0.0957
Second Step	-0.0174 0.0641	-0.0160 0.0724	-0.0076 0.0838	-0.0036 0.1113

Table 3.4: Dependent y's: N=1024, γ=0.1

True Parameter Values				
λ	0	0.3	0.5	0.8
β_0	0.04	0.04	0.04	0.04
β_1	0.16	0.16	0.16	0.16
	Bias and RMSE	Bias and RMSE	Bias and RMSE	Bias and RMSE
λ				
First Step	-0.0192 0.4180	-0.0326 0.4644	-0.0805 0.4290	-0.1086 0.3229
Second Step	0.0045 0.4680	0.0193 0.4890	0.0100 0.4726	-0.0214 0.2950
β_0				
First Step	0.0004 0.0519	0.0045 0.0739	0.0165 0.0924	0.0567 0.1752
Second Step	-0.0036 0.0559	-0.0079 0.0789	-0.0063 0.0981	0.0048 0.1477
β_1				
First Step	-0.0055 0.0387	-0.0040 0.0431	-0.0008 0.0511	0.0048 0.0647
Second Step	-0.0059 0.0395	-0.0050 0.0440	-0.0025 0.0541	0.0026 0.0704

Table 3.5: Dependent y's: N=400, γ=0.5

True Parameter Values				
λ	0	0.3	0.5	0.8
β_0	0.04	0.04	0.04	0.04
β_1	0.16	0.16	0.16	0.16
	Bias and RMSE	Bias and RMSE	Bias and RMSE	Bias and RMSE
λ				
First Step	0.1063 0.6060	0.0126 0.6696	-0.0656 0.6326	-0.1493 0.5257
Second Step	0.0436 0.7857	0.0593 0.8126	0.0041 0.7927	-0.0218 0.5337
β_0				
First Step	-0.0224 0.0826	-0.0060 0.1049	0.0078 0.1400	0.0729 0.2881
Second Step	-0.0152 0.0990	-0.0140 0.1269	-0.0125 0.1674	0.0016 0.2780
β_1				
First Step	-0.0103 0.0603	-0.0116 0.0690	-0.0130 0.0804	-0.0035 0.0909
Second Step	-0.0135 0.0655	-0.0152 0.0747	-0.0174 0.0867	-0.0109 0.1014

Table 3.6: Dependent y's: N=1024, γ=0.5

True Parameter Values				
λ	0	0.3	0.5	0.8
β_0	0.04	0.04	0.04	0.04
β_1	0.16	0.16	0.16	0.16
	Bias and RMSE	Bias and RMSE	Bias and RMSE	Bias and RMSE
λ				
First Step	0.0662 0.4157	0.0223 0.4477	-0.0142 0.4436	-0.0753 0.3127
Second Step	0.0137 0.4898	0.0201 0.4969	0.0378 0.4820	-0.0113 0.2762
β_0				
First Step	-0.0106 0.0536	-0.0068 0.0752	0.0005 0.0987	0.0352 0.1614
Second Step	-0.0067 0.0627	-0.0077 0.0849	-0.0142 0.1032	-0.0003 0.1412
β_1				
First Step	-0.0057 0.0351	-0.0034 0.0405	-0.0030 0.0485	0.0023 0.0616
Second Step	-0.0071 0.0372	-0.0045 0.0418	-0.0066 0.0518	0.0008 0.0647

3.4 Conclusion

The spatial linear probability model that I discussed includes in addition to most models spatial interdependence in the choices, and is computationally more tractable than the Bayesian or Simulation methods. Further, the Monte Carlo experiment I performed suggests that the two step estimator for the spatial linear probability model has reasonable finite sample properties with a small bias that decreases with an increase in the sample size under both dependent and correlated outcomes.

Chapter 4

A Model Of Interrelated Patent Renewals

4.1 Introduction

The patent system was established in order to offer incentives for innovation and technological progress by allowing for temporary monopolistic rents. These rents arise from the possibility to exclude other firms from the market of products, which utilize the patented technology, and/or from the licensing or selling of the patent. Patents are a source of economic returns to research and development (R&D), which is important for economic growth. Theoretical studies have explored the impact of R&D on strategic interaction among firms and long run growth.[1] While R&D spending is related to the inputs that go into the innovative process, patents and their value are indicators of the output and the quality of innovative activity. A better understanding of the determinants of the economic value of a patent helps policy makers improve the rules governing the patent system to foster R&D investment, and target Government funding to high risk projects in technological areas of high patent return. Moreover, the value of a patent provides investors with a dollar measure for the value of innovation, an otherwise intangible asset.

Under the current United States patent policy, the level of protection declines

[1]See, for example, Romer (1990), Aghion and Howitt (1992), Spence (1984), and Reinganum (1989); and Griliches (1992) and Keller (2004) for surveys of the literature.

suddenly at the end of the patent's statutory life unless the patent is renewed.[2] In order to renew a patent, the patent owner must pay periodically non trivial maintenance fees. Starting with Pakes and Schankerman (1984) this feature of the patent system has been used in studies to estimate the returns to patented innovations.[3] Patent renewal reveals the implicit returns to a patent, since the patent holder assesses whether the returns exceed the maintenance fees or not, when making the choice to keep a patent in good standing or let it expire. The existing empirical literature focuses on estimating returns based on renewal fees and proxy variables that capture the quality of the patent.[4] This literature assumes that patent-holders make the decision to renew a patent as independent from renewal decisions regarding other patents. The assumption that renewal decisions are made independently may not hold when strategic effects are in place. As a result, the existing estimates in the literature that are based on models assuming independence between an assignee's decision to renew a patent and other assignees' decisions to renew similar patents may exhibit biases, and hence be misleading.

In this paper, I introduce an economic model that allows for the renewal decisions for, say, n patents to be interdependent, and where firms renew a patent if the expected revenues exceed expected costs. The economic model implies that the probability of renewing patent i will depend on the probabilities of renewing

[2]Hopenhayn et al. (2006) show that the life of a patent should only end if something better replaces the existing patent, for optimal R&D investment.

[3]See Lanjouw et al. (1998) for a review of this literature. More recent additions include Serrano (2006) and Baudry and Dumont (2006).

[4]See Lanjouw and Schankerman (2004), Harhoff et al. (2003) and Bessen (2008).

technologically similar patents and other covariates. In my case, patents are considered technological more or less similar based on the number of overlapping citations made. Patent citations can offer a detailed description of the technology patented and can relate patents that build upon the same underlying technological trajectory. This is a widespread idea originating in Jaffe et al. (1993) and further investigated in Jaffe et al. (2000). Jaffe et al. (1993)'s experiment on the geographic localization of spillovers, used citations as a paper trace for knowledge flows from cited inventors to citing inventors.

My model of interrelated renewal decisions is a better representation of reality when patented innovations can be considered strategic complements. Survey evidence shows firms interacting strategically and using patenting and patent renewal to enforce proprietary rights and appropriate additional returns.[5]

To empirically implement my model for the probabilities of patent renewal I show that it can be recast in terms of a interdependent linear probability model. This in turn can be viewed as a spatial Cliff and Ord (1973, 1981) type model, where space refers to technological space rather than geographical space. More specifically, the model can be estimated as a linear probability model with a spatial lag. In my estimation, I use the instrumental variables technique to try to overcome spurious dependence in renewal arising from: obsolescence, unobserved heterogeneity in demand and cost factors, or information flows from predecessor to successive inventions. I find that there is significant positive dependence in the renewal decisions

[5] According to Cohen et al. (2000) the main explanations for strategic renewal are the prevention of copying, patent blocking, use of patents in negotiations and prevention of law suits.

for patents related at origin.

A possible explanation for the interdependence in renewals that I find in this study is that it is driven by firm strategic behavior. Under asymmetric information, an established firm may have better knowledge of market demand. In this case, the firm may find optimal to renew a patent as a signal to deter an entrant, even if immediate returns do not exceed the maintenance fees. My theoretical framework permits estimation in the case of asymmetric information about the demand in a certain market. Under perfect information, Cohen et al. (2000) discusses scenarios where a firm would use renewal to protect its own market. As a defensive tactic in order to enforce its own innovation a firm might patent several substitutes to its product, known as patent "fencing". Alternatively, a firm may build up a large patent portfolio or "thicket" to prevent a rival firm from entering the market altogether, which is a strategy called patent blocking. Further, thickets of patents are used to force innovators to share rents through cross-licensing. Anecdotal evidence exists about firms building large patent thickets and using them in negotiations and the prevention of lawsuits .[6] All these cases predict dependence in renewals, which when ignored leads to biased estimates of the impact proxies for patent quality have on the renewal decision.

The organization of the paper is as follows. The next section defines the model, discusses some related literature and the meaning of dependence in this

[6]Firms in semiconductors, electronics and computers negotiate based on the relative heights of the stacks of all their related patents and they license entire portfolios for a technology field, Grindley and Teece (1997) and Hall and Ziedonis (2001)

paper. Section 3 introduces the econometric framework, and Section 4 presents the data and introduces the estimation model for my application. Section 5 presents the econometric results, and Section 6 concludes.

4.2 A Model For Patent Renewal Under Interdependence

4.2.1 Background

My work contributes to the empirical strand of literature that uses patent renewal to uncover characteristics of the value of intellectual property.[7] Economists are interested in private returns to patents since they represent incentives firms face for undertaking investments in R&D. Bessen and Meurer (2008) compare the estimated gross private benefits of patents to litigation costs. The authors show that by the end of the 1990s the two were of the same magnitude, and call for some improvement in property protection to reduce litigation costs. Investors are also interested in private returns to patents since this information helps them make investment decisions. Bessen (2009) and Hall et al. (2005) try to evaluate empirically the knowledge stock of firm, which is an intangible asset. The authors proceed by estimating the impact of patent rents and other knowledge assets on the firm's market value. Bessen (2009) find a 6% increase in market value coming from patent

[7]The patent renewal literature falls generally into two categories: the theory of optimal policy for incentive schemes to promote investment in R&D, and the empirical estimation of the return to R&D. My work does not focus on the former strand which includes Kremer (1998), Scotchmer (1999) Cornelli and Schankerman (1999), and Baudry and Dumont (2006). These authors study methods to implement incentives to R&D by requiring maintenance fees.

rents per R&D, which is a measure of the subsidy that patents provide to R&D, in the computers and other machinery industry. Hall et al. (2005) find that if the "quality" of patents of a firm increases so that on average these patents receive one additional citation, the market value of the firm would increase by 3%.

Most studies use "hedonic" econometric models to estimate patent returns, where observed patent and owner characteristics proxy the quality of a patent. In addition, the decision to pay the renewal fee is used to pin down the dollar counterpart for these estimates. These models start with Schankerman and Pakes (1986) who assume that the likelihood of renewal should increase with the associated profit to the patent at any age. Pakes and Schankerman (1984) show that renewal establishes a rough correspondence to how appropriable revenues are to R&D. This literature considers returns to patent rights as a form of "subsidy" to R&D. Lanjouw et al. (1998) use renewal and application data and measure the value of patent protection without offering an upper bound for the returns to patents expiring at the maximum term. Harhoff et al. (2003) do not have the upper truncation limitation, since they use survey data instead when modeling the value of patents. Other work of Bessen (2008), based on patent renewal data across countries, determined the extent of monopolistic patent rents to be 3% of R&D. In a later study Bessen (2009) finds that the weighted mean return across industries is 18% of R&D, when adjusting for the upper tail bias, with an extreme of 80% for the pharmaceutical industry. Baudry and Dumont (2009) show that the cost to society of patenting is ten-fold the amount collected in patent renewal fees.[8] In the above mentioned

[8]It is worth mentioning that a patent need not be a source of incumbency rents, when the

empirical studies, as well as in Lanjouw and Schankerman (2004) and Hall et al. (2005), the value of patents is examined through proxy variables for "quality" such as: the number of references made and of citations received, the patent's outcome in litigation and the size of their international patent family. I will be using some of these standard proxies for patent quality in my estimation.

A common approach in the previous literature is to model the firm's decision to renew a patent as independent of the decision to renew other patents. One study in the patent renewal literature of Liu et al. (2008) draws upon evolutionary economics and shows that patent renewal is more likely if the patent belongs to a sequence of patents. The authors consider patents to be part of a sequence if they are "related patents".[9] The authors show that sequential innovation within a firm is an additional mechanism that enhances the value of US pharmaceutical and biotechnology patents. A study outside the renewal literature where the connection between patents matters is that of Belenzon (2006) who looks at the private market value of R&D spillovers. He finds evidence that private returns to innovation increase when spillovers feed back into the sequential research of the initial investor, but

opportunity cost equals its value (e.g. if it can be resold for a price that equals the value to the owner.)

[9]The authors use the USPTO definition for related patents: a focal patent, its divisional patents (if applicable), and its continuation-in-part patents (if applicable). When a company files an application that contains technologies appropriate for more than one patent, and the patent examiner asks the company to divide the application into multiple applications to obtain *divisional patents*. While, *continuation-in-part patents* cover innovations that build upon, but also add substantially new knowledge to the innovations of the original parent patents.

if others benefit from the advancements returns decrease. Firms who internalize spillovers from innovations are found to invest more in R&D. Another study where the closeness among patents matters is that of Bloom et al. (2007). The authors use a range of firm performance indicators (market value, patents, productivity, and R&D) and show that both technology and product market spillovers are significant. However, using a data panel of US firms over 1980-2001 they find under-investment in R&D from the social perspective, since the technological spillovers exceed market rivalry effects.

4.2.2 Dependence

Liu et al. (2008) is the only previous study that shows that the decision to renew a patent takes into account other related patented innovations. In their study the renewal probability increases if the patent is part of a sequence of related patented innovations within the firm. I consider how renewal depends on the decision to renew, and implicitly returns, of other genealogically related patents within and across firms. My approach brings new insights into the decision process showing that assignees choose to renew patents interdependently. To capture the interdependence in renewal I make use of Spatial Autoregressive Model (SAR). The SAR model consists of introducing in a linear model a spatial lag of the dependent variable to be used as an explanatory variable. The spatial lag is a weighted average of the other patents' renewal status based on a distance metric. In my case, the distance metric captures similarity between patents' scope, and not geographical location.

The distance between patent i and j, d_{ij}, is defined in the following manner:

$$d_{ij} = \frac{(\text{citations made by patent } i + \text{citations made by patent } j)/2}{\text{overlapping citations made between patent } i \text{ and patent } j} \tag{4.1}$$

The more overlap in the number of citations made exists between two patents, the "closer" they are found to be by my metric.

This structure, based on "prior art", relates patented innovations and is particularly appropriate for my setting: the computers and data processing system industry. Firms in this industry are engaged in rapidly advancing cumulative technologies. Due to the pace of the development, it is very likely that a new product overlaps with technologies previously or simultaneously patented by the same or external parties. This metric is in line with Liu et al. (2008)'s finding that patents are more likely to be renewed if they are part of a citation sequence. They show in their analysis that the likelihood of renewal for each patent in a citation-determined sequence is correlated, as high as 0.98, with the renewal of the previous patent. Bloom et al. (2007) define technological closeness based on the firms' distribution of patenting over technological fields, and product market closeness based on the industries of the firms' sales activity. The measures of technological and product closeness they provide are more broad since these measures are at the firm level. My work on the other hand uses a more detailed measure of similarity at the patent level.

Moffitt (2001) raised a criticism about there being an identification problem between correlation of outcomes within a group arising from interaction, and correlations arising for other reason, in particular correlated unobservables. Lee (2007) addresses this criticism and shows that the parameter on the spatial lag is identified when different groups have different number of members. In my case there is variation in group size, but the identification might be weak when the number of group members is large. Moreover, disregarding interdependence in empirical models can lead to biased estimates and incorrect inference regarding the parameters of interest.

4.2.3 Model

This section introduces a static model for patent renewal behavior, in which firms' renewal decisions are interdependent. The model corresponds to a game where the renewal decision for a patent depends upon the renewal decision of similar patents. In this game one player's decision affects the incentives of other players. The measure of "closeness" capturing similarity of patents is based on the overlap in references made to respective patents, and will be discussed in more detail below. The set up of the game considers simultaneous patent renewal decisions of two or more players. Such games where players' decisions mutually reinforce, rather than offset, one another have the supermodularity propriety.

Suppose there are n patents, and let $R = (r_1, r_{2,}..., r_n)'$, with $r_i \in \{0, 1\}$ denote the vector of patent renewal choices, where $r_i = 1$ if the i-th patent has been renewed. While some firms are assigned more than a patent, I assume that the decision to renew patent i is made at the patent level.

Further, I consider the genealogy of patents and I make parametric assumptions over the form of the metric that describes how close patents are in the innovation space. I measure the technological "closeness", between patents i and j to be a function of the number of common references made. This metric describes patents to be assigned for an innovation with a similar underlying functionality if they cite directly the same patent. Intuitively, the more overlap there is in the patents' citations made, the closer the patents are in the technological space and the more similar the innovations are considered to be. The network structure is described by an $n \times n$ matrix $W = (w_{ij})$ with $w_{ii} = 0$, and where the weights w_{ij} reflect the "closeness" between patent i and j. I define the weights

$$w_{ij} = \frac{\text{overlap in citations made by patent } i \text{ and patent } j}{\sum_j \text{overlap in citations made by patent } i \text{ and patent } j} \quad \text{for } i \neq j \qquad (4.2)$$

which ensures that the weights are increasing in the inverse distance metric.

The realized profits for patent i are a function of its renewal decision r_i and the renewal decisions for other patents. It proves convenient to introduce the notation $R_{-i} = (r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_n)'$. The revenue, Rev_i, and the cost, C_i, of renewal of the i-th patent are, respectively, given by

$$\begin{aligned} Rev_i &= Rev_i(r_i, R_{-i}, \xi_i^{r_i}) = \begin{cases} a_i + b\sum_{j \neq i} w_{ij} r_j + \xi_i^1 & \text{if} \quad r_i = 1 \\ \xi_i^0 & \text{if} \quad r_i = 0 \end{cases}, \\ C_i &= C_i(r_i, \epsilon_i^{r_i}) = \begin{cases} c_i - \epsilon_i^1 & \text{if} \quad r_i = 1 \\ \epsilon_i^0 & \text{if} \quad r_i = 0 \end{cases}, \end{aligned}$$

or more compactly

$$\begin{aligned} Rev_i &= a_i r_i + b \sum_{j \neq i} w_{ij} r_j r_i + \xi_i^{r_i}, \\ C_i &= c_i r_i - \epsilon_i^{r_i} \end{aligned}$$

where a_i and c_i are fixed (non-stochastic) and observed by all patent-holders and the econometrician. One pays these renewal costs c_i and obtains benefits a_i and b

only if renewal happens $r_i = 1$. If the revenues and costs for each patent i, e.g. a_i and respectively c_i, were determined exclusively by the patent's own quality and characteristics it would be the case that $b = 0$. However, if patents are strategic complements we would expect to have $b > 0$ and the returns to patent i to increase in the renewal of the other patents.

In the above, a_i measures revenues based on proxies for the quality of the patent such as citations received, citations made, claims, and also based on other observed characteristics such as origin of inventor, size of entity, R&D spending, technological class fixed effects, time fixed effects etc.; while, c_i is a measure of the non-stochastic renewal costs such as legal and royalty fees that apply; b measures the strategic effect;[10] and w_{ij} are weights measuring similarity or technological closeness as defined by (4.2).

The patent and decision specific random components $\xi_i^{r_i}$ and $\epsilon_i^{r_i}$ are unobserved by the econometrician or other patent-holders. Assignee i knows only his own $\xi_i^{r_i}$ and $\epsilon_i^{r_i}$, but he does not observe $\xi_j^{r_j}$ and $\epsilon_j^{r_j}$ for $j \neq i$. I assume that the stochastic components $\xi_i^{r_i}$ and $\epsilon_i^{r_i}$ are independent of $\xi_j^{r_j}$ and $\epsilon_j^{r_j}$ for $j \neq i$. Further, I assume that $E\xi_i^{r_i} = 0$, $E\epsilon_i^{r_i} = 0$ and that $\xi_i^{r_i}$ and $\epsilon_i^{r_i}$ are identically distributed. Even though ξ_i^0 and ϵ_i^0 can be set to zero, I am following the random utility theory and allow for the errors to be non null.

Profits associated with the i-th patent are then given by

$$\Pi_i(r_i, R_{-i}, \xi_i^{r_i}, \epsilon_i^{r_i}) \;=\; Rev_i(r_i, R_{-i}, \xi_i^{r_i}) - C_i(r_i, \epsilon_i^{r_i}) \qquad (4.3)$$

[10]Alternatively, in a non strategic model b is "just a reflection of spatially correlated technological opportunities" as stated by Griliches (1998)

$$= \begin{cases} a_i - c_i + b\sum_{j\neq i} w_{ij}r_j + \xi_i^1 + \epsilon_i^1 & \text{if} \quad r_i = 1 \\ \xi_i^0 + \epsilon_i^0 & \text{if} \quad r_i = 0 \end{cases}, \tag{4.4}$$

or more compactly

$$\Pi_i(r_i, R_{-i}, \xi_i^{r_i}, \epsilon_i^{r_i}) = a_i r_i + b\sum_{j\neq i} w_{ij} r_j r_i - c_i r_i + \xi_i^{r_i} + \epsilon_i^{r_i}. \tag{4.5}$$

According to Bulow et al. (1985), if the inventors regard the innovations as strategic complements marginal profits have to be increasing in the decision variable of the others. A positive interdependence in renewals, e.g. positive b, insures that this condition holds.[11] The case of strategic complements arises when there are sufficient strong aggregate increasing returns to scale and/or the demand curves for the firms' products have a sufficiently low own-price elasticity to allow for Bertrand competition.

In this model, as in other discrete choice literature, player heterogeneity is captured by private information. Let $F_i = \sigma(\xi_i^0, \xi_i^1, \epsilon_i^0, \epsilon_i^1)$ be the patent-holder's information set for the i-th patent. This information is revealed before any action is made, and is observed only by the ith patent assignee and not by assignee j, $j \neq i$, or by the econometrician.[12] Under incomplete information, the players make the decision simultaneously without observing rival choices. The assignee evaluates the returns to patent i and makes its decision r_i by maximizing the conditional expected

[11]Under strategic complementarity, in a two player game one needs to check that the following condition holds: $\frac{\partial^2 \Pi_i}{\partial r_i \partial r_j} = b$ is positive. For simplicity of notation, in the latter formula I am using renewal as a continuous variable to check the condition, even though renewal is a discrete variable.

[12]Langinier (2004) studies strategic patent renewal under asymmetric information; while Pakes (1986) and Lanjouw et al. (1998) study post-patent learning duration under incomplete information.

payoff of renewing patent i:

$$E\Pi_i(r_i, R_{-i}, \xi_i^{r_i}, \epsilon_i^{r_i}|F_i) = ERev_i(r_i, R_{-i}, \xi_i^{r_i}|F_i) - EC_i(r_i, \epsilon_i^{r_i}|F_i). \tag{4.6}$$

Observe that

$$ERev_i(r_i, R_{-i}, \xi_i^{r_i}|F_i) = a_i r_i + b\sum_{j\neq i} w_{ij} E(r_j|F_i) r_i + \xi_i^{r_i}, \tag{4.7}$$

and that $\mathrm{EC}_i(r_i|F_i) = c_i r_i + \epsilon_i^{r_i}$. Renewal happens when:

$$E\Pi_i(r_i = 1, R_{-i}, \xi_i^1, \epsilon_i^1|F_i) \geq E\Pi_i(r_i = 0, R_{-i}, \xi_i^0, \epsilon_i^0|F_i) \tag{4.8}$$

or equivalently

$$a_i + b\sum_{j\neq i} w_{ij} E(r_j = 1|F_i) - c_i + \xi_i^1 + \epsilon_i^1 \geq \xi_i^0 + \epsilon_i^0 \tag{4.9}$$

since $E\Pi_i(r_i = 0, R_{-i}, \xi_i^0, \epsilon_i^0|F_i) = \xi_i^0 + \epsilon_i^0$, by functional form. Due to the binary nature of the renewal variable $E(r_j|F_i) = \Pr(r_j = 1|F_i)$ so that

$$a_i + b\sum_{j\neq i} w_{ij} \Pr(r_j = 1|F_i) - c_i + \xi_i^1 + \epsilon_i^1 \geq \xi_i^0 + \epsilon_i^0. \tag{4.10}$$

In Appendix B it is shown that

$$\Pr(r_j = 1|F_i) = \Pr(r_j = 1). \tag{4.11}$$

The intuition in equality (4.11) is that the content of the information set F_i is sufficient to permit player i to determine objectively the unconditional probability assessments for the possible value of r_j, but that player i has no additional information that is private to player j.

Substituting (4.11) into (4.10) yields

$$\xi_i^0 - \xi_i^1 + \epsilon_i^0 - \epsilon_i^1 \leq a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i. \tag{4.12}$$

The above equation can furthermore be rewritten as

$$\eta_i \leq a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i \tag{4.13}$$

where $\eta_i = \xi_i^0 - \xi_i^1 + \epsilon_i^0 - \epsilon_i^1$. By previous assumptions the η_i are statistically independent in i and identically distributed. Also, observe that under the maintained assumptions $a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i$ is non-stochastic.

Now, let $G(.)$ denote the c.d.f. of η_i, then it follows from (4.13) that the probability of patent i being renewed is

$$\begin{aligned} \Pr(r_i = 1) &= \Pr(\eta_i \leq a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i) \\ &= G(a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i) \end{aligned}$$

Motivated by Heckman and Snyder (1997), we need η_i to be i.i.d. uniformly distributed which is satisfied under the assumption that $\xi_i^{r_i}$ and $\epsilon_i^{r_i}$ are i.i.d . Since $E\eta_i = E\xi_i^0 - E\xi_i^1 + E\epsilon_i^0 - E\epsilon_i^1 = 0$, it follows that if η_i is uniformly distributed, it has to be distributed symmetrically around 0. Suppose η_i is uniformly distributed in the interval $[-M, M]$, then

$$\text{G(z)=}\begin{cases} \frac{z+M}{2M} & \text{for} \quad -M \leq z \leq M \\ 0 \text{ or } 1 & \text{else} \end{cases}$$

so that

$$\Pr(r_i = 1) = \begin{cases} 0 & \text{for} \quad a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i \leq -M \\ a_i^* + b^* \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i^* + \frac{1}{2} & \text{for} \quad -M \leq a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i \leq M \\ 1 & \text{for} \quad a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i \geq M \end{cases}$$

with $a_i^* = (a_i)/(2M)$, $b^* = b/(2M)$ and $c_i^* = c_i/(2M)$. Consequently, the parameters in the model can only be identified up to the scale M.

It is sufficient for the scale parameter M to satisfy

$$M \geq \max_i \left| a_i + b \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i \right|$$

to guarantee that

$$a_i^* + b^* \sum_{j \neq i} w_{ij} \Pr(r_j = 1) - c_i^* + \frac{1}{2} \in (0, 1). \tag{4.14}$$

Now let $b^* = \lambda$ and $a_i^* - c_i^* + \frac{1}{2} = x_i\beta$, and rewrite the model

$$\Pr(r_i = 1) = \lambda \sum_{j \neq i} w_{ij} \Pr(r_j = 1) + x_i\beta \tag{4.15}$$

where x_i is a vector of observed exogenous and non-stochastic characteristics of patent i. The patent characteristics in x_i and the constant are used to model linearly the benefits a_i and the costs c_i.

On a side note, the model can be written as $a_i^* - c_i^* + \frac{1}{2} = (1-\lambda)G(x_i\beta)$, where $G(.)$ was defined above as the uniform c.d.f., to satisfy the condition $\Pr(r_i = 1) \in (0, 1)$.

4.3 Estimation Methodology

The discrete choice model that I am estimating, i.e.,

$$\Pr(r_i = 1) = \lambda \sum_{j=1, j \neq i}^{n} w_{ij} \Pr(r_j = 1) + x_i\beta, \qquad i = 1, \ldots, n, \tag{4.16}$$

represents a linear simultaneous equation system. The solution it generates is unique, provided that $I - \lambda W$ is non-singular.[13] More specifically, let $P = [\Pr(r_1 =$

[13] The linear model in this case has a single equilibrium, while a nonlinear discrete choice model describing strategic complements can have multiple equilibria.

$1), \ldots, \Pr(r_n = 1)]'$, and let $X = [x_1', x_2', ..., x_n']'$, then $P = (I - \lambda W)^{-1} X\beta$. One sufficient condition for $I - \lambda W$ to be non-singular is that the spatial weights matrix is row normalized and that λ is $\in$ (-1,1); for more general conditions see, e.g., Kelejian and Prucha (2010).

For estimation we reformulate model (4.16) as an interdependent linear probability model. For that purpose let

$$u_i = r_i - P(r_i = 1) = r_i - E(r_i = 1) \tag{4.17}$$

then $Eu_i = 0$ and $Var(u_i) = P(r_i = 1)(1 - P(r_i = 1))$. Model (4.16) can then be re-written as the regression equation $r_i - u_i = \lambda \sum_{j \neq i} w_{ij}(r_j - u_j) + x_i\beta$, or

$$r_i = \lambda \sum_{j \neq i} w_{ij} r_j + x_i \beta + v_i \tag{4.18}$$

with $v_i = u_i - \lambda \sum_{j \neq i} w_{ij} u_j$. It proves helpful to rewrite the above model in matrix notation as:

$$R = \lambda W R + X\beta + v \tag{4.19}$$

$$v = (I - \lambda W) u \tag{4.20}$$

where $R = [r_1, \ldots, r_n]$, $v = [v_1, \ldots, v_n]$, and $u = [u_1, \ldots, u_n]$. By construction $Eu = 0$, $Euu' = \Omega_u = (\omega_{ij,u})$ with $\omega_{ii,u} = P(r_i = 1)(1 - P(r_i = 1))$. The theoretical model leaves the covariances $\omega_{ij,u}$ for $i \neq j$ unspecified. Observe further that $R = (I - \lambda W)^{-1} X\beta + u = P + u$, and that $ER = P$, and $VC(R) = VC(u) = \Omega_u$.

In the generalized linear probability model (4.19) the interdependence in renewal decisions is reflected by $\overline{R} = WR$. Model (4.19) is essentially of the form of a SAR model considered in Kelejian and Prucha (2007), with $Ev = 0$, $Evv' =$

$(I-\lambda W)\Omega_u(I-\lambda W')$. In the spatial literature the r.h.s. variable $\overline{R} = WR$ is called a spatial lag. Note that $COV(\overline{R}, v) = E\left[Wuu'(I-\lambda W')\right] = W\Omega_u(I-\lambda W') \neq 0$, i.e., $\overline{R}$ is endogenous. To estimate the model we apply the generalizes spatial two stage least squares (GS2SLS) estimator of Kelejian and Prucha (1998, 2007), and refer to (4.19) as a spatial linear probability model (SLPM). As suggested in the above cited literature, I will use $H = (X, WX, W^2X)$ as instruments, which include, in addition to the covariates X first and second order spatial lags (weighted averages) of the covariates. Maintaining assumptions as in Kelejian and Prucha (1998, 2007) the GS2SLS estimator provides consistent estimates for the vector of unknown parameters $\theta = (\lambda, \beta')'$. Furthermore, the estimator is asymptotically normal, and the asymptotic variance covariance matrix of the GS2SLS estimator can be estimated consistently by

$$\hat{\Phi} = n^2(\widehat{Z}'\widehat{Z})^{-1}Z'H(H'H)^{-1}\hat{\Psi}(H'H)^{-1}H'Z(\widehat{Z}'\widehat{Z})^{-1}$$

where $Z = [WR, X]$, $\widehat{Z} = H(H'H)^{-1}H'Z$ and $\hat{\Psi}$ is a spatial heteroskedasticity and autocorrelation consistent (SHAC) for $\Psi = n^{-1}H'Evv'H$ as discussed in Kelejian and Prucha (2007) and Kim and Sun (2010). In more detail, the typical element of Ψ is $\psi_{rs} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ir}h_{js}\sigma_{ij}$,where $\sigma_{ij} = Ev_iv_j$, while the typical element of its estimate $\hat{\Psi}$ is: $\hat{\psi}_{rs} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ir}h_{js}\hat{v}_i\hat{v}_jK(d_{ij}/d)$ where $\hat{v}_i$ denote 2SLS residual, $K(.)$ is a kernel function, and the d_{ij} denotes the distance defined in (4.1). I use

$$K(x) = \begin{cases} (1-x)^2 & \text{, for } 0 \leq x \leq 1 \\ 0 & \text{, otherwise,} \end{cases}$$

as my kernel function and the bandwidth is chosen such that $d = [n^{1/3}]$.[14]

Moreover, the error term in equation (4.19) is derived based on the expectation of the unobserved renewal decisions made by other patent-holders. When the decision of some to renew is observed, the expectation becomes the actual decision and the error u_j is zero, for those j patent-holders. In the extreme case, when i's decision to renew comes after that of all the other holders of similar patents, the regression equation becomes: $r_i = \lambda \sum_{j \neq i} w_{ij} r_j + x_i \beta + u_i$ such that the errors in (4.18) are instead $v_i = u_i$. Nevertheless, the estimation procedure from above still holds.

4.4 Data Description and Empirical Model

In the follow I apply the above model and estimation strategy to analyze a subset of US patented inventions. The underlying patent data are available to the public on the United States Patent and Trademark Office (USPTO) website. The USPTO is the US government agency responsible for examining patent applications and issuing patents. My sample consists of 28,138 US utility patents issued from 1994 through 1997.[15] My sample ends in 1997, since it is the last year, at the time of

[14] [z] stands for nearest integer less than or equal to z.

[15] There are three patent categories: Utility, Plant and Design. Utility patents is the largest category of issued patents out of the three. Section 101 of the U.S. Patent Act sets forth the general requirements for a utility patent: "Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvements thereof, may obtain a patent, subject to the conditions and requirements of this title." Design patents differ from utility patents in that a design patent covers only the ornamental appearance of a useful

this writing, for which I can observe the final renewal decision for the patents. My data includes only patents classified under technological classes denoting computers and related innovations, which are listed in the Appendix A1. These twenty out of over 400 technology classes were selected according to the category Computer Hardware and Software in Hall et al. (2001).

One part of my data consists of data that I collected on patent renewal status, entity type and a detailed listing of citations to present date, from the USPTO website.[16] The other source of data is the NBER patent data archive, which provided me with additional claims and assignee specific information already collected from the USPTO.[17] [18]

For an invention to be patentable it must be statutory, new, useful, and non-obvious. In the US a patent is awarded for a maximum of 20 years from the application date. In some cases the lifetime can be extended if the issuance process was unnecessarily lengthy. While 20 years is the maximum legal protection, a US patent is required to be renewed three times over its lifetime to maintain its validity. All utility patents issued from applications filed after December 12, 1980 are subject

---

product. Plant patents are a small category of patents which protect certain types of botanical plants, such as flowers, fruits, shrubs and vines.

[16]The data consists of Patent Number, International Class, Inventors, Status, Field of Search, Entity, Assignee, Application Number, Application Date, Issue date, Expiration, US Class, References Made, References Received

[17]See http://www.nber.org/patents/ for the 1999 NBER data, and https://sites.google.com/site/patentdataproject/Home/downloadsfor the 2006 NBER data.

[18]I have not used some of the data found in previous studies (e.g. measure of originality, measure of generality), since it would have decreased the size of my sample.

to fees that keep the patent in force. These maintenance fees are due by 3 1/2, 7 1/2, and 11 1/2 years after the patent is granted, to remain in force after 4, 8, and respectively 12 years. In case of a missed payment, the assignee can petition and have the patent reinstated for a penalty fee in addition to the maintenance one. I will be modeling the *decision to renew* a patent, at each of the three renewal stages, based on patent and assignee characteristics, as well as the renewal decision for other similar patents. The binary variable indicating renewal can be constructed based on the patent issue and expiration date. In case the patent expired at 4, 8 or 12 years, then the renewal decision becomes zero from that stage and on. There is sufficient variation in patent renewal in my sample, such that 11,262 out of 28,138 patents expired by 2010, see Table 4.1.[19]

Table 4.1: Sample Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Renewed at age 4	28138	.9144929	.2796399	0	1
Renewed at age 8	28138	.7475656	.4344168	0	1
Renewed at age 12	28138	.5997583	.4899559	0	1
Cites received	28138	26.89356	37.3611	0	1015
Claims	28138	16.17208	13.65686	1	309
Cites made	28138	10.29483	10.72908	1	270
Entity large	28138	.8959059	.305388	0	1
Unassigned	28138	.0544815	.2269693	0	1
Assignee: US non-Gov	28138	.5640415	.4958905	0	1
Assignee: non-US non-Gov	28138	.3656621	.4816239	0	1
Assignee: US individual	28138	.0028076	.0529132	0	1
Assignee: non-US individual	28138	.0014926	.0386066	0	1
Assignee: US Gov	28138	.009951	.0992587	0	1
Assignee: non-US Gov	28138	.0015637	.0395137	0	1
Log R&D scaled by patent count	14061	-.3924817	.797324	-6.319969	1.941478

The US fee schedule was created with the purpose to support the US Patent

[19]See Figure 1. and Figure 2. for year and age at expiration

Office and protect small entities. A small entity is an independent inventor (or inventors), a small business, or a non-profit. The normal government definition of a small business is a company with fewer than 500 employees. So many high tech start-ups are small entities. A non-profit could be for example a university, a scientific organization, or many other educational organizations. The fees are differentiated based on the size of the entity to which it is issued, e.g. small or large, and also the age of the patent, see Appendix A2 for how fees are structured today.

In my model for patent renewal (4.18) I use patent and owner characteristics as covariates, see Table 4.1 for Summary Statistics. I follow the literature that has considered claims, citations received, citations made, and technological classes as proxies for the quality of a patent, and that has used entity size, institutional classification of the owner, and country of origin of the first inventor as other factors which make renewal more likely. In my sample, on average a patent cites 10 patents, receives 26.5 citations, and shares some feature with 69.7 other patents.[20] Further, I discuss each of the covariates I use when modeling the decision to renew a patent or not.

From the NBER source I use data on the *number of claims* for each patent to measure the extent of the innovation. Claim data is an important tool that describes the scope of the protection conferred by a patent and is used during prosecution and litigation. Moore (2005) suggest that the more claims a patent has the broader the degree of protection. Lanjouw and Schankerman (2004) show for the majority

[20]See Figure 3, Figure 4. and Figure 5. for the distribution of the number of backwards references, forward references and connectivity.

of industries, including the computer industry, the number of claims is the most important indicator of research quality. In case corrections are made on claims that are too narrow or too broad, a patent is reissued. Other cases under which a patent is reissued are failure to correctly reference prior documents. I eliminate from my sample patents that were reissued.

Individual patent data contains high quality information on the network structure citations create. Citations are informative links between patents, and are used as an indicator of knowledge spillovers in Caballero and Jaffe (1993) and Hall et al. (2001). Researchers are required to cite patents upon which they build. Others are also allowed to suggest additional references to an existing patent. The list of references conveys how a specific technology infringes on the rights of others. Ultimately the decision on which patents to cite is made by the patent agent during the examination.

I use the collected data on *count of citations received* by each patent until 2010 as an indicator for the cited patent's importance, impact or even revenues. These citations are identified by parties other than the citing inventor and may convey valuable information about the size of the technological "footprint" of the cited patent. Jaffe et al. (2000) provide evidence that there is a significant correlation between the number of citations a patent received and its importance (both economic and technological) as perceived by the inventor. According to the statement of the Office of Technology Assessment and Forecast (OTAF) from the USPTO (1976): "[...] if a single document is cited in numerous patents the technology revealed in that document is apparently involved in many developmental efforts. Thus the number of times a patent document is cited may be a measure of its technological significance."

In terms of empirical evidence, Lanjouw and Schankerman (2004) use citation counts as weights to create quality adjusted measures of patents counts. The authors show that an index of research quality is more correlated with R&D than patent counts alone. Citations counts are inherently truncated, for some patents even after 50 years. In another study by Hall et al. (2005) the number of subsequent citations received by a firm's patent was used to get a measure of R&D success. The authors claim that the prime citation years are roughly 3 to 10 and that those years give an accurate estimate of lifetime citations. In light of this study, I can assume the truncation problem to be negligible since I count citations over 13 to 16 years from the issue date.

Also, I include the *count of citations made* as another proxy variable for the quality and costs of a patent, see Jaffe et al. (2000) for evidence from a survey of inventors on the role of citations.[21] One could argue that patents citing less are more reliant, and perhaps more "original". Moreover, Harhoff et al. (2003) note that lawyers say a patent application seeking to protect an invention with broad scope might induce the examiner to delineate the patent claims by inserting more references to the relevant patent literature. A counter point is that an assignee

[21] Tranjenberg et al. (1997) build a measure of originality referring to the percentage of citations made by a patent that belong to a certain patent class, out of all technology classes. By this measure if a patent cites previous patents that belong to a narrow set of technologies the originality score will be low.

might have to pay more royalty fees when citing patents granted to other firms. Given these arguments, it is not clear whether the impact on renewal of the number of citations made should be positive or negative.

The scope of a patent may be an important determinant of the efficacy of patent protection. While it is difficult to measure, in my estimation, I use *3-digit technology classes* to control for the technological opportunities which would make patents granted in those technological areas more valuable. In my regressions, I include *cohort fixed effects*, namely dummy variables for the year the patent was granted, to eliminate the effect of possible changes in legislation that would disrupt returns to patents and costs of renewal.

I include characteristics of the patent owner such as *institutional status*, and *geographical location of the first inventor* (i.e. country) to isolate some of the effect of intra institution and intra country propensity to cite. Finally, I control for the *size of the entity.* Table 4.2 summarizes that large entities have a higher propensity to renew their patents, even though renewal fees are higher for such entities. While this could imply that larger entities have higher quality patents, it could also be that the renewal fees are negligible. If the latter is true there should be a more optimal restructuring of the fees. However, this latter explanation does not seem to be supported by facts, since the total renewal fees amount to more than 7,500 USD for large entities and the average firm in my sample is granted 100-150 patents per year. Moreover, for the sample I analyze some firms have granted more than 700 patents in one year. These numbers are an underestimate for the total number of patents granted to a firm, since they do not include the patents in the other technological classes. Renewal fees could add up to millions of dollars over the lifetime of a cohort and therefore would not be negligible for a firm with hundreds of patents.

Table 4.2: Renewal Summary Statistics by Entity Size

Variable	Large Entity	Small Entity
First renewal	.9287556	.7917378
Second renewal	.7725019	.5329464
Last renewal	.627395	.3618983
No. obs.	25209	2929

These covariates do not take into account that patented innovations share the same functionality with other patents and are likely to have interdependent renewal values. My sample consists of patents that have at least one citation in common with another patent.[22] Explicitly, I include the *weighted average of the renewal decisions* as a determinant of patent renewal. The weighted average of the renewal decisions captures returns to a patent resulting from the specific interaction with other patents described as related at origin.

4.5 Empirical Results

In Table 4.3, I report results on patent renewal at each of the three stages depending on the decision to renew of other patent assignees. For comparison, in Table 4.4 I show the results for a model where the decision to renew a patent is

[22] I excluded 1010 patents from the original sample that did not meet this criterion.

taken independently from other similar patent. [23] Table 4.3 shows 2SLS parameter estimates of the spatial linear probability model in (4.18) for patents' "contemporaneous" renewal decisions. The estimates in the second and third columns of the table are conditional on having renewed the patent at the previous stage. Therefore, the estimation sample includes only patents that have not expired. Moreover, the renewal of patents is only dependent on decisions made for patents that have also renewed at the previous stage, i.e. have not yet expired.

In Table 4.3 the explanatory variable of interest is the weighted average of the observed and unobserved decision to renew patents that are related technologically. The weighted average of other similar patents renewal explains a large percentage, 14% to 28%, of the propensity to renew a patent. Graphs of the predicted probabilities from these regressions can be found in the Appendix A in Figure 6, Figure 7 and Figure 8. For these 2SLS regressions, I computed the F-statistic against the null that the excluded instruments are irrelevant in the first-stage regression. The F-statistics are unilaterally larger than 10 in my case, and show no evidence that the instruments are weak.[24]

In order to put these estimates in context, I include a model that assumes independent renewal decisions, see the estimates in Table 4.4. In most cases, the absolute value of the coefficients are larger than in the model with interaction in

[23]Renewal timing is exogenous, since renewal follows every four years from the issue date of a patent. While the application date could be endogenous, the issue date will be exogenously determined by the patent office.

[24]This is Staiger and Stock (1997) rule of thumb for models with one endogenous regressor.

renewals. I stress the existence of an upward bias in the estimation of the coefficients of a model that assumes independent decision making.

Table 4.3: Renewal Decisions under Dependence

Renewal	1st	2nd	3rd
W*renewal	0.28026***	0.18386***	0.14637***
	(0.07449)	(0.03242)	(0.02254)
Cites received	0.00029***	0.00067***	0.00064***
	(0.00004)	(0.00006)	(0.00006)
Cites made	0.00042***	-0.00029	-0.00008
	(0.00011)	(0.00019)	(0.00022)
Claims	0.00086***	0.00124***	0.00121***
	(0.00010)	(0.00015)	(0.00017)
Entity Large	0.13179***	0.17428***	0.15426***
	(0.00922)	(0.01106)	(0.01285)
Unassigned	-0.02795*	-0.05527**	-0.07004***
	(0.01356)	(0.01716)	(0.01935)
US non-Gov	-0.01055	-0.04686***	-0.03163*
	(0.00937)	(0.01290)	(0.01430)
US indiv	0.02283	-0.07423	0.03271
	(0.03674)	(0.05281)	(0.05191)
Non-US indiv	-0.06929	-0.06949	-0.02469
	(0.06304)	(0.08207)	(0.09760)
US Gov	-0.35202***	-0.41567***	-0.43614***
	(0.03403)	(0.04157)	(0.05952)
Non-US Gov	-0.04658	-0.02456	0.12694*
	(0.05687)	(0.07440)	(0.05622)
1st Stage F-stat	14.89	79.08	295.38
R-squared	0.059	0.052	0.068
N	28138	25732	21035
	country FE	country FE	country FE
	cohort FE	cohort FE	cohort FE
	tech class FE	tech class FE	tech class FE

* p<0.05, ** p<0.01, *** p<0.001,
HAC standard errors are reported in parenthesis.

A consequence of my model with interdependent renewals is that the average marginal effect is different from the model that does not consider this interdepen-

Table 4.4: Renewal Decisions under Independence

Renewal	1st	2nd	3rd
Cites received	0.00031***	0.00070***	0.00067***
	(0.00004)	(0.00006)	(0.00006)
Cites made	0.00048***	-0.00024	0.00001
	(0.00011)	(0.00020)	(0.00022)
Claims	0.00087***	0.00126***	0.00123***
	(0.00010)	(0.00015)	(0.00018)
Entity Large	0.13584***	0.17901***	0.15798***
	(0.00842)	(0.01114)	(0.01303)
Unassigned	-0.03149*	-0.05794***	-0.07160***
	(0.01307)	(0.01730)	(0.01956)
US non-Gov	-0.01111	-0.04738***	-0.03190*
	(0.00924)	(0.01300)	(0.01437)
US indiv	0.02038	-0.07833	0.02670
	(0.03598)	(0.05312)	(0.05416)
Non-US indiv	-0.06609	-0.06983	-0.03685
	(0.06400)	(0.08308)	(0.09831)
US Gov	-0.36248***	-0.42687***	-0.44273***
	(0.03082)	(0.04113)	(0.06022)
Non-US Gov	-0.04482	-0.03289	0.12409*
	(0.05813)	(0.07353)	(0.05692)
R-squared	0.053	0.043	0.061
N	28138	25732	21035
	country FE	country FE	country FE
	cohort FE	cohort FE	cohort FE
	tech class FE	tech class FE	tech class FE

* p<0.05, ** p<0.01, *** p<0.001,
Robust standard errors are reported in parenthesis.

dence. Consider my model in (4.16) and rewrite it as:

$$P(r_i = 1) = \sum_{k=1}^{K} S_{i.}(W)x_{.k}\beta_k$$
$$S(W) = (I_n - \lambda W)^{-1}$$

Unlike in a model without a spatial lag , i.e. $\lambda = 0$, where the marginal effect or the derivative of r_i with respect to x_{ik} is β_k, in the interdependent renewal model this effect is not constant: $\frac{\partial P(r_i=1)}{\partial x_{ik}} = S(W)_{ii}\beta_k$, where $S(W)_{ii}$ is the element in the

ith column and ith row of $S(W)$. The effect includes the "feedback loops" where observation i affects observation j also affects observation i as well as longer paths which might go from observation i and j to m and back to i. Another implication of the model under interdependence is that a change in the characteristic for a single patent can also affect the renewal for all other patents, i.e. $\frac{\partial P(r_i=1)}{\partial x_{jk}} = S(W)_{ij}\beta_k \neq 0$.

In Table 4.5, I compare the average marginal effects for the model with interdependence in renewal with the marginal effects for the non-spatial linear probability model. The first three columns contain the average over all patents of the total impact that an equal change in the characteristic k of each patent has on the probabilities of renewing the patents:

$$n^{-1}\iota_n\prime S(W)\iota_n\beta_k = n^{-1}\sum_i\sum_j S(W)_{ij}\beta_k, \tag{4.21}$$

while the last three columns include straightforwardly the ordinary least squares coefficient estimates. The average marginal effects for the model with independent renewal are underestimated in absolute terms since the model ignores the network effect.

The coefficients on claims and citations received are positive, and stay significant and comparable across regressions. We can infer that these two quality measures are significantly associated with the private value of patents. Claims, as discussed, describe the patent quality in terms of breadth of property law protection. Citations received allows me to control for the depreciation of the innovation. In case of obsolescence it would happen that newer technologies replace older technologies, and that the latter receive fewer citations. In terms of marginal effects, an

Table 4.5: Average Marginal Effects under Dependence and Marginal Effects under Independence

Renewal	SLPM 1st	SLPM 2nd	SLPM 3rd	LPM 1st	LPM 2nd	LPM 3rd
Cites received	0.00040	0.00082	0.00075	0.00031	0.00070	0.00067
Cites made	0.00058	-0.00036	-0.00009	0.00048	-0.00024	0.00001
Claims	0.00119	0.00152	0.00142	0.00087	0.00126	0.00123
Entity Large	0.18311	0.21347	0.18050	0.13584	0.17901	0.15798
Unassigned	-0.03883	-0.06770	-0.08196	-0.03149	-0.05794	-0.07160
US non-Gov	-0.01466	-0.05740	-0.03701	-0.01111	-0.04738	-0.03190
US indiv	0.03172	-0.09092	0.03828	0.02038	-0.07833	0.02670
Non-US indiv	-0.09627	-0.08512	-0.02889	-0.06609	-0.06983	-0.03685
US Gov	-0.48909	-0.50915	-0.51034	-0.36248	-0.42687	-0.44273
Non-US Gov	-0.06472	-0.03008	0.14854	-0.04482	-0.03289	0.12409

additional 100 citations received, at the mean, result in a 4-8.2% percent increase in the predicted probability of renewal; while an additional 100 claims is required for an increase of 11.9-15.2% in the same probability.

I include citations made as a proxy for the costs resulting from the payment of royalty fees. I find that the count of citations made have a positive impact at first renewal, and no effect at the next two stages. Since no negative significant effect is found, it would suggest that royalty fees per citations are insignificant on average. A plausible explanation for the positive coefficient at first renewal can be that of an assignee who wrongly believes his invention is valuable, documents well the patent, and also renews it initially, to learn later on about its low value.

One variable with a large explanatory power that is missing from many of the empirical studies is the size of the entity. Being a large firm impacts the probability of renewal by up to 21.3%. A result which supports Bessen (2008)'s finding that small entities have patent with values that are on average less than half as large as the values obtained by large corporations. It is not surprising given the summary statistics in Table 4.2 showing large companies to be more likely to renew their patents at any stage. Besides the explanation that large entities own higher quality patents since they invest more in R&D, some other potential explanations for this higher propensity to renew are that large entities face cheaper legal and royalty cost, and/or build patent patent portfolios that serve strategic purposes. As discussed in the previous section, I believe that there is strong evidence against total renewal fees being modest. Additionally, large entities are known to commercialized more products resulting from patents, and therefore make those patents more profitable. Further, when a small entity licenses a patent to a large entity to increase the returns on that patent, the small entity is required to pay large entity fees.

In terms of the assignee status, patents that are unassigned or assigned to the US Government have a much lower probability to renew, holding the other factors fixed, compared to foreign corporations. While at the second and last renewal, US corporation also have significant lower propensity to renew their patents relative to the foreign counterparts. However, it is expected that US patents assigned to foreign corporations are of a higher value, since most likely they are the result of a selection rather than a random patent application from that foreign country.

A plausible story for the observed dependence in renewals is that of strategic effects from holding patents on the same technology. Since patents enforce market power, a typical oligopolistic price competition with two substitute goods that are strategic complements would explain the result. Furthermore, firms could act strategically about renewal: cross citation used in negotiation and patent infringement lawsuits, patent fencing, patent blocking, etc.

While the spatial lag could capture an omitted variable bias reflecting demand, costs or obsolescence, an instrumented approach would eliminate this a bias. Now, I will relate the dependence found to the time series literature. Heckman (1981) notes that persistence in outcome can be the result of true state dependence or spurious state dependence. In my case, an omitted variable story translates into spurious dependence, and the strategic effects translate into true dependence. It holds under both true and spurious state dependence that: $\Pr(r_i|x_i, R_{-1}) \neq \Pr(r_i|x_i)$. However, under spurious dependence the difference in probabilities is caused by unobserved heterogeneity. When conditioning on the unobserved heterogeneity behind patent i, denoted by ζ_i, the dependence on the renewal of other patents disappears: $\Pr(r_i|x_i, R_{-1}, \zeta_i) = \Pr(r_i|x_i, \zeta_i)$, Since heterogeneity is unobserved and cannot be controlled for, the latter equality could also be achieved by using a set of instruments. Therefore, my identification strategy relies on the instrumental variable technique and provides evidence for the existence of true state dependence.

To disentangle strategic effects, I consider the case when the renewal decision for each patent is made on one hand with respect to the renewal decisions for patents owned by its own firm, and on the other hand with respect to the renewal decisions for patents owned by other firms, see Table 4.6. For this I defined two spatial lags one taking into account only the renewal of patents owned by other firms, $W_{ext} * R$, and one for the renewal of patents owned by the same firm, $W_{int} * R$. Dependence is significant with respect to patents of other firms at every stage of renewal, and

Table 4.6: Interdependent Renewal Decisions Between Firms and Within a Firm

Dep. Variable	1st Renewal	2nd Renewal	3rd Renewal
W_ext*Renewal	0.09602*	0.12596***	0.14398***
	(0.04151)	(0.02886)	(0.02185)
W_int*Renewal	0.00019*	0.00038**	-0.00037*
	(0.00007)	(0.00013)	(0.00018)
Cites Received	0.00030***	0.00068***	0.00064***
	(0.00004)	(0.00006)	(0.00006)
Cites Made	0.00044***	-0.00028	0.00004
	(0.00012)	(0.00019)	(0.00022)
Claims	0.00087***	0.00125***	0.00122***
	(0.00010)	(0.00015)	(0.00018)
Entity Large	0.13478***	0.17748***	0.16014***
	(0.00875)	(0.01121)	(0.01323)
Unassigned	-0.03203*	-0.06624***	-0.09803***
	(0.01287)	(0.01730)	(0.01989)
US non-Gov	-0.01120	-0.04678***	-0.02893*
	(0.00922)	(0.01314)	(0.01412)
US individual	0.01956	-0.08429	-0.00928
	(0.03508)	(0.05313)	(0.05583)
Non-US individual	-0.07632	-0.06220	0.00061
	(0.06572)	(0.08547)	(0.09660)
US Gov	-0.36072***	-0.42219***	-0.44006***
	(0.03103)	(0.04118)	(0.05896)
Non-US Gov	-0.04489	-0.02572	0.12579*
	(0.05964)	(0.07392)	(0.05674)
R-squared	0.053	0.043	0.062
N	28077	25680	20989
	country FE	country FE	country FE
	cohort FE	cohort FE	cohort FE
	tech class FE	tech class FE	tech class FE

* p<0.05, ** p<0.01, *** p<0.001.
HAC standard errors are reported in parenthesis.

smaller than for the whole sample. Moreover, dependence in renewals within the firm is much smaller and of low significance level.

Some interpretations of the within firm dependence for patent renewal, in the context of strategic behavior to preempt entry, are: patent fencing or patent blocking. It is not surprising that within firm renewal dependence is weak, since when a patent expires it does not lose its value but can still protect the company against imitation. By law if another company copies the design of a product even after the patent expires, that company can be sued for trade dress infringement by the "holder" of the expired patent.

While, the interpretation for strategic interaction in patent renewal between firms is patent thickets. Patent thickets occur when each product may involve many patents, and when there is excessive patenting for that product to induce cross-citation with related products owned by other firms. Consequently, thickets can be used as threat of litigation in negotiation to extract rents from rival innovators. This type of patenting is abusive and imposes transaction costs, holdup or vertical monopoly problems, mentioned by Shapiro (2001).

Further, I merged the patent data with balance sheet data on R&D from the Compustat database. My sample size decreases since it includes only publicly traded US companies. My next set of results in Table 4.7 show much more interdependence in renewals for this sample. I joined the data sets based on the NBER file that matches assignees to Compustat corporate entities. These regressions include per patent research spending one year prior to patent grant. As a measure of research inputs, I used log annual R&D scaled by the resulting patent count granted in the four years following the R&D spending.[25] I am making the assumption that the returns to one year R&D spending come from rents to patents granted over the

[25] I chose this scaling knowing that the recipients NIST Technology Innovation Program and Advanced Technology Program Grants are funded for 3-5 years per project. See http://www.nist.gov/tip/faqs.cfm

Table 4.7: Returns to per patent R&D spending made one year prior to patent grant

Dep. Variable	1st Renewal	2nd Renewal	3rd Renewal
W*renewal	0.46367***	0.35154***	0.26416***
	(0.10033)	(0.09869)	(0.07972)
Cites Received	0.00028***	0.00076***	0.00077***
	(0.00005)	(0.00008)	(0.00009)
Cites Made	0.00021	-0.00031	-0.00039
	(0.00017)	(0.00034)	(0.00043)
Claims	0.00071***	0.00143***	0.00173***
	(0.00014)	(0.00023)	(0.00027)
Entity Large	0.02182	0.11648**	0.05222
	(0.02532)	(0.03960)	(0.04747)
US non-Gov	-0.04100*	-0.04856	-0.00111
	(0.01762)	(0.02546)	(0.02884)
log Scaled R&D	0.01097***	0.02785***	0.02114***
	(0.00238)	(0.00396)	(0.00446)
1st Stage F-stat	9.05	9.73	17.22
R-squared	0.019	0.043	0.077
N	14061	13111	10724
	country FE	country FE	country FE
	cohort FE	cohort FE	cohort FE
	tech class FE	tech class FE	tech class FE

* p<0.05, ** p<0.01, *** p<0.001.
HAC standard errors are reported in parenthesis.

following 4 years. Liu et al. (2008) used another measure of research intensity which is the log value of the R&D expenditure scaled by the total revenues. Notice that the significance of the entity size variable disappears in this regressions since only about 100 small entities were left in the sample. By employing the log of scaled R&D I get an estimate of 1-2.8% over different renewal stages. Implying that at the mean of 0.675 million dollars R&D per patent, an increase of 100% increases the probability of renewal by 2-4.3%, when taking into account the multiplier effect. This result confirms the study of Bessen (2008) who finds that the ratio of patent value to R&D – a measure of the subsidy that patents provide to R&D investment – is only about 3%. My results are robust when scaling R&D by the patent count granted over only two years following the spending, see Table B.1 in the Appendix B. For another robustness check in Table B.2 in the Appendix B, I used scaled R&D spending made two years before patent grant and I find returns to be smaller and less precisely estimated.

4.6 Conclusion

In this paper, I introduce a semi structural model of interdependent renewal decisions that can be estimated as a spatial linear probability model. I show that the returns to a patent are not only a function of patent and owner characteristics, but also of the renewal decisions for related patented innovations. Previous studies that ignored the interrelatedness of patent renewal decisions are likely to have biased results. My estimation methodology does not make assumptions on the error term distribution, while other studies which make parametric assumptions could use misspecified models, e.g. probit, logit, ordered probit. Ultimately, I suggest firm strategic behavior, consisting of patent fencing, patent blocking and patent thickets, to be a potential explanation for the positive dependence in renewals that I find. This type of strategic behavior can impose large costs to society when it brings about property rights lawsuits between rival companies. To reduce litigation costs policy makers should consider awarding less patents for products or processes that are not highly differentiated. Another case when interdependence in renewals would also arise is if the patent owner is uncertain about the value of its patent. Thus if related patents are being renewed, it may indicate to the patent holder that its patent is worth renewing as well which would result in interdependent renewals for similar patents. Moreover, when it comes to estimating the intangible value of a transfer of patents from one firm portfolio to another, the taxpayers and the IRS should also take into account additional returns coming from the interaction between the patents in the traded portfolio.

Chapter 5

Conclusion

The literature review from Chapter 1 covers different spatial discrete models. I present some consistency results for estimators of discrete choice models with spatially correlated unobserved components, e.g. Pinkse and Slade (1998) and Wang et al. (2009). The estimator of Pinkse and Slade (1998) adjusts for the spatially introduced heteroskedasticity, while the one of Wang et al. (2009) is a partial maximum likelihood estimator. For models with spatial dependence in the observed discrete choices there are various Bayesian estimators, e.g. Bolduc et al. (1997) and LeSage (2000), and Simulations estimators, e.g. Beron and Vijverberg (2004), whose small sample properties are investigated. There is still no formal theory for the consistency of these latter estimators where discrete choices are made interdependently. Chapter 3 proposes a model that can be consistently estimated using Kelejian and Prucha (1998, 2007) techniques from the spatial literature with a continuous dependent variable.

The Spatial Linear Probability Model (SLPM) discussed in Chapter 3 has a different specification from the models that already exist in the spatial discrete choice literature. Rather than having the interdependent discrete choices be a result of a truncation in some latent variable, the model is formulated based on the probability of making interdependent choices. I investigate the small sample properties of a two step estimator where the binary outcomes are generated independently and dependently. The results show a considerable bias for the first step estimator. However, the bias decreases significantly for the second step estimator. Moreover both the bias and root mean squared errors decrease with an increase in the sample size offering evidence for the consistency of the procedure.

Chapter 4 presents a game among patent holders who make the decision to renew technologically similar patents interdependently and simultaneously. It turns out that this model for the probability of renewing a patent can be cast as the SLPM in Chapter 3. In terms of the estimation strategy for the patent renewal model, I applied the first step estimator from the previous chapter. The simulation results in Chapter 3 show good small sample properties for the first step estimator even when the Monte Carlo sample size is less than one tenth of the sample size used in the application of Chapter 4. I find that there is positive interdependence in the renewal decision for patents and that ignoring it would lead to biased coefficient and underestimated marginal effects. The positive interdependence can be explained by positive network effects, as well as strategic behavior behind patent renewal.

The theoretical model of interdependent choices that I developed and the estimation strategy are not confined to my application, but rather can have other applications in which firms or people are making strategic and simultaneous decisions.

Appendix A

Data Description

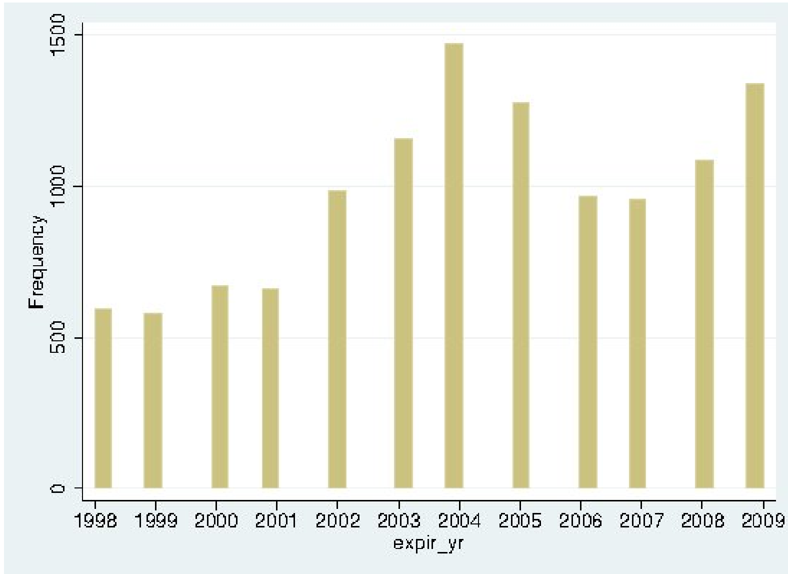

Figure A.1: Expiration Years

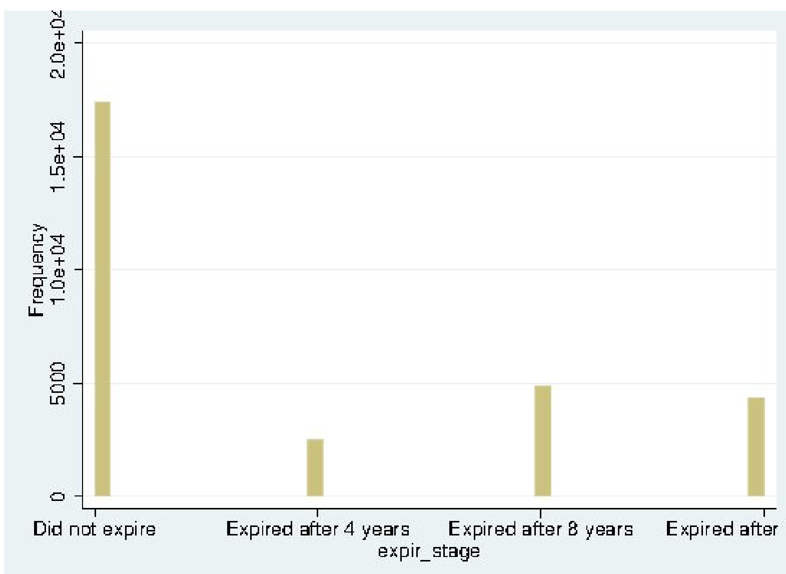

Figure A.2: Expiration Stages

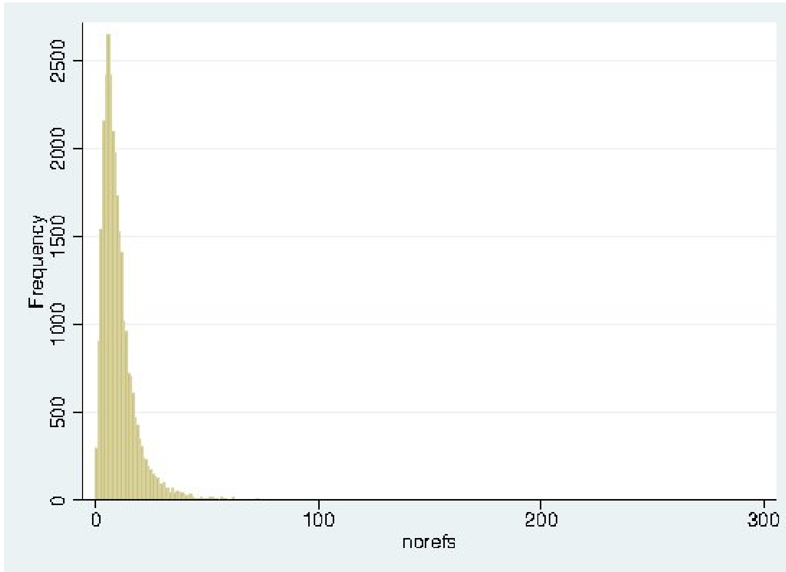

Figure A.3: Frequency of Cited References

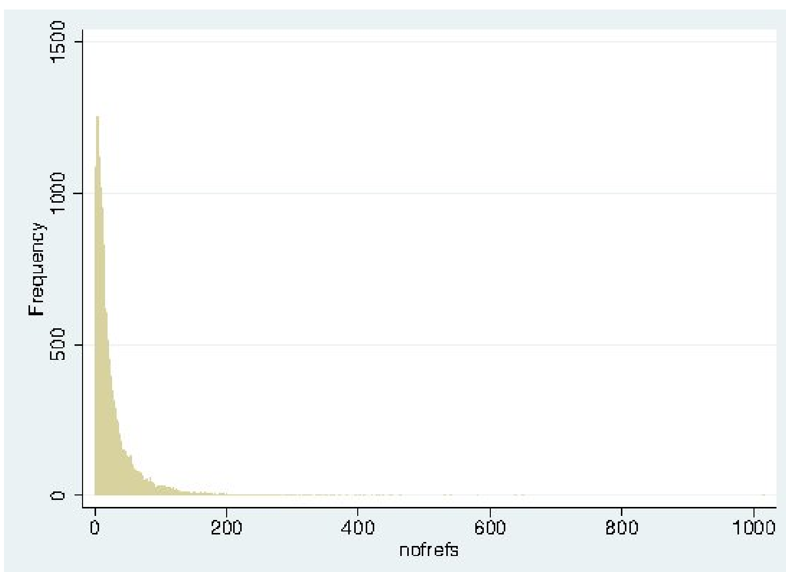

Figure A.4: Frequency of Citations Received

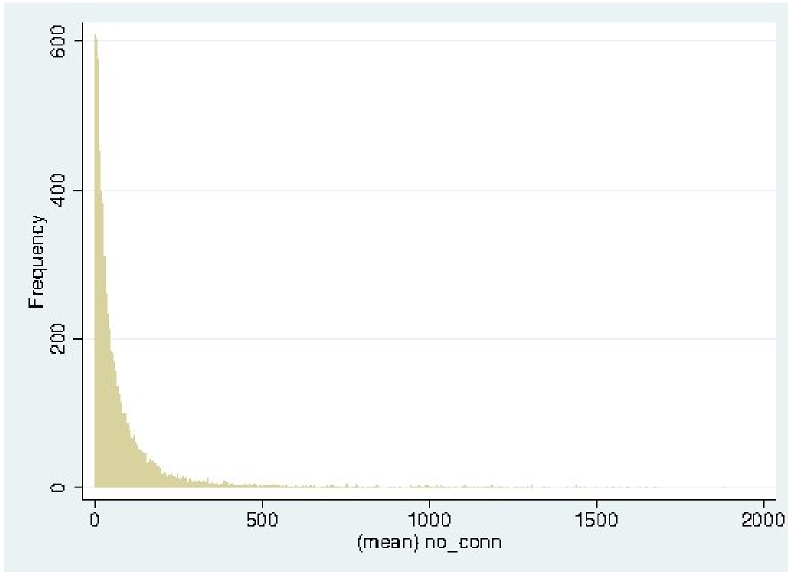

Figure A.5: Frequency of Number of Patents to which One is Similar to

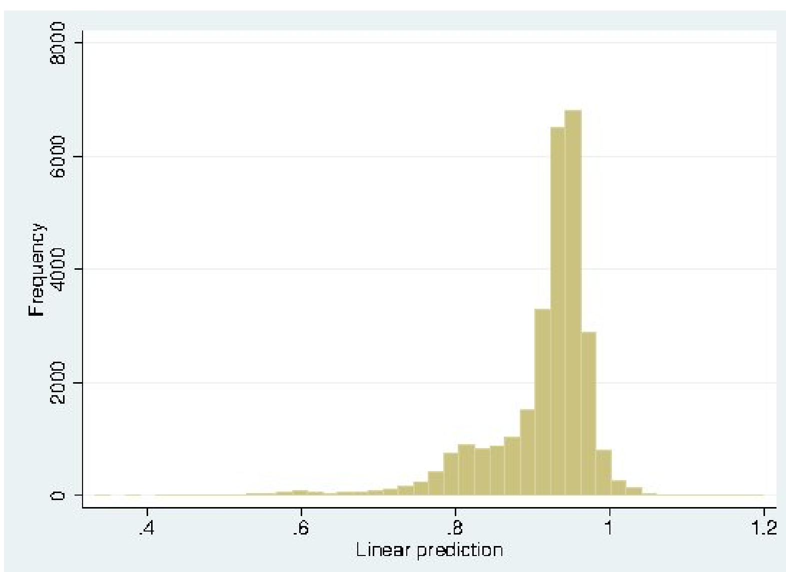

Figure A.6: Frequency of Predicted Probabilities of First Renewal

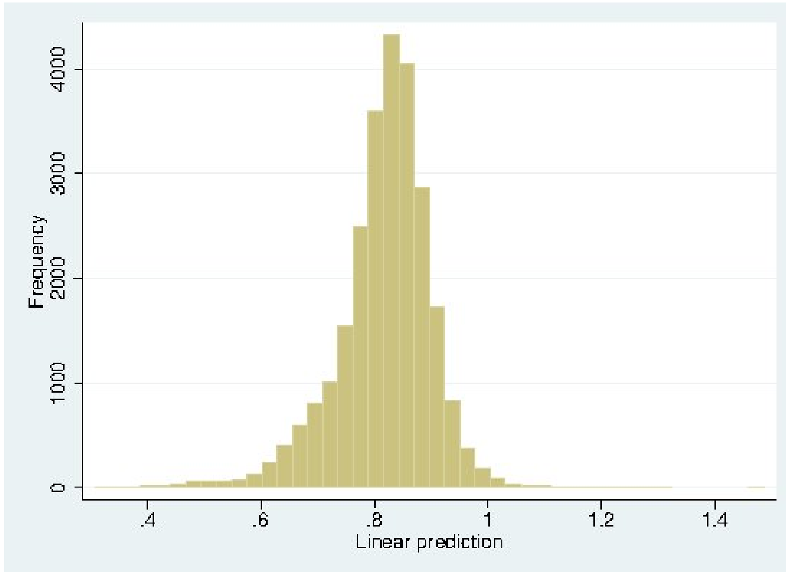

Figure A.7: Frequency of Predicted Probabilities of Second Renewal

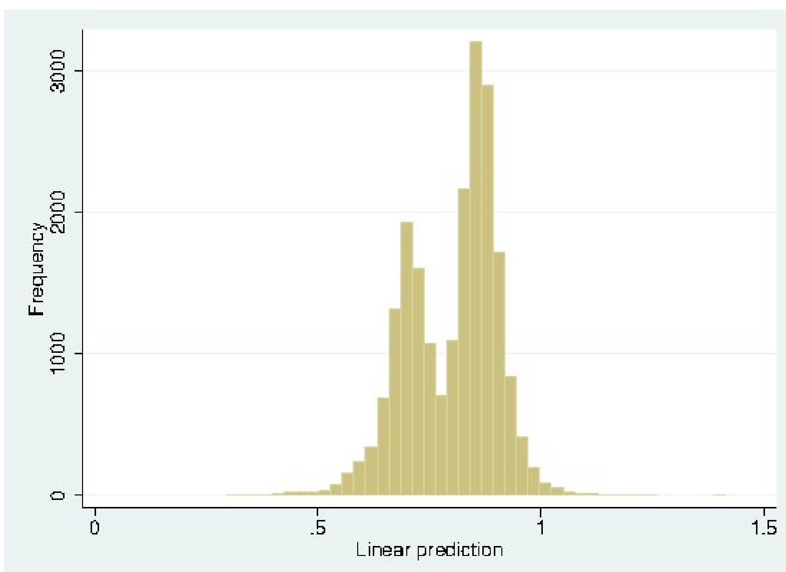

Figure A.8: Frequency of Predicted Probabilities of Third Renewal

A.1 Appendix A1

TECHNOLOGY CLASSES FOR CALCULATORS, COMPUTERS, OR DATA PROCESSING SYSTEMS

341. Coded Data Generation or Conversion

380. Cryptography

382. Image Analysis

395. Information Processing System Organization

700. Data Processing: Generic Control Systems or Specific Applications

701. Data Processing: Vehicles, Navigation, and Relative Location

702. Data Processing: Measuring, Calibrating, or Testing

704. Data Processing: Speech Signal Processing, Linguistics, Language Translation, and Audio Compression/Decompression

705. Data Processing: Financial, Business Practice, Management, or Cost/Price Determination

706. Data Processing: Artificial Intelligence

707. Data Processing: Database and File Management or Data Structures

708*. Electrical Computers: Arithmetic Processing and Calculating

709. Electrical Computers and Digital Processing Systems: Multicomputer Data Transferring

710. Electrical Computers and Digital Data Processing Systems: Input/Output

711. Electrical Computers and Digital Processing Systems: Memory

712. Electrical Computers and Digital Processing Systems: Processing Archi-

tectures and Instruction Processing (e.g., Processors)

713. Electrical Computers and Digital Processing Systems: Support

714. Error Detection/Correction and Fault Detection/Recovery

718. Electrical Computers and Digital Processing Systems: Virtual Machine Task or Process Management or Task Management/Control

719. Electrical Computers and Digital Processing Systems: Interprogram Communication or Interprocess Communication (IPC)

A.2 Appendix A2

POST-ISSUANCE FEES IN 2010

Maintenance fees

For maintaining an original or reissue patent, except a design or plant patent, based on an application filed on or after after Dec. 12, 1980, in force beyond 4 years; the fee is due by three years and six months after the original grant:

By a small entity $490.00

By other than a small entity $980.00

For maintaining an original or reissue patent, except a design or plant patent, based on an application filed on or after Dec. 12, 1980 in force beyond 8 years; the fee is due by seven years and six months after the original grant:

By a small entity $1,240.00

By other than a small entity $2,480.00

For maintaining an original or reissue patent, except a design or plant patent, based on an application filed on or after Dec. 12, 1980 in force beyond 12 years; the fee is due by eleven years and six months after the original grant:

By a small entity $2,055.00

By other than a small entity $4,110.00

Late Payment Penalties

Surcharge for paying a maintenance fee during the 6 month grace period following the expiration of three years and six months, seven years and six months, and eleven years and six months after the date of the original grant of a patent based on an application filed on or after Dec. 12, 1980:

By a small entity $65.00

By other than a small entity $130.00

(i) Surcharge for accepting a maintenance fee after expiration of a patent for non-timely payment of a maintenance fee where the delay is shown to the satisfaction of the Commissioner to have been:

(1) unavoidable $700.00

(2) unintentional $1,640.00

Source: http://www.uspto.gov/web/offices/com/sol/og/1997/week50/patmfee.htm

Appendix B

Derivations and Tables

In light of (4.9) the probability of renewal of patent j is given by:

$$\Pr(r_j = 1) = \Pr\left(\xi_j^0 - \xi_j^1 + \epsilon_j^0 - \epsilon_j^1 \leq a_j + b\sum_{j\neq i} w_{jk}E(r_k = 1|F_j) - c_j\right). \tag{B.1}$$

Recall that $F_j = \sigma(\xi_j^0, \xi_j^1, \epsilon_j^0, \epsilon_j^1)$, and thus $E(r_k = 1|F_j) = g_k(\xi_j^0, \xi_j^1, \epsilon_j^0, \epsilon_j^1)$ for some measurable function $g_k(.)$. The probability of renewal of patent j, conditional on the information set F_i, is thus given by:

$$\begin{aligned}\Pr(r_j &= 1|F_i) = \Pr\left(\left[\xi_j^0 - \xi_j^1 + \epsilon_j^0 - \epsilon_j^1 \leq a_j + b\sum_{j\neq i} w_{jk}E(r_k = 1|F_j) - c_j\right]|F_i\right)\\ &= \Pr\left(\left[\xi_j^0 - \xi_j^1 + \epsilon_j^0 - \epsilon_j^1 - b\sum_{j\neq i} w_{jk}g_k(\xi_j^0, \xi_j^1, \epsilon_j^0, \epsilon_j^1) \leq a_j - c_j\right]|F_i\right).\end{aligned}$$

Since $\xi_j^0, \xi_j^1, \epsilon_j^0, \epsilon_j^1$ and $\xi_i^0, \xi_i^1, \epsilon_i^0, \epsilon_i^1$ are assumed to be independent it follows further that

$$\begin{aligned}\Pr(r_j &= 1|F_i) = \Pr\left\{\left(\xi_j^0 - \xi_j^1 + \epsilon_j^0 - \epsilon_j^1 - b\sum_{j\neq i} w_{jk}g_k(\xi_j^0, \xi_j^1, \epsilon_j^0, \epsilon_j^1) \leq a_j - c_j\right)\right\}\\ &= \Pr(\left(\xi_j^0 - \xi_j^1 + \epsilon_j^0 - \epsilon_j^1 \leq a_j + b\sum_{j\neq i} w_{jk}E(r_k = 1|F_j) - c_j\right) = \Pr(r_j = 1),\end{aligned}$$

i.e., $\Pr(r_j = 1|F_i) = \Pr(r_j = 1)$as claimed.

Table B.1: Returns to log R&D spending made a year prior to patent grant scaled by the number of patents awarded in the following 2 years

Dep. Variable	SLPM 1st Renewal	SLPM 2nd Renewal	SLPM 3rd Renewal
W*renewal	0.43862***	0.33658***	0.25852**
	(0.10227)	(0.09765)	(0.07927)
Cites received	0.00028***	0.00077***	0.00077***
	(0.00005)	(0.00008)	(0.00009)
Cites made	0.00021	-0.00027	-0.00042
	(0.00017)	(0.00031)	(0.00039)
Claims	0.00071***	0.00141***	0.00172***
	(0.00013)	(0.00022)	(0.00026)
Entity large	0.02706	0.13102**	0.05146
	(0.02682)	(0.04046)	(0.04474)
US non-Gov	-0.03955*	-0.05087*	-0.00066
	(0.01715)	(0.02542)	(0.02803)
Log scaled R&D	0.01413***	0.02735***	0.02158***
	(0.00236)	(0.00391)	(0.00440)
R-squared	0.024	0.043	0.076
N	14062	13116	10717
p<0.05, ** p<0.01, *** p<0.001			
	country FE	country FE	country FE
	cohort FE	cohort FE	cohort FE
	tech class FE	tech class FE	tech class FE

Table B.2: Returns to log R&D spending made two years prior to patent grant scaled by the number of patents awarded in the following 4 years

Dep. Variable	SLPM 1st Renewal	SLPM 2nd Renewal	SLPM 3rd Renewal
W*Renewal	0.52473***	0.42650***	0.25232**
	(0.10835)	(0.09912)	(0.08044)
Cites received	0.00028***	0.00076***	0.00075***
	(0.00006)	(0.00008)	(0.00009)
Cites made	0.00023	-0.00028	-0.00054
	(0.00017)	(0.00032)	(0.00039)
Claims	0.00074***	0.00151***	0.00156***
	(0.00013)	(0.00023)	(0.00026)
Entity large	0.02204	0.12266**	0.07741
	(0.02742)	(0.04103)	(0.04711)
US non-Gov	-0.04656**	-0.05996*	-0.02120
	(0.01718)	(0.02559)	(0.02575)
Log scaled R&D	0.00532*	0.01953***	0.00697*
	(0.00226)	(0.00362)	(0.00294)
R-squared	0.016	0.039	0.073
N	13705	12779	10991
p<0.05, ** p<0.01, *** p<0.001			
	country FE	country FE	country FE
	cohort FE	cohort FE	cohort FE
	tech class FE	tech class FE	tech class FE

Bibliography

Aghion, Philippe and Peter Howitt, "A model of growth through creative destruction," *Econometrica*, 1992, *60*, 323–51.

Aitkin, M. and I. Aitkin, "A hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions," *Statistics and Computing*, 1996, *6*, 127–130.

Avery, R.B., L.P. Hansen, and V.J. Hotz, "Multiperiod Probit Models and Orthogonality Condition Estimation," *International Economic Review*, 1983, *24*, 21–35.

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov, "Computing Equilibria in Static Games of Incomplete Information Using the All-Solution Homotopy," *Working Paper*, 2010.

Baudry, Marc and Beatrice Dumont, "Patent renewal as options: Improving the mechanism for weeding out lousy patents," *Review of Industrial Organization*, 2006, *28*, 41–61.

_ and Batrice Dumont, "A Bayesian Real Option Approach to Patents and Optimal Renewal Fees," Technical Report, HAL 2009.

Belenzon, Sharon, "Knowledge Flow and Sequential Innovation: Implications for Technology Diffusion, R&D and Market Value," CEP Discussion Papers, Centre for Economic Performance, LSE May 2006.

Beron, K. J. and W.P.M. Vijverberg, "Probit in a Spatial Context: A Monte Carlo Analysis.," *In Anselin, L., Florax, R.J., Rey, S.J. (Eds.), Advances in spatial econometrics: methodology, tools and applications. Business & Economics*, 2004, pp. 169–195.

Beron, K.J., J.C. Murdoch, and W.P.M. Vijverberg, "Why cooperate? Public goods, economic power, and the Montreal Protocol," *Unpublished*, 1997.

Bessen, James, "The value of U.S. patents by owner and patent characteristics," *Research Policy*, 2008, *37*, 932–945.

_ , "Estimates of patent rents from firm market value," *Research Policy*, 2009, *38*, 1604–1616.

_ and Michael Meurer, "The private costs of patent litigation," *Boston University School of Law Working Paper*, 2008.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying techonology spillovers and product market rivalry," *NBER WP 13060*, 2007.

Blume, L. and S. Durlauf, "Equilibrium Concepts for Social Interaction Model," *WP*, 2002.

Bolduc, D., "A practical technique to estimate multinominal probit models in transportation," *Transportation Research Part B: Methodological*, 1999, *33*, 63–67.

_ , B. Fortin, and S. Gordon, "Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques," *International Regional Science Review*, 1997, *20*, 77–101.

Bramoulle, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, 2009, *150*, 41–55.

Brock, W. and S. Durlauf, "Interactions-based models," *In Heckman, J., and Leamer, E. (Eds.) Handbook of Econometrics*, 2001b, *5*, 3297–3380.

Brock, W.A. and S.N. Durlauf, "Discrete Choice with Social Interactions," *The Review of Economic Studies*, 2001a, *68*, 235–260.

_ and _ , "A Multinomial-Choice Model of Neighborhood Effects," *American Economic Review*, 2002, *92*, 298–303.

_ and _ , "Indentification of binary choice models with social interactions," *Journal of Econometrics*, 2007, *140*, 52–75.

Bulow, Jeremy, John Geanakoplos, and Paul Klemperer, "Multimarket oligopoly: strategic substitutes and strategic complements," *Journal of Political Economy*, 1985, *93*, 488–511.

Caballero, Ricardo J. and Adam B. Jaffe, ""How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth,"," *NBER Macroeconomics Annual 1993*, 1993, *8*, 15–86.

Case, A., "Neighborhood Influence and Technological Change," *Regional Science and Urban Economics*, 1992, *22*, 491–508.

Cliff, A. and J. Ord, *Spatial Autocorrelation*, London: Pion, 1973.

_ and _ , *Spatial Processes*, London: Pion, 1981.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh, "Protecting their intellectual assests: appropriability conditions and why US manufacturing firms patent (or not)," *NBER WP 7552*, 2000.

Cornelli, Francesca and Mark Schankerman, "Patent Renewals and R&D incetives," *RAND Journal of Economics*, 1999, *30*, 197–213.

Dempster, A., N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, 1977, *39*, 1–38.

Dubin, R.A., "Spatial autocorrelation and neighborhood quality," *Regional Science and Urban Economic*, 1992, *22*, 433–452.

Egger, P. and M. Larch, “Interdependent preferential trade agreement memberships: An empirical analysis,” *Journal of International Economics*, 2008, *76*, 384–399.

Fleming, M.M., “Techniques for Estimating Spatially Dependent Discrete Choice Models,” *In Anselin, L., Florax, R.J., Rey, S.J. (Eds.), Advances in spatial econometrics: methodology, tools and applications. Business & Economics*, 2004, pp. 145–168.

Glaeser, Edward, Bruce Sacerdote, and Jose Scheinkman, “The Social Multiplier,” *Harvard Institute Research Working Paper No. 1968.*, 2002.

Gourieroux, C. and A. Monfort, “Simulation-based inference: A survey with special reference to panel data models,” *Journal of Econometrics*, 1993, *59*, 5–33.

Gourieux, C., A. Monfort, and A. Trognon, “Estimation and Test in Probit Models with Serial Correlation,” *In Florens, J. P., Mouchart, M., Raoult, J. P. and Simar L. (eds.) Alternative Approaches to Time Series Analysis (Brussels: Publications des Facultes Universitaires Saint-Louis)*, 1984.

Graham, Bryan and Jinyong Hahn, “Identification and estmation of the linear-in-means model of social interactions,” *Economics Letters*, 2005, *88*, 1–6.

Griliches, Zvi, “The Search for R&D Spillovers,” *Scandinavian Journal of Economics*, 1992, *94, suplement*, S29–S47.

_ , “R&D and productivity: The unfinished business,” in “R&D and Productivity: The Econometric Evidence,” The University of Chicago Press, Chicago and London, 1998.

Grindley, Peter and David Teece, “Managing Intellectual Capital: Licensing and Cross-Licensing in Semiconductors and Electronics,” *California Management Review*, 1997, *29*, 8–41.

Hajivassiliou, V. and P. Ruud, “Classical estimation methods for LDV models using simulation,” *In R. Engle and D. McFadden (eds.) Handbook of Econometrics, North-Holland, Amsterdam*, 1994, p. 23832441.

Hall, Bronwyn and Rosemary Ham Ziedonis, “The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995,” *RAND Journal of Economics*, 2001, *32*, 101–128.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Tranjenberg, “The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools,” *NBER WP 8498*, 2001.

_ , Adam Jaffe, and Manuel Tranjenberg, “Market value and patent citations,” *RAND Journal of Economics*, 2005, *36*, 16–38.

Harhoff, Dietmar, Frederic Scherer, and Katrin Vopel, "Citations, family size, opposition and the value of patent rights," *Research Policy*, 2003, *32*, 1343–1363.

Hautsch, N. and S Klotz, "Estimating the neighborhood influence of decision makers: theory and an application on the analysis of innovation decisions," *Journal of Economic Behaviour and Organization*, 2003, *52*, 97–113.

Heckman, James, "Heterogenity and State Dependence," in S. Rosen, ed., *Studies in Labor Market*, Chicago University Press, Chicago, 1981, p. 91139.

Heckman, James J. and James M. Snyder, "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preference of Legislators," *Rand Journal of Economics*, 1997, *28*, 14289.

Hopenhayn, Hugo, Gerard Llobet, and Matthew Mitchell, "Rewarding Sequantial Innovators: Prizes, Patents, and Buyouts," *Journal of Political Economy*, 2006, *114*, 1041–1068.

Iglesias, E.M. and G.D. Phillips, "Asymptotic bias of GMM and GEL under possible nonstationarity spatial dependence," *Economics Letters*, 2008, pp. 393–397.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty, "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *The American Economic Review*, 2000, *90:2*, 215–218.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 1993, *108:3*, 577598.

Kelejian, Harry and Ingmar Prucha, "A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 1998, *17*, 99–121.

_ and _ , "HAC estimation in a spatial framework," *Journal of Econometrics*, 2007, pp. 131–154.

_ and _ , "Specification and Estimation of Spatial Autoregressive Models with Autoregressive Disturbances," *Journal of Econometrics*, 2010, *157*, 53–67.

_ and _ , "Spatial Linear Probability Model," *Working Paper*, 2011.

Keller, Wolfgang, "International Technology Diffusion," *Journal of Economic Literature*, 2004, *42(3)*, 752–782.

Kim, Min Seong and Yixiao Sun, "Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix," *Journal of Econometrics*, 2010, *Forthcoming.*

Kremer, Michael, "Patent Buyouts: A mechanism for Encouraging Innovation," *QJE*, 1998, *113*, 1137–67.

Krummenauer, Frank, "Efficient Simulation of Multivariate Binomial and Poisson Distributions," *Biometrica*, 1998, *40*, 823–832.

Lahiri, S.L., "On Inconsistency of Estmators Based on Spatial Data under Infill Asymptotics," *The Indian Journal of Statistics*, 1996, *58*, 403–417.

Langinier, Corinne, "Are patents strategic barriers to entry?," *Journal of Economics and Business*, 2004, *56*, 349–361.

Lanjouw, Jean and Marc Schankerman, "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators," *The Economic Journal*, 2004, *114*, 441–465.

Lanjouw, Jean O., Ariel Pakes, and Jonathan Putnam, "How to Count Patents and Value Intellectual Property: The uses of patent renewal and application data," *Journal of Industrial Economics*, 1998, *46*, 405–32.

Lee, Lungfei, "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models," *Econometric Theory*, 1992, *8*, 518–552.

_ , "Asymptotic bias in simulated maximum likelihood estimation of discrete choice models," *Econometric Theory*, 1995, *11*, 437–438.

_ , "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 2004, *72(6)*, 18991925.

_ , "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 2007, *140*, 333–374.

LeSage, James, "Bayesian Estimation of Limited Dependent variable Spatial Autoregressive Models," *Geographical Analysis*, 2000, *32*, 19–35.

Levine, R. and G. Casella, "Implementation of the Monte Carlo EM algorithm," *Journal of Computational and Graphical Statistics*, 2001, *10*, 422439.

Li, J. and L.F. Lee, "Binary Choice under Social Interaction: An Empirical Study with and Without Subjective Data of Expectation," *WP*, 2006.

Liu, Kun, Jonathan Arthurs, John Cullen, and Roger Alexander, "Internal sequantial innovations: How does interrelatedness affect patent renewal?," *Research Policy*, 2008, *37*, 946–953.

Manski, Charles F, "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 1993, *60* (3), 531–42.

McFadden, D., “A method of simulated moments for estimation of discrete response models without numerical integration,” *Econometrica*, 1989, *57*, 995–1026.

McMillen, D.P., “Probit with Spatial Autocorrelation,” *Journal of Regional Science*, 1992, *32* (375), 335–348.

_ , “Spatial Effects in Probit Models: A Monte Carlo Investigation,” *In Anselin, L. and Florax, R.J. (Eds.), New Directions in Spatial Econometrics. Springer, New York*, 1995, pp. 189–228.

Moffitt, Robert, *Social Dynamics*, MIT Press, Cambridge, MA, 2001.

Moore, Kimberly, “Worthless patents,” *Berkley Technology Law Journal*, 2005, *20.*

Pakes, A. and D. Pollard, “Simulation and asymptotics of optimization estimators,” *Econometrica*, 1989, *57*, 1027–1057.

Pakes, Ariel, “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks,” *Econometrica*, 1986, *54*, 755–784.

_ and Mark Schankerman, “The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources,” in “R & D, Patents, and Productivity,” National Bureau of Economic Research, Inc, 1984.

Pinkse, J. and M. Slade, “Contracting in space: An application of spatial statistics to discrete-choice models,” *Journal of Econometrics*, 1998, *85*, 125–154.

_ and _ , “Semi-Structural Models of Advertising Competition,” *Journal of Applied Econometrics*, 2007, *22*, 1227–1246.

_ , _ , and L. Shen, “Dynamic Spatial Discrete Choice Using One Step GMM: An Application to Mine Operating Decisions,” *Spatial Economic Analysis*, 2006, *1.*

Poirier, D.J. and P.A. Ruud, “Probit with dependent observations,” *Review of Economic Studies*, 1988, *55*, 593–614.

Reinganum, Janet, *The timing of Innovation: Research, Development and Diffusion*, Handbook of Industrial Organization, 1989.

Robinson, P.M., “On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables,” *Econometrics*, 1982, *50*, 27–41.

Romer, Peter, “Endogenous Technological Change,” *Journal of Political Economy*, 1990, *98*, S71–S102.

Ruud, P., “Extensions of estimation methods using the EM algorithm,” *Journal of Econometrics*, 1991, *49*, 305–341.

Schankerman, Mark and Ariel Pakes, “Estimates of the value of patent rights in European countries during the post-1950 period,” *Economic Journal*, 1986, *96*, 1052–76.

Scotchmer, Suzanne, "On the optimality of the patent renewal system," *RAND Journal of Economics*, 1999, *30*, 181–196.

Serrano, Carlos, "The Market for Intellectual Property: Evidence From the Transfer of Patents.," *Working Paper*, 2006.

Shapiro, Carl, "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting," in Adam Jaffe, Joshua Lerner, and Scott Stern, eds., *Innovation Policy and the Economy*, NBER, 2001.

Smirnov, O., "Spatial Discrete Choice Model," *WP*, 2008.

Soetevent, A.R. and P. Kooreman, "A Discrete Choice Model with Social Interactions: With An Application to High School Teen Behavior," *Journal of Applied Econometrics*, 2007, *22*, 599–624.

Spence, Michael, "Cost Reduction, Competition and Industry Performance," *Econometrica*, 1984, *52*, 101–121.

Staiger, Douglas and James Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 1997, *65*, 557–86.

Su, Che-Lin and Kenneth Judd, "Constrained Optimization Approaches to Estimation of Structural Models," *Work*, 2010.

Tranjenberg, Manuel, Adam Jaffe, and Rebecca Henderson, "University versus Corporate Patents: A Window on the Basicness of Invention," *Econonomics of Innovation and New Technology*, 1997, *5*, 19–50.

Vijverberg, W.P.M., "Monte Carl evaluation of multivariate normal probabilities," *Journal of Econometrics*, 1997, *76*, 281–307.

Wang, Honglin, Emma M. Iglesias, and Jefferey M. Wooldridge, "Partial Maximum Likelihood Estimation of Spatial Probit Models," *WP*, 2009.

Weeks, D. and K. Lange, "Trials, tribulations, and triumphs of the EM algorithm in pegigree analysis," *Journal of Mathematics Applied in Medicine and Biology*, 1989, *6*, 209–232.

Yatchew, A. and Z. Grilitches, "Specification Error in Probit Models," *The Review of Economics and Statistics*, 1985, *18*, 134–139.