December 2014

Three Essays in Applied Econometrics: with Application to Natural Resource and Energy Markets

Chong Li
Syracuse University

Abstract

Chapter 1

This essay examines the effect of state renewable energy policies in inducing innovation and the spillover effect of these policies on innovation in neighboring states. The analysis is conducted with patent data related to renewable technology using wind power for the United States over the period 1983-2010. We run a panel data regression of a log transformation of states' yearly patent counts on state renewable energy policies and spatially weighted average of renewable energy policies in neighboring states using the Tobit model with individual effects. The results show that renewable energy rules, regulation and mandates such as interconnection standards, net metering and renewable portfolio standard enacted in neighboring states have shown a statistically significant positive spillover effect in increasing the number of patent applications in that state. However, financial policies such as tax incentives and subsidy policies implemented by neighboring states have shown statistically significant negative effects on technological innovation within that state.

Chapter 2

In this essay, we have conducted a Monte Carlo Study of the prediction performance of various nonparametric estimation methods for spatially dependent data, such as the nonparametric local linear kernel estimator, the Nadraya-Watson estimator, and the k-Nearest Neighbors method developed by Hallin et al. (2004b), Lu and Chen (2002), P.M. Robinson (2011) and Li and Tran (2009). With data sampled on a rectangular grid in a nonlinear random field, the results show that nonparametric local linear kernel method has the best performance in terms of mean squared prediction error. The Nadaraya-Watson estimation method also performs well. In general, these two nonparametric methods consistently outperform the k-Nearest Neighbors method and the

maximum likelihood method regardless of the data generating process and sample size. However, the maximum likelihood method does not perform well because the spatial weight matrix can only be used to estimate linear structures while the true data generating process is nonlinear. This also gives some support to the idea of using nonparametric methods when various misspecification may exist either in the functional form or spatial weight matrix for spatially dependent data.

We use these methods to predict county-level crop yields with spatially weighted precipitation. The results are generally consistent with the simulation results. The nonparametric local linear kernel estimator has the best prediction performance. The Nadaraya-Watson estimator also performs better than the k-Nearest Neighbors method and the maximum likelihood estimator. However, with an inverse distance weighting matrix, the maximum likelihood estimator outperforms the k-Nearest Neighbors method in predicting crop yield.

Chapter 3

This essay uses the "exceedances over high threshold model" of Davidson and Smith (1990) to investigate the univariate tail distribution of the returns on various energy products such as Crude Oil, Gasoline, Heating Oil, Propane and Diesel. The bivariate threshold exceedance model of Ledford and Tawn (1996) is also used to study the tail dependence between returns on various pairs of selected energy products. Tail index estimates for univariate threshold exceedance models show that these returns generally have fat tails similar to those of a Student's t-Distribution with 2 to 5 degrees of freedom except that for Crude Oil where the tail index estimates are closer to that of a normal distribution. We also estimate the tail dependence index for four pairs of energy products, crude oil/gasoline, crude oil/heating oil, crude oil/propane, crude oil/diese. The correlation coefficients implied by the dependence index estimates show that correlations conditional on threshold exceedance are generally higher than the unconditional correlation between crude oil/heating oil and crude oil/gasoline. However, there is some variation in the implied correlation between crude oil/propane and crude oil/diesel. Whether the extreme correlation will be higher or lower than the unconditional correlation depends on the threshold chosen.

Three Essays in Applied Econometrics: with Application to Natural Resource and Energy Markets

by

Chong Li

B.S., Huazhong University of Science and Technology, 2006
M.A., Xiamen University, 2009

DISSERTATION

Submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy in Economics

in the Graduate School of Syracuse University

December 2014

Acknowledgements

My deepest gratitude goes to all those who helped me complete this dissertation and contributed to my intellectual progress at Syracuse University.

In particular, I am extremely indebted to my advisor, Prof. William Horrace, who beefed me up when I was "wallowing in the valley of despair". Without his support, this dissertation would be a mission impossible. I also would like to thank my other committee members. Prof. Alfonso Flores-Lagunes, Prof. Chihwa Kao, Prof. David Popp and Prof. Peter Wilcoxen have spent their precious time on reading this dissertation. I have benefited a lot from their comments and suggestions. I am especially grateful to Prof. Lihong Liang for chairing my oral defense.

I would like to extend my thanks to all those professors and directors in this community for their dedicated teaching and helpful suggestions, which is essential for me to accumulate knowledge and develop skills for later research. I have also benefited a lot from those constructive conversations with my fellow classmates. In addition, I really appreciate the assistance of Sue Lewis, Laura Sauta and Faye Shephard in the Economics Department at Syracuse University.

Last but not least, I would like to sincerely thank my family and friends in China. Especially, I want to thank my father, my mother and my younger sister. Without their love and support, this dissertation would be impossible. It is my grandfather this dissertation is dedicated to.

Table of Contents

List of Tables

Chapter 1: Technological Spillover Effects of State Renewable Energy Policy: Evidence from Patent Count

Chapter 2: Nonparametric Prediction for Spatially Dependent Data

Chapter 3: Modeling Extreme Price Movements in the Energy Markets

List of Figures

Chapter 1: Technological Spillover Effects of State Renewable Energy Policy: Evidence from Patent Count

Chapter 2: Nonparametric Prediction for Spatially Dependent Data

Chapter 3: Modeling Extreme Price Movements in the Energy Markets

Chapter 1: Technological Spillover Effects of State Renewable Energy Policy: Evidence from Patent Count

> In a climate change speech at Georgetown University, President Obama argued that: "confronting climate change need not threaten economic growth: that investing in windmills, solar panels and other types of clean-energy technology could spur scientific innovation and generate jobs" (New York Times, July 2, 2013).

1.1 Introduction

Over the past three decades, green energy has emerged as an important topic of our social and economic life. It is now more widely accepted that the adoption of renewable energy sources such as wind, solar, geothermal, ocean, biomass, and waste-to energy can significantly contribute to environmental protection. Also, the diversification resulted from increased shares of renewable energy sources could also lead to greater energy security in the face of uncertainty in the fossil fuel markets. During this period, various environmental policies, both at the federal level and the state level, have been implemented to encourage the development of renewable energy. It is hoped that these policies will not only accelerate the pace of switching from conventional fossil fuels to renewable energy sources, but also cultivate innovation in environmentally friendly renewable technologies that will speed-up 'green growth'. In this paper, we will examine whether renewable energy policies enacted by a state will induce innovation within the state. In addition, we will investigate whether these policies have any spillover effects in facilitating innovation in neighboring states. The relative impact of state renewable energy policies on innovation within state and on innovation of neighboring states has important implications, since the existence of spillover effects in inducing innovation may change the relative competitive advantages that could be obtained by within state firms when competing with firms in neighboring states. If the spillover effects do exist, state policy makers may have to think about implementing renewable energy policies in coordination with policy makers from neighboring states rather than acting alone in order to achieve better overall effects.

Although the adoption of renewable energy sources has been increasing very rapidly and renewable energy sources have become a significant part of energy production for some states in the US; the use of renewable energy sources remain limited and public acceptance still needs to improve in most states. In the absence of government intervention favoring their development, renewable energy production costs are still relatively higher than conventional fossil fuels. For example, over the period from 2001 to 2010, the average proportion of electricity produced from wind power was almost zero in 12 states in the United States; and another 12 states with an average percentage somewhere between 0 and 0.56%. The 13 states in the lead all have an average percentage higher than 5.866%, and the highest is 23.587%. A geographic map showing the percentage of electricity produced from wind power in each state is given in Figure 1.1. In order to increase the share of renewable sources in the total energy supply, most of the states have introduced some form of renewable energy policy in an effort to increase renewable technology deployment, such as tax credits, subsidies, tradable renewable energy certificates, renewable energy portfolio standard, interconnection standard and net-metering. By either decreasing the relative price of renewable energy relative to fossil fuels, or increasing the demand for electricity generated from renewable sources, these policy measures will improve the relative cost or benefit of electricity generation using renewable energy sources compared to traditional fossil fuels. Additionally, it is well documented that renewable energy policies at the state level have increased market deployment of renewable technology, for example, see Carley (2009), Yin and Powers (2010), and Sarzynski et.al. (2012).

The adoption of renewable energy sources could contribute to environment protection and energy security. However, we still have to make sure that it is done in a way that is beneficial to economic growth. If the increased deployment of renewable technology will translate into innovation in renewable technology as reflected by patent filings, entrepreneurs would be more willing to devote effort to innovative activities in the beginning. This induced innovation would in return improve productivity in the energy industry and lead to 'green growth', which would make the overall benefit more likely to exceed the cost of implementing these renewable energy policies. It is hoped that these renewable energy policies will spur technological innovation related to renewable energy. The effectiveness of these different types of policy measures and the relative effectiveness of policies from within and outside of state in boosting innovation still need to be tested empirically. In this paper, we will examine whether state renewable energy policies will lead to technological innovation and investigate whether renewable energy policies enacted in neighboring states will have a spillover effect on technology innovation in that state.

To study the effect of renewable energy policy on innovation, we first need to find a measure for technological innovation. Many studies have used research and development (R&D) expenditures and number of scientific personnel as indicators of innovation in an economy. However, they are imperfect in the sense that these measures only focus on the input of innovative activity and it is not clear that the input will be transformed into innovation efficiently. In this case, people have turned to patent information as an output measure reflecting the innovative performance of a firm or an economy, see Grilliches (1992). Patent information used in this study considers patents issued by national patent offices. A patent gives the holder exclusive rights to produce a specific good in the way specified in the patented invention for a defined period of time. To be eligible for a patent, the invention must be novel, non-obvious and be of commercial usage, see Dernis and Kahn (2004). Griliches (1992) also showed that patents, sorted by their year of application, are strongly correlated with R&D expenditure, and thus make a good proxy for innovative activity.

In this study, based on detailed information about the patent class, source country, inventor address and application year for each patent, we count patent applications for every state in each year from 1983 to 2010. We use a Tobit model with random effects or fixed effects to regress a nonlinear transformation of patent counts on a series of policy variables representing the existence or level of renewable energy policies and a spatially weighted average for each of these policies implemented in neighboring states. State energy consumption, production and prices are used to control for state energy market conditions. State social, economic and political variables such as population growth, state per capita income and state LCV voting scores are also included. A time invariant state renewable energy technical potential variable is used to control for state heterogeneity and a variable for global wind power capacity is used to control for trends in global renewable energy development[1]. The regression results show that regulation rules and mandates such as interconnection policies, net metering and renewable portfolio standards in neighboring states tend to show a significantly positive effect, with the positive effect of renewable portfolio standards consistently highly significant. While financial incentives such as tax incentives and subsidy policies in neighboring states have a negative effect on patent application in that state. At the same time, state electricity consumption, population growth and

[1] The variable for state renewable energy technical potential is included when the random effects specification is used; the variable for global wind power capacity is included when time dummy variables are not used.

electricity price all have positive effects in increasing the number of patent applications within that state. However, states' own renewable policies don't show any significant effect on patent application. For robustness checks, we also constructed another version of cumulative policy variables[2] analogous to Yin and Powers (2010). The results are more or less the same as those when variables characterizing policy existence are used.

The rest of this paper is structured as follows. Section 1.2 gives a brief overview of the related literature. Section 1.3 presents the testing hypothesis and estimation strategy. Section 1.4 describes the data used in this study. Section 1.5 presents the empirical results. Section 1.6 concludes and provides some issues for further discussions.

1.2 Related Literature

As to the effect of environmental policy on technology innovation at the federal level, there is some empirical evidence supporting the argument that environmental policies lead to innovation, as reflected in increased patenting activity. For instance, Lanjouw and Mody (1996) used pollution abatement expenditures as a measure of environmental policy stringency in Japan, the US and Germany and found that the environmental patenting activity measured by the number of granted patents is correlated to abatement costs. With US environmental technology manufacturing data, Brunnermeier and Cohen (2003) also found that environment-related patent counts increase as pollution abatement expenditures increase. When it comes to the role played by specific policy instruments, Popp (2003) examined the effects of the 1990 Clean Air Act, which introduced a market for sulfur dioxide (SO_2) permits. He found that this market oriented

[2] A series of cumulative policy variables are also used as an alternative robustness check. For a specific policy, we assign 0 to the variable for the periods before it was first enacted, and add 1 to the variable every time it's amended. So the value of the policy variable would look like a step function over years.

environmental regulation hasn’t induced more innovation compared to the previous regulation scheme as measured by patent counts related to SO_2 pollution-control technologies. However, innovation occurred after 1990 tends to be more environmentally friendly and more efficient in removing new scrubbers.

On the effect of renewable policies at the state level, most of the existing literature has concentrated on their impact on renewable technology deployment or renewable energy production. For example, using cross-sectional time-series data from 1997 to 2009, Sarzynski et al. (2012) found that states with a subsidy program like rebates or grants have experienced more rapid growth in the capacity of grid-tied PV technology than states without these policies. They also found that renewable portfolio standards also affect the market deployment of grid-tied solar PV technology. With US state level data from 1998 to 2006, Carley (2009) found that RPS implementation has not increased the percentage of electricity generated from renewable energy sources relative to the total electricity generation, yet it has increased the total amount of electricity generated from renewable sources. By constructing a new measure for policy stringency that could more accurately characterize the incentives provided by RPS, Yin et al. (2010) found that RPS policy has significantly increased in-state renewable energy development.

In the absence of studies on the effect of renewable energy policies on renewable technology innovation at the state level, there are some related cross-country studies. De Vries and Withagen (2005) investigated the relationship between environmental policy and innovation related to SO_2 abatement, as measured by the number of patent applications in relevant patent classes. Using three different models of policy stringency, their results showed some evidence that strict environmental policies tend to lead to more innovation. In another cross-country study by Johnstone, Hascic and Popp (2010), using a panel data set comprised of 25 countries in 26 years, they examined the effect of a wide variety of policy tools on innovations of renewable technology, such as tradable energy certificates, feed-in-tariffs, production quotas and public R&D. By studying patent activity related to a number of different sources of renewable energy, they have shown that the effectiveness of different policy tools varies with the relative cost of different renewable technology sources compared to fossil fuels. Dechezlepretre and Glachant (2013) also tried to investigate the technology diffusion in wind power across OECD countries. Using patent data as an indicator of innovation activity, their results indicate that both domestic and foreign demand-pull renewable policies positively affect renewable technology innovation. However, the marginal effect of policies implemented at home is 12 times higher.

In the current study, we want to examine the effect of a variety of state renewable energy policies on innovation related to wind power and the technological spillover effect of these policies. We calculate patent counts for each state over time by identifying the state and year of application of every US patent related to wind energy using information from the database Delphion[3]. Lagged policy variables are constructed over time for each state with at least one patent application recorded over the period from 1983 to 2010. To characterize the spillover effect of neighboring states' renewable energy policy, spatially weighted policy variables are calculated with a spatial weight matrix assigning weights only to neighboring states or assigning relatively larger weights to states that are geographically closer. Then we regressed a nonlinear transformation of patent count on lagged policy variables and the spatially weighted average of neighboring states' policy variables to check whether renewable energy policies induce

[3] http://www.delphion.com/. Among commercially available on-line databases, Delphion offers the advantage of allowing exports of large amounts of data into easily readable files, such as Microsoft Excel spreadsheets. This database allows searching by patent classification, and provides detailed descriptive information on each patent, including the date of application, country of origin, and inventor address.

innovation within state and whether they have any spillover effects in boosting innovation in neighboring states.

1.3 Testing Hypothesis and Model Specification

In order to study the role played by various renewable energy policies from both within state and out of state in inducing innovation, we have considered two major categories of renewable energy policies: 1) financial incentives such as tax credits and various subsidy policies and 2) renewable energy related regulation rules and mandates such as interconnection standards, net-metering and renewable portfolio standards. In order to isolate the spillover effects of renewable technology policy, we have included in the models both a series of variables representing the existence or level of a state's own renewable energy policies and a spatially weighted average of each of these variables implemented in neighboring states. Some other factors that could potentially affect the incentives for innovative behaviors are also controlled for by including variables for energy demand, supply and price, state social, economic and political factors (Detailed definitions for these variables are provided in the data section). The benchmark reduced-form regression equation is specified as[4]:

$$f(\#Patents_{it}) \sim \beta_0 + (Policy_{it-1})\boldsymbol{\beta_1} + W_i(Policy_{t-1})\boldsymbol{\beta_2} + \beta_3(R\&D_{it-1})$$

$$+ \beta_3(ElectricityConsum_{it-1}) + \beta_4(ElectricityPrice_{it-1})$$

$$+ \beta_5(ElectricityConsumGrowth_{it-1}) + \beta_6(PopulationGrowth_{it-1})$$

$$+ \beta_7(PerCapitaIncome_{it-1}) + (PoliticalIndex_{it-1})\,\beta_8 + \varepsilon_{it}$$

[4] In this equation, explanatory variables are lagged by 1 year, we also have robustness checks with states' own policy variables and neighboring states' policy variables lagged by different number of years.

where $i = 1, \dots, 48$ are indices for the cross sectional unit (state) and $t = 1983, \dots, 2010$ are indices for time. The dependent variable is a function of patent activity measured by the number of wind technology-related patent applications for a given state in a given year. The policy variables include a set of renewable technology policies such as a tax incentives index, a subsidy policy index, interchanging rules, net metering rules and renewable portfolio standards. R&D expenditures are used to control for changes in input for innovative activities; electricity consumption, electricity consumption growth and electricity price are used to control for the energy market demand and supply; population growth nd state per capita income may also affect the demand for or affordability of renewable energy; LCV senate score and LCV house score are used to represent the state political mood toward pro-environmental legislation.

The coefficient of special interest is β_2, which characterizes the spillover effect of renewable energy policies implemented in neighboring states. W is the spatial weight matrix used to calculate neighboring states' average renewable energy policy. In this paper, we have used two versions of spatial weight matrices, the first one follows from Aichele and Gelbermayr (2012).

$$W_{ijt} = \frac{\frac{Population_{jt-1}}{Distance_{ij}}}{\sum_{j=1}^{N} \frac{Population_{jt-1}}{Distance_{ij}}} \qquad i, j = 1, \dots, N, i \neq j, t = 1, \dots, T^{5}$$

Here we use W_{ijt} to denote the element in row $(i-1) * T + t$ and column $(j-1) * T + t$ of the spatial weight matrix W, which characterizes the effect of renewable energy policy implemented in state j on innovation in state i in year t. We argue that how policy in state j will affect state i

[5] All the diagonal elements of W are zero, so this formula would not be used to calculate diagonal elements. Log of the population level is used in weight matrix construction here.

is determined by population in state j and the geographic distance between state i and state j. Distance between state i and state j is approximated as[6],

$$Distance_{ij} = R\sqrt{\left(lat_i - lat_j\right)^2 + \left(\cos\left(\left(lat_i + lat_j\right)/2\right) * \left(long_i - long_j\right)\right)^2}$$

where lat_i and $long_i$ are the latitude and longitude of the center of state i, respectively. After constructing the spatial weight matrix by dividing the population in state j by the approximate geographic distance between state i and state j, we do a row normalization so that left multiply the spatial weight matrix to a policy variable (which is a vector) is in effect to take the spatially weighted average of the values of the policy variable in neighboring states.

The second version of spatial weight matrix is the contiguity spatial weight matrix created by Luc Anselin (1988):

$$W_{ijt} = Neighbor_{ij} / \sum_{j=1}^{N} Neighbor_{ij}$$

$$Neighbor_{ij} = \begin{cases} 1 & if\ state\ i\ and\ state\ j\ are\ neighbors \\ 0 & otherwise \end{cases}$$

In this spatial weight matrix, the spillover effect only happens between states with common borders. Compared to the other spatial weight matrices, this one puts relatively more weight on policies implemented in direct neighboring states.

We will estimate the above regression equation using a Tobit model with individual effects. Most of the existing literature in this area use count data models to fit the data; censoring in the

[6] R is the radius of the earth, which equals 6,371.009 kilometers. All latitude and longitude are denominated in unit of radians.

Tobit model we used could have the same effect as the zero-inflated Poisson or zero-inflated negative binomial model, for example see Lambert (1992). More importantly, we don't have to assume a Poisson distribution or negative binomial distribution in the error term. We could circumvent this issue by referring to the Central Limit Theorem[7]. As pointed out in Baltagi (2008), if the N individuals can be seen as randomly drawn from a large population, the random effects model would be the appropriate specification. Since we are studying US continental states, we might expect the Tobit model with fixed effects is the proper specification. Given that random effects model usually gives more efficient coefficient estimates but may suffer from the endogeneity resulted from the correlation between the unobserved heterogeneity and other regressors, we use the random effects specification to get the main estimation results. The fixed effects specification is also used as a robustness check. Following the use of the random effects panel data model with attrition in Hausman and Wise (1979), the model with random effects is specified as

$$y_{it} = \log(\,1\, + \#Patents_{it}\,)$$

$$y_{it} = \begin{cases} {y_{it}}^{*} & if \quad {y_{it}}^{*} > 0 \\ 0 & otherwise \end{cases}$$

$${y_{it}}^{*} = x_{it}\beta + \epsilon_{it} = x_{it}\beta + \mu_i + u_{it} \qquad i = 1, 2, \dots, N, t = 1, 2, \dots, T$$

$$\mu_i \sim IID\ \mathcal{N}\left(0, \sigma_\mu^2\right), \quad u_{it} \sim IID\ \mathcal{N}(0, \sigma_u^2)$$

We first take a log transformation of the patent counts plus 1 because quite some states had only zero patent applications in most of the years. After the log transformation, zero patent

[7] Although we used the normal distribution in the Tobit mode with random effects specification, the results don't really rely on the normal assumption. Based on the estimation methods described below, Tobit model with random effects can be estimated with the simulated maximum likelihood method with individual random effects following any probability distribution. A distribution assumption is not needed at all for the Tobit model with fixed effects.

counts are still zero; that's why the observed patent counts after transformation are still censored at zero. All the explanatory variables are included in x_{it}, which also includes a state renewable technology technical potential and a variable for global wind capacity or a series of time dummy variables. The state individual effects are characterized by μ_i, which are assumed to follow a normal distribution with variance σ_μ^2.

The Tobit model with random effects could be estimated using the method of simulated likelihood, see Baltagi (2008). The method of simulated likelihood transforms the original likelihood function as an expectation of some function with respect to the density function of μ_i.

$$L_i = \Pr(y_{i1}, y_{i2}, \dots, y_{iT}|X) = \int \left[\prod_{t=1}^{T} \int f_1(\epsilon_{it}|\mu_i) d\,\epsilon_{it} \right] f_2(\mu_i) d\mu_i$$

$$= E_{\mu_i} \left[\prod_{t=1}^{T} \int f_1(\epsilon_{it}|\mu_i) d\,\epsilon_{it} \right] = E_{\mu_i}(h(\mu_i))$$

This function is continuously differentiable. If the expectation is finite, then the conditions for the Law of Large Numbers would be satisfied. The expectation could be calculated as the average of the function with a sample of observations $\mu_{i1}, \dots, \mu_{iR}$ plugged into the function.

$$\frac{1}{R} \sum_{r=1}^{R} h(\mu_{ir}) \xrightarrow{p} E_{\mu_i}(h(\mu_i))$$

Then a sample $\mu_{i1}, \dots, \mu_{iR}$ is drawn from the population of μ_i with a random number generator. The log of the simulated likelihood function can be written as,

$$lnL_{simulated} = \sum_{i=1}^{N} \left[ln \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \int f_1(\epsilon_{it}|\mu_{ir}) d\,\epsilon_{it} \right\} \right]$$

which is similar to the probit function used in Buttler and Moffitt (1982). Maximizing this simulated likelihood will give the estimator for β. With estimates for the coefficients in the model, the marginal effect of neighboring states' renewable policies in inducing innovation in a state would be calculated in a way similar to that used in McDonald and Moffitt (1980). For more details on the calculation of marginal effects in this nonlinear Tobit model, see Appendix 1.1.

As a robustness check, we also estimate the benchmark regression equation using the Tobit model with fixed effects. Following Maddala (1987), the Tobit model with fixed effects applied to the transformed patent counts is specified as

$$y_{it} = \log(\, 1 \, + \#Patents_{it}\,)$$

$$y_{it} = \begin{cases} {y_{it}}^* & if \quad {y_{it}}^* > 0 \\ 0 & otherwise \end{cases}$$

$$ {y_{it}}^* = x_{it}\beta + \mu_i + u_{it} \qquad u_{it} \sim IID\ \mathcal{N}(0, \sigma^2)$$

Unlike the case of a linear model, in a Tobit model with fixed effects there is no sufficient statistics for the fixed effects μ_i to condition on. So the MLE estimators of β and σ will depend on the fixed effect μ_i. Since the number of observations per cross-section unit is fixed, it is not possible to consistently estimate the fixed effects μ_i and this inconsistency will result in inconsistency in the estimators of β and σ. Green (2004) has shown that the usual MLE estimator is biased and inconsistent through Monte Carlo methods. Given this invalidity of MLE, we estimated the Tobit model with fixed effects following Honore (1992). Based on the assumption that the error terms are independent and identically distributed conditional on the explanatory variable and the individual fixed effects, Honore (1992) derived orthogonality conditions that must be satisfied by the true parameters. The minimum distance method is in turn used to estimate the parameter values by either minimizing a least absolute deviations objective function or a least squares objective function. It is shown that these estimators are consistent and asymptotically normal under suitable regularity conditions.

1.4 Data

In this study, we use patent count data for 48 continental states spanning 28 years from 1983 to 2010. The states and time period are chosen so that each of these states has at least one patent application during this time span, and for a geographic analysis we also constrained the analysis to the 48 contiguous states by excluding data for Alaska and Hawaii. Then state-level R&D data, electricity consumption, production and price information, state demographic, economic and political factors are used as control variables. In addition, time-invariant state renewable energy technical potential is used to control for state heterogeneity that might affect renewable energy deployment or renewable energy technical innovation, and a variable for aggregate global wind capacity over this period is used to control potential developing trends that may have occurred in the renewable energy industry. Details about the source, collection and manipulation of these data are given in this section.

1.4.1 Dependent Variable

The dependent variable in this paper is a log transformation of the number of patent applications for a specific state in each year. Data on relevant patent information comes from the on-line database provided by Delphion. Detailed descriptive information includes the patent class, source country, inventor address and application date. Based on a large amount of data downloaded from Delphion, we choose wind patents with the United States as the source country; then we assign a patent to a state or several states for a specific year using information on the inventors' address and the first application date. If there is only one inventor or if multiple inventors for a patent are from the same state, then the patent is assigned to that state. However, if there are multiple inventors for a patent and they are from different states, the patent will be assigned to each of those different states. The year of patent application is decided by the earliest application date. All these patents are counted as 1,582, with some of the patents having multiple inventors from different states. After identifying these patents, I count the number of patents each state has filed in each year over this period. The frequency table of patent application counts for the 48 states and the 28 year period is given in Table 1.1. As can be seen from Table 1.1, the state yearly patent counts are nonnegative. For all these states and years, 843 out of the 1,344 patent counts, which is about 62%, are 0. Most of the patent counts are 0 or 1 for these states over the 28 year period, which accounts for about 80%. The largest number of patent applications for a state in a single year is 32, which appears in South Carolina in 2009. The distribution of those non-zero patent counts seems to be skewed. These facts motivate us to use a Tobit model with individual effects to a nonlinear transformation[8] of the patent counts. Figure 1.2 shows a map with the total number of patent applications for each state from 1983 to 2010. The figure shows that nine states (CA, TX, FL, SC, VA, NY, PA, CT, MA) have more than 50 patent applications in total, on average more than 2 patents every year. Figure 1.3 through Figure 1.8 show the patent counts for each state every other five years from 1984 to 2009. These maps show that the Southwestern and Eastern parts of the country are becoming greener and greener over the years, which means the number of patent applications are increasing.

---

[8] We do the transformation $\log(1 + number\ of\ patents)$ so that the transformed patent counts are still censored at 0.

1.4.2 Policy Variables

With available data from the Database of State Incentives for Renewables and Efficiency (DSIRE), which outlines operational policy instruments across the country and the date of enactment and amendment for each policy instrument by each state, we first constructed two variables for the two major broad classes of financial incentives for renewable energy following Caley (2009) and Sarzynski et. al. (2012). One represents a number of different types of existing subsidy policies and the other tax incentives. The subsidy policies include grants, loans and rebates. If there is a grant program in a state for a given year in place, then assign one to the grant variable for the state in that year, otherwise zero. The variables for loan programs and rebate programs are defined similarly. After defining these variables, we construct the subsidy index by summing up the variables for loans, grants and rebates. So the subsidy index ranges from zero to three, indicating the number of kinds of subsidy policies in existence. Based on state corporate, personal, property and sales tax, the tax incentive index is constructed similarly ranging from zero to four indicating the number of types of tax incentives in existence for a state in a given year.

We have also used three variables for rules, regulations and mandates for renewable energy at the state level: interconnection rules, net metering and renewable portfolio. The interconnection standards usually give clear technical rules such as maximum capacity, connection voltage and connection procedure so that on-site distributed generations can connect to the electric utility grid conveniently and safely. Under the net metering policy, electricity meters could accurately record both energy inflows and outflows so that distributed generators can save excess electricity production for future credit. This provides incentive or convenience for the consumer-based small-scale renewable energy facilities such as wind or solar power to interconnect with the grid.

The last policy is renewable portfolio standard, which makes it an obligation that electricity supply companies produce a specified proportion of the increased electricity production from renewable energy sources, such as wind, solar, biomass, or geothermal. Renewable energy certificates (REC) are granted to certified generators for every unit of electricity produced from renewable sources. Earned REC can then be sold to electricity suppliers bundled with the electricity they represent. Supply companies then could use the certificates to demonstrate that they are in compliance with the regulatory obligations. Two versions of variables are constructed for these policy standards in a similar way to Yin and Powers (2010). One version represents the existence of these policies. For example, for a given year, if a renewable portfolio standard is in effect in the state; then the variable for renewable portfolio standard equals one for that state in that year, otherwise it equals zero. Based on the variable characterizing the existence of RPS, two alternative variables for RPS are constructed to account for the stringency of the policy. The first alternative considers how well states are meeting their RPS obligations. The variable is the same as the RPS existence variable except that the variable will take value 0 if the RPS are met with 100% compliance in the first year of implementation and stay at 100% afterwards. The second alternative is different from the RPS existence variable in that the variable will take value 0 if states with RPS have a target less than 5% in 2010, which is the last year of the data sample. As suggested by Yin and Powers (2010) and Menz and Vachon (2006), we also constructed a series of cumulative policy variables for these policies taking into consideration the amendment of these policies over time. Under the assumption that amended policies would be more stringent and thus could be more effective in promoting renewable energy adoption and innovation, the variable starts from zero if there is no policy in effect, one if the considered policy is in effect in the initial year (1983); then adding one to the policy variable every time the policy is amended.

In this way, the policy variable is illustrated with a step function for a given state over years. All the data on financial incentives and regulation rules and mandates for renewable energy is extracted from DSIRE.

By the end of our sample, 39 states have adopted RPS to promote renewable energy. 43 states had interconnection standards in place, and 46 states had regulations covering net metering. From the summary statistics in Table 1.2, the mean that value of the tax incentive index is 0.917. Because the tax incentive index characterizes the number of kinds of corporate, personal, sales and property taxes that are present in a state, this means on average states have about one out of the four kinds of tax incentives in all these years. Similarly, a mean value of 0.321 for the subsidy policy index means that states on average have one kind of subsidy policy in 34.6% of the years. For variables indicating the existence of regulatory rules and mandates, the renewable portfolio standard existence variable has a mean of 0.166. This means states on average have the renewable portfolio standard policy implemented in 16.6% of the years. For interconnection existence and net metering existence, the mean values of 0.192 and 0.313 indicate that states on average have these policies implemented in 19.2% and 31.3% percent of the years, respectively.

To study the spillover effect of policies enacted in neighboring states on innovation in that state, we also included a series of spatially weighted averages of policy variables in the regression. If a large proportion of a state's neighboring states have adopted some form of renewable energy policy, this might have some "demonstration effects" on the renewable technology adoption in that state. At the same time, it would be more convenient and cost-effective to learn from neighboring area's experience related to renewable technology. This provides another incentive to renewable energy development and technological innovation. To calculate the spatially weighted average of neighboring states' policy variables, we left multiply the policy variable constructed above by one of the three spatial weight matrix. See R. Leenders (2002) for a discussion of how to use a spatial weight matrix to model social influence.

1.4.3 R&D Data

When it comes to factors affecting technical innovation, we first used state R&D investment to control for the general scientific capacity of the state[9]. R&D data from 1983 to 2007 come from the Industrial Research and Development Information System (IRDIS) database[10], which contains all of the statistics produced by the National Science Foundation's Survey of Industry Research and Development (SIRD) from 1953 to 2007. The IRDIS database contains total funds spent for business R&D performed in each state over the period from 1953 to 2007, which measures the cost of firms' R&D activity. After 2007, the National Science Foundation's SIRD has been replaced by the Business Research and Development and Innovation Survey (BRDIS)[11]. The R&D data for 2008 and 2009 are from this database. The R&D time series data for each state from 1983 to 2009 are then adjusted to 2009 US dollars by using the consumer price index. In 2009, the total business spending on R&D activity in the United States was $282 billion, of which $225 billion was paid by the company. Businesses in California lead in R&D investment, accounting for over 23% of the nation's business R&D expenditures.

1.4.4 State Electricity Information

When estimating a model that accounts for the main drivers of renewable technology innovation, it is necessary to control for those potential underling trends in a state's electricity

---

[9] We want to isolate the effect of renewable policy on innovation from that resulted from increased R&D investment. Although firms may increase R&D investment in renewable technology in response to renewable policy; with the existence of spillover effect from policies implemented in neighboring states, a state may have more innovation by learning from the experience of neighboring states with the same amount of instate R&D investment.

[10] http://www.nsf.gov/statistics/iris/history_data.cfm

[11] http://www.nsf.gov/statistics/infbrief/nsf12309/

markets. So we include state electricity consumption, state electricity consumption growth, and electricity price per British Thermal Unit (BTU) in the regression. These data are extracted from the EIA's State Energy Data System (SEDS)[12], which contains detailed information on state energy consumption, production and prices by source from 1960 to 2010. The first two variables included are state electricity consumption and state electricity consumption growth. These two variables reflect the demand for electricity and to how large an extent a state needs to establish new capacity to satisfy the increase in electricity demand. If electricity consumption is high and increasing at a relatively faster pace, then the state would be under pressure to build more capacity. This could possibly provide incentives for renewable energy deployment and hopefully the increased deployment would lead to renewable technology innovation. Another variable included is electricity price, which is adjusted to 2009 US dollars by the consumer price index. It is possible that high or volatile conventional energy prices improve the potential return of investment for renewable technology, which in return provides incentives for innovation in renewable technology.

1.4.5 State Social, Economic and Political Factor

Another factor that might affect state demand for electricity is state population growth. A state with high population growth would be under more pressure to construct more capacity for electricity generation. If part of this additional capacity is built from renewable energy resources, increased deployment may lead to technology innovation improving the efficiency of electricity generation. Wealthier consumers may have a higher valuation for a good environment and thus be more likely to prefer energy produced from renewable sources. Given the usual high upfront

[12] http://www.eia.gov/beta/state/seds/seds-data-complete.cfm

cost, renewable technology is more likely to be afforadable for a consumer in states with higher per capita income, see Sarzynski et al. (2012). For example, Rodberg and Schachter (1980) have found evidence that higher income households are more likely to claim solar income tax credits. To control for state per capita income, we construct a variable by taking the log of the state per capita income adjusted to 2009 US dollars. The state population and state personal income data are from the Regional Economic Account of the Bureau of Economic Analysis (BEA)[13].

Research in political studies and public administration has found evidence that state institutional framework and other political structures could affect policy adoption and the outcomes of policy implementation. For example, Steinmo and Tolbert (1998) found that state political and economic institutions could explain state tax policy variation. Ringquist and Clark (2007) argued that state policy efforts could be affected by interparty competition and interest groups in the states. Sapat (2004) claimed that the beliefs of political figures and organizational culture could also impact government performance. In this study we use League of Conservation Voters (LCV) voting scores to account for institutional factors that may affect pro-environmental legislation that is important for renewable energy development and renewable technological innovation. The LCV voting scores data are from the National Environmental Scorecard[14] published yearly by the Congress since 1970.

1.4.6 Renewable Energy Technical Potential

We also include wind energy technical potential as an upper-bound estimate of wind energy development potential. The National Renewable Energy Laboratory (NREL)[15] estimates the

[13] http://www.bea.gov/regional/downloadzip.cfm

[14] http://scorecard.lcv.org/scorecard/archive

[15] http://www.nrel.gov/gis/

technical potential of different renewable energy sources based on factors such as availability, quality, and land-use constraint. It is important to point out that these estimates do not consider economic or market factors like the availability of transmission infrastructure, production cost, and policy or regulatory impacts; and therefore do not represent the level of actual renewable energy deployment. Figure 1.9 and Figure 1.10 show maps of two major sources of renewable energy technical potential for each state. From these maps, the Midwest and the Southwest states of the country are rich in solar and wind power.

1.4.7 Global Wind Power Capacity

To control for the possible underlying trends in the development of the renewable energy industry worldwide, we also included a variable for world cumulative installed wind power capacity as an indicator for the development level of wind power technology in the regression. Data on world cumulative installed wind power capacity and net annual addition are from the International Energy Agency[16].

1.5 Empirical Results

From my main results presented in Table 1.4, the coefficients for neighboring states' regulatory rules and mandates are almost all positive and all the coefficients for the financial incentives are negative. The most exciting result is that coefficients for neighboring states' renewable portfolio standards are all positive and significant at the 1% confidence level when the random effects or fixed effects model without time dummies is used. This means renewable portfolio standard policies implemented in neighboring states have a positive spillover effect on renewable technology innovation in that state. Without time dummy varialbes, coefficients for

[16] http://www.iea.org/

neighboring states' net metering policies are significantly positive at the 1% and 5% confidence level for the random effects Tobit model and fixed effects Tobit model, respectively[17]. However, coefficients of variables characterizing financial incentives are in general negative and significant. Specifically, the coefficient for neighboring states' tax incentives is significantly negative when random effects Tobit model without time dummies is used and close to significant for fixed effects Tobit model without time dummies. When time dummies are included in the model, the coefficient of neighboring states' tax incentives is significantly negative at the 5% and 10% confidence level for the random effects and fixed effects Tobit model including time dummies, respectively. The coefficients for neighboring states' subsidy policy are significantly negative at the 1% confidence level when the random effects Tobit or fixed effects Tobit model without time dummies is used. These results suggest that regulatory rules and mandates implemented in neighboring states have a positive spillover effect, while financial incentives implemented in neighboring states tend to have a negative spillover effect on innovation in that state.

To understand these two counter effects on innovation resulted from renewable energy policies implemented in neighboring states, we have the following tentative explanation. Renewable energy policy regulatory rules and mandates such as interconnection policy and net-metering implemented in one state would make it more convenient to connect to the grid not only for generation facilities from within state but also neighboring states. Renewable portfolio standards implemented in one state will provide incentive for utility companies to purchase electricity

[17] The coefficient for neighboring states' renewable portfolio standards is positive and close to significant at the 10% confidence level and neighboring states' net metering is significantly positive at the 10% confidence level when a set of time dummy variables for each year are included in the Tobit model with random effects. When time dummies are not included in the regression, a log global wind power capacity variable is included to control for the possible underlying trends in the development of the renewable energy industry.

produced from renewable sources from neighboring states. This could be the possible mechanism that renewable energy regulatory rules and mandates would lead to increased deployment of renewable technology in neighboring states and thus are beneficial to technology innovation in neighboring states. For tax incentives and subsidy policies, although these policies will make renewable technology more affordable to the general public and ensure a better market prospective for companies developing renewable technology, to benefit from these tax incentives or subsidy policies, one has to be resident of a state where these financial incentives are present. So if one state has implemented some form of financial incentives, it is possible to attract potential renewable technology users who would otherwise locate in neighboring states to switch to that state. This kind of negative spillover effects resulted from neighboring states' financial incentives may be seen as competing effects.

For the marginal effects of neighboring states' renewable energy policies, we will first calculate marginal effects using the sum of μ_i and u_{it} as one single normal distribution in the error term. Since both individual state random effect μ_i and the disturbance term u_{it} are assumed to be normal, their summation is normal. Then we will also calculate the marginal effect conditional on different realizations of individual state random effect μ_i. In the main results, the estimated standard deviation of μ_i and u_{it} are 0.393 and 0.952, respectively. So the variance of the sum of these two normal distributions is 1.061 (with standard deviation 1.030). With the formula for marginal effects in Appendix 1.1, plug in the mean value of the explanatory variables from the summary statistics table and their corresponding estimated coefficients from the Tobit random effects model. The marginal effects of neighboring states' tax incentive, subsidy policy, interconnection existence, net metering existence and renewable portfolio standard existence calculated as the expected changes in the number of patent applications in response to a one unit change in the corresponding policy variable are -0.5206, -1.5242, 1.5097, 1.8927, and 1.4770, respectively[18]. Next, marginal effects are also calculated when the individual random effect is evaluated at the mean value, one standard deviation below and one standard deviation above the mean value[19]. After the realization of individual random effects, random effect can be combined with the estimated constant of the model. When the realization of the random effect is at the mean level (0), the marginal effects of neighboring states' tax incentive, subsidy policy, interconnection existence, net metering existence and renewable portfolio standard existence are -0.4616, -1.3516, 1.3388, 1.6784, and 1.3098, respectively. When the random effect is realized one standard deviation below the mean (-0.393), the marginal effects of these policies are -0.2483, -0.7269, 0.7200, 0.9027, and 0.7044. The marginal effects of these policies are -0.7914, -2.3170, 2.2951, 2.8773, and 2.2454 when the random effect is realized at one standard deviation above the mean level (0.393).

Results shown in Table 1.4 also indicate that electricity consumption is significantly positive at the 1% and inflation adjusted R&D is significantly positive at the 5% confidence levels for the random effects specification, respectively. These results show that states with higher electricity consumption or higher R&D investments tend to have more innovation in renewable energy technology. The coefficient for state wind power technical potential is also significantly positive. This means that renewable technology is more likely to be deployed in states rich in wind power where the marginal cost could be lower or scale economies are more likely to appear and this increased deployment could lead to relevant technology innovation. However, in all the

[18] The direction of the changes in the number of patent applications will depend on the sign of the coefficients of neighboring states' policy variables.

[19] The standard deviation is the estimated standard deviation of μ_i. Then the variance of disturbance term used to calculate marginal effects is that of u_{it}.

regressions state's own renewable policies have not shown any statistically significant effect on technology innovation in that state.

As an alternative, we also run the Tobit model with either fixed effect or random effect for the original count variable. From results in Table 1.5, we can see although the coefficients are much larger, they show the similar sign pattern as those reported in Table 1.4. The coefficients for neighboring states' tax incentive and subsidy policy are in general negative, and the coefficients for regulatory rules and mandate are in general positive. The coefficients for neighboring states RPS are significantly negative for most of the specifications. For controlling variables, the coefficients for electricity price are significantly positive for all the specifications, and the coefficients for electricity consumption are also all positive. From the results, wind power technical potential has also shown a significantly positive effect on technology innovation. To characterize how large are the effects of neighboring states renewable policy in inducing innovation in a state, marginal effects for neighboring states' tax incentive, subsidy policy, interconnection, net metering and RPS are calculated as -0.5932, -1.9798, 1.3834, 2.2081, and 2.1825, respectively[20].

For robustness checks, we have run these regressions again with the contiguity spatial weight matrices for the log transformed count variable. Although the magnitude of the coefficients is different, the results show very similar patterns as those we get with the population/distance weighting matrix. Specifically, the coefficients for neighboring states' regulatory rules and mandates tend to be positive with the coefficient of neighboring states' renewable portfolio standards significantly positive. We also try a series of regressions with alternative RPS variables

---

[20] These marginal effects are calculated assuming all the explanatory variables are evaluated at their corresponding mean values. The average marginal effects for the five neighboring states' policy variables over all the observations of the explanatory variables are calculated as -1.9406, -6.4769, 4.5257, 7.2237, 7.1400, respectively.

and different lag structures in both the within state renewable policy variables and neighboring states' policies. Again, the results are also consistent with our baseline results that neighboring states' financial incentives have a negative spillover effect while neighboring states' regulatory rules and mandate have a positive spillover effect[21]. As robustness checks, we also run a zero inflated Poison regression for the original count variable using state patent applications from 1974 to 1982 to control for state individual effects and Poison regression with fixed or random effects. The results for these regressions are again consistent with those we have got from running a Tobit regression. All these robustness checks results are presented in Table 1.6 through Table 1.10.

From these regression results, states' own renewable energy policies haven't shown any significant effects on technology innovation so far. One suspection is that there might be multicolinearity between states' own policy variables and neighboring states' policy variables. To see whether this is the cause of the nonsignificance for states' own policy variables, we drop neighboring states' policy variables and run the regression again. The results reported in Table 1.11 show that most of the coefficients for the policy variables are positive except for RPS, although most of them are not statistically significant. This means states' own renewable energy policy could have some positive effect on technology innovation, we didn't see any significant effect in the previous regressions bacause of the multilinearity problem.

One last thing we need to worry about is the policy endogeneity issue, one possibility is that states with stronger green power characterized by more patent applications may lobby for environmental legislation more intensively. Thus the causal relationship is inverse, that's to say

[21] The coefficients in general show similar sign patterns, although their magnitude may vary with the specifications.

states with more patent applications will be more likely to have some form of renewable energy policy in place. To exclude this possibility, we run a probit or ordered probit regression on the five policy variables using similar explanatory variables as those used in the previous regressions. One variable of special interest is the state wind power technical potential variable. Since state wind power technical potential is exogenous and has shown a significant positive effect on technology innovation, we could use this variable as an instrument for technology innovation and see if it will affect renewable energy policy enactment. If it does not show any effect on whether a state will have some form of renewable energy policy, we can exclude the causal effect of technology innovation on enactment of renewable energy policy. From regression results in Table 1.12, none of the coefficients of state wind power technical potential are statistically significant except for the regression on net metering, which is significant at 10% level. This means that technology innovation will not affect the enactment of renewalbe energy policies in most of the cases.

In summary, the empirical results show that renewable energy regulatory rules and mandates such as interconnection policy, net metering and renewable portfolio standard enacted by neighboring states have a statistically significant positive spillover effect on technology innovation in that state. However, financial incentives such as tax incentives and subsidy policies implemented in neighboring states have a statistically significant negative spillover effect on technological innovation in that state. Other factors, such as R&D investment, electricity consumption, and per capita income may also affect innovation related to renewable technology. In general, the results show that states with higher electricity consumption or per capita income would have a faster pace of technological innovation.

1.6 Conclusions and Discussion

In this study, we have examined the role of state renewable energy policies in inducing renewable technological innovation and the spillover effect of renewable energy policies enacted in neighboring states on innovation in that state. We have constructed a variable for state yearly patent counts as a measure of technological innovation using patent information from the national patent office provided through Delphion. Then we do a log transformation of the state yearly patent counts and use it as the dependent variable. Explanatory variables include a series of dummy variables characterizing the existence of renewable policy variables or a series of cumulative policy variables taking into account any amendments in these policies and the spatially weighted average of these policy variables. We run the regression using a Tobit model with individual effects using energy market demand and supply factors, and state demographic, economic and political factors as controling variables. The results show that renewable technology regulatory rules and mandates such as interconnection standards, net metering and renewable portfolio standards have a significant positive spillover effect on technological innovation in that state. However, neighboring states' financial incentives, like tax incentives and subsidy policies have statistically significant negative effects on patenting activities in that state. Other factors such as electricity consumption, electricity price and population growth in a state also show significantly positive effects on technology innovation within a state. These results consistently hold water when we change the regression specification, such as changing the Tobit regression with random effects to a Tobit regression with fixed effects, including time dummy variables in the regression or using a series of alternative policy variables to take into account amendments in these policies or the stringency of state RPS.

As can be seen from the previous regression results, while neighboring states' spatially weighted average renewable energy policies have a significant effect on technological innovation in that state, the states' own policies do not show any significant effect on technological innovation within the state. We suspect this might be caused by the mulcolinearity between states' own policies and the corresponding policies enacted in neighboring states. We have tried to prove that this is the case by running a regression excluding neighboring states' policy variables. Another possibility is that renewable energy policies enacted within a state have such a strong effect on innovation in neighboring states that states surrounded by a larger proportion of neighboring states with renewable policy in place will show increased patent applications while states with only within state renewable policies will not. However, the mechanism of state's own policy enactment and its effect on innovation needs to be explored further.

Another issue that needs to be addressed is the underlying assumption in the analysis of this study that there is no policy endogeneity. Specifically, the enactment of state renewable energy policy is the result of technological innovation related to renewable energy in that state. We have ruled out the possible of endogenous policy by running a Probit or ordered Probit regression on the policy variables using state wind power technical potential as an instrument for technology innovation. Another possible explanation is that these renewable policies are implemented as a result of public understanding of the consequences of pollution and global warming. Thus it is reasonable to assume that renewable policies both at the state level and federal level are implemented in response to a larger global trend, such as a greater sense of awareness and responsibility in environmental issues.

1.7 References

Aichele, Rahel, and Gabriel Felbermayr. "Kyoto and the carbon footprint of nations." Journal of Environmental Economics and Management 63, no. 3 (2012): 336-354.

Anselin, Luc. Spatial econometrics: methods and models. Vol. 4. Springer, 1988.

Baltagi, Badi. Econometric analysis of panel data. Vol. 1. John Wiley & Sons, 2008.

Bennear, Lori Snyder. "Are management - based regulations effective? Evidence from state pollution prevention programs." Journal of Policy Analysis and Management 26, no. 2 (2007): 327-348.

Brunnermeier, Smita B., and Mark A. Cohen. "Determinants of environmental innovation in US manufacturing industries." Journal of environmental economics and management 45, no. 2 (2003): 278-293.

Cameron, A. Colin, and Pravin K. Trivedi. "Regression analysis of count dataCambridge University Press." Nueva York (1998).

Carley, Sanya. "State renewable energy electricity policies: An empirical evaluation of effectiveness." Energy Policy 37, no. 8 (2009): 3071-3081.

Carley, Sanya. "The era of state energy policy innovation: A review of policy instruments." Review of Policy Research 28, no. 3 (2011): 265-294.

Chamberlain, Gary. "Panel data." Handbook of econometrics 2 (1984): 1247-1318.

De Vries, Frans P., and Cees Withagen. "Innovation and environmental stringency: the case of sulfur dioxide abatement." (2005).

Van Pottelsberghe, Bruno, H. Denis, and Dominique Guellec. Using patent counts for cross-country comparisons of technology output. No. 2013/6227. ULB--Universite Libre de Bruxelles, 2001.

Greene, William. "Fixed effects and bias due to the incidental parameters problem in the Tobit model." Econometric Reviews 23, no. 2 (2004): 125-147.

Griliches, Zvi. "Patent statistics as economic indicators: a survey." In R&D and productivity: the econometric evidence, pp. 287-343. University of Chicago Press, 1998.

Hausman, Jerry A., and David A. Wise. "Attrition bias in experimental and panel data: the Gary income maintenance experiment." Econometrica: Journal of the Econometric Society (1979): 455-473.

Hausman, Jerry A., Bronwyn Hall, and Zvi Griliches. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." Econometrica: Journal of the Econometric Society (1984): 909-938.

Honoré, Bo E. "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects." Econometrica: journal of the Econometric Society (1992): 533-565.

Jaffe, Adam B., Richard G. Newell, and Robert N. Stavins. "Technological change and the environment." Handbook of environmental economics 1 (2003): 461-516.

Jaffe, Adam B., and Karen Palmer. "Environmental regulation and innovation: a panel data study." Review of economics and statistics 79, no. 4 (1997): 610-619.

Jaumotte, Florence, and Nigel Pain. From ideas to development: the determinants of R&D and patenting. No. 457. OECD Publishing, 2005.

Johnstone, Nick, Ivan Haščič, and David Popp. "Renewable energy policies and technological innovation: evidence based on patent counts." Environmental and Resource Economics 45, no. 1 (2010): 133-155.

Lambert, Diane. "Zero-inflated Poisson regression, with an application to defects in manufacturing." Technometrics 34, no. 1 (1992): 1-14.

Leenders, Roger Th AJ. "Modeling social influence through network autocorrelation: constructing the weight matrix." Social Networks 24, no. 1 (2002): 21-47.

McDonald, John F., and Robert A. Moffitt. "The uses of Tobit analysis." The review of economics and statistics (1980): 318-321.

Maddala, GS. Limited-dependent and qualitative variables in econometrics. No. 3. Cambridge university press, 1983.

Maddala, GS. "Limited dependent variable models using panel data." Journal of Human Resources (1987): 307-338.

Mazur, Allan, and Eric W. Welch. "The geography of American environmentalism." Environmental Science & Policy 2, no. 4 (1999): 389-396.

Menz, Fredric C., and Stephan Vachon. "The effectiveness of different policy regimes for promoting wind power: Experiences from the states." Energy policy 34, no. 14 (2006): 1786-1796.

Popp, David. "Induced Innovation And Energy Prices." American Economic Review (2002): 160-180.

Popp, David. "Pollution control innovations and the Clean Air Act of 1990." Journal of Policy Analysis and Management 22, no. 4 (2003): 641-660.

Popp, David. "International innovation and diffusion of air pollution control technologies: the effects of NO_x and SO_2 regulation in the US, Japan, and Germany." Journal of Environmental Economics and Management 51, no. 1 (2006): 46-71.

Ringquist, Evan J., and David H. Clark. "Issue definition and the politics of state environmental justice policy adoption." International Journal of Public Administration 25, no. 2-3 (2002): 351-389.

Rodberg, L., Schachter, M. State conservation and solar-energy tax programs: incentives or windfalls? The Council of State Planning Agencies, Washington, DC, 1980.

Sapat, Alka. "Devolution and innovation: the adoption of state environmental policy innovations by administrative agencies." Public Administration Review 64, no. 2 (2004): 141-151.

Sarzynski, Andrea, Jeremy Larrieu, and Gireesh Shrimali. "The impact of state financial incentives on market deployment of solar technology." Energy Policy 46 (2012): 550-557.

Schmookler, Jacob. "Invention and economic growth." (1966).

Steinmo, Sven, and Caroline J. Tolbert. "Do institutions really matter? Taxation in industrialized democracies." Comparative political studies 31, no. 2 (1998): 165-187.

Yin, Haitao, and Nicholas Powers. "Do state renewable portfolio standards promote in-state renewable generation?." Energy Policy 38, no. 2 (2010): 1140-1149.

Table 1.1 Frequency Table for State Yearly Patent Application Count (1983-2010)

Count	Frequency	Percent	Cumulative Frequency	Cumulative percent
0	843	62.72	843	62.72
1	223	16.59	1066	79.32
2	113	8.41	1179	87.72
3	49	3.65	1228	91.37
4	29	2.16	1257	93.53
5	21	1.56	1278	95.09
6	15	1.12	1293	96.21
7	9	0.67	1302	96.88
8	5	0.37	1307	97.25
9	6	0.45	1313	97.69
10	6	0.45	1319	98.14
11	1	0.07	1320	98.21
12	5	0.37	1325	98.59
13	1	0.07	1326	98.66
14	4	0.3	1330	98.96
15	2	0.15	1332	99.11
17	1	0.07	1333	99.18
18	1	0.07	1334	99.26
20	1	0.07	1335	99.33
22	2	0.15	1337	99.48
23	1	0.07	1338	99.55
24	1	0.07	1339	99.63
25	1	0.07	1340	99.7
26	1	0.07	1341	99.78
28	1	0.07	1342	99.85
31	1	0.07	1343	99.93
32	1	0.07	1344	100

Note: The data source is Delphion. The data used in this paper covers 48 states for 28 years, so the total number of state yearly counts is 1344.

Table 1.2 Summary Statistics

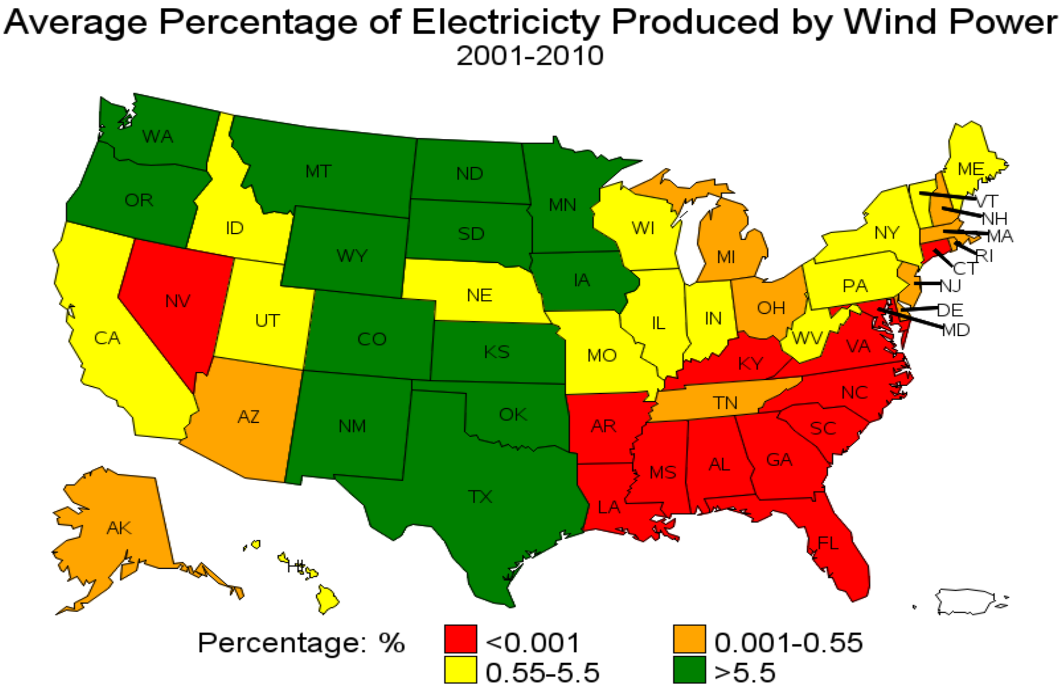
Variable	Mean	Std. Dev.	Min	Max
Log R&D Investments (Millions 2009 USD)	7.1542	1.7887	1.1886	11.1166
Log Electricity Consumption (Billion BTU)	11.8411	0.9911	9.4507	13.9846
Electricity Consumption Growth (%)	2.0597	3.7896	-21.4907	41.4432
Log Electricity Price (2009 USD per BTU)	3.3374	0.2985	2.7048	4.0979
Population Growth (%)	0.9974	1.0234	-5.9861	7.325
Log Per Capita Income (2009 USD)	10.3713	0.1921	9.8144	10.9647
Log Wind Power Technical Potential (GWh)	12.1868	2.5072	4.9972	15.5297
Log Global Wind Capacity (Megawatts)	8.7641	1.9122	4.4998	11.9765
LCV Senate Score	0.483	0.3022	0	1
LCV House Score	0.4719	0.2512	0	1
Tax Incentive Index	0.9167	1.11	0	4
Subsidy Policy Index	0.3214	0.5809	0	3
Interconnection Existence	0.192	0.394	0	1
Net Metering Existence	0.3125	0.4637	0	1
Renewable Portfolio Standard Existence	0.1659	0.3721	0	1
Interconnection Cumulative	0.2753	0.6516	0	5
Net Metering Cumulative	0.3981	0.6628	0	3
Renewable Portfolio Standard Cumulative	0.2567	0.6483	0	4
Renewable Portfolio Standard Alternative 1	0.1138	0.3177	0	1
Renewable Portfolio Standard Alternative 2	0.1503	0.3575	0	1

Note: Units are in parenthesis.

Table 1.3 Summary Statistics for Neighboring States' Average Policy

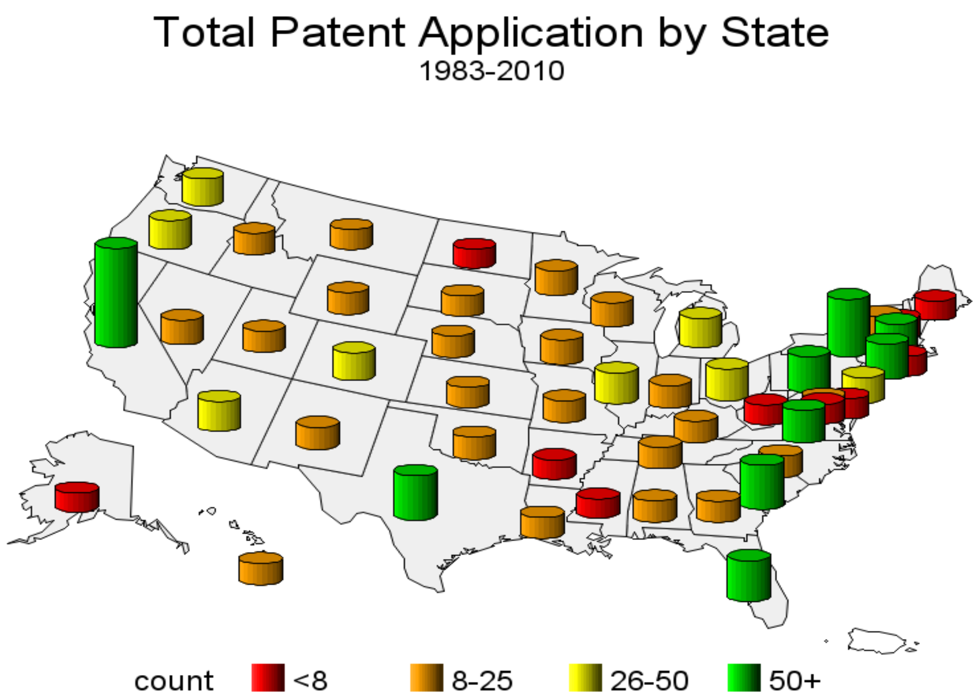
	Variable	Mean	Std.Dev.	Min	Max
Contiguity Weighting	Tax Incentive	0.9586	0.7715	0	3.6667
	Subsidy Policy	0.3254	0.3598	0	2
	Interchange Existence	0.2011	0.3031	0	1
	Net Metering Existence	0.3238	0.3515	0	1
	Renewable Portfolio Standard Existence	0.1624	0.2708	0	1
	Interconnetion Cumulative	0.2952	0.5387	0	5
	Net Metering Cumulative	0.4208	0.5387	0	3
	Renewable Portfolio Standard Cumulative	0.2521	0.4600	0	2.6667
	Renewable Portfolio Standard Alternative 1	0.1087	0.2217	0	1
	Renewable Portfolio Standard Alternative 2	0.1495	0.2490	0	1
Population/Distance Weighting	Tax Incentive	0.9450	0.5461	0.29979	2.6329
	Subsidy Policy	0.3178	0.2587	0	1.3733
	Interconnection Existence	0.1915	0.2484	0	0.9469
	Net Metering Existence	0.3091	0.2727	0	0.9593
	Renewable Portfolio Standard Existence	0.1641	0.2186	0	0.8893
	Interconnetion Cumulative	0.2755	0.3992	0	1.8743
	Net Metering Cumulative	0.3920	0.3972	0	1.6655
	Renewable Portfolio Standard Cumulative	0.2521	0.3745	0	1.9005
	Renewable Portfolio Standard Alternative 1	0.1129	0.1706	0	0.7424
	Renewable Portfolio Standard Alternative 2	0.1476	0.1846	0	0.8049

Figure 1.1 Average Percentage of Electricity Produced by Wind Power (2001-2010)



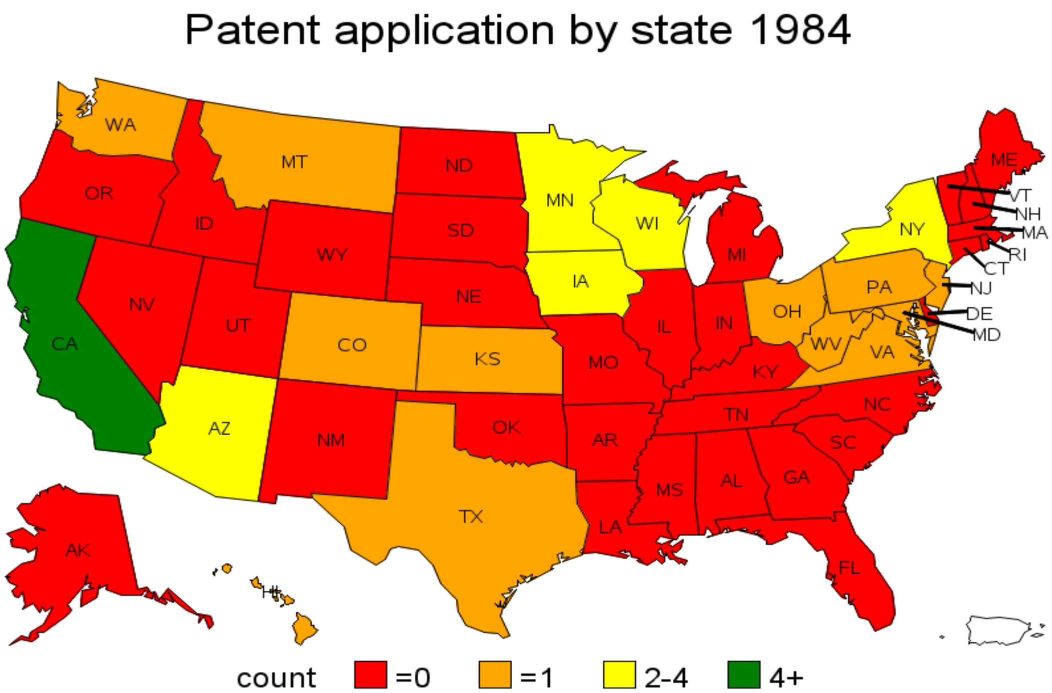
Data Source: State Energy Data System, EIA

Figure 1.2 Total Number of Patent Application by State (1983-2010)



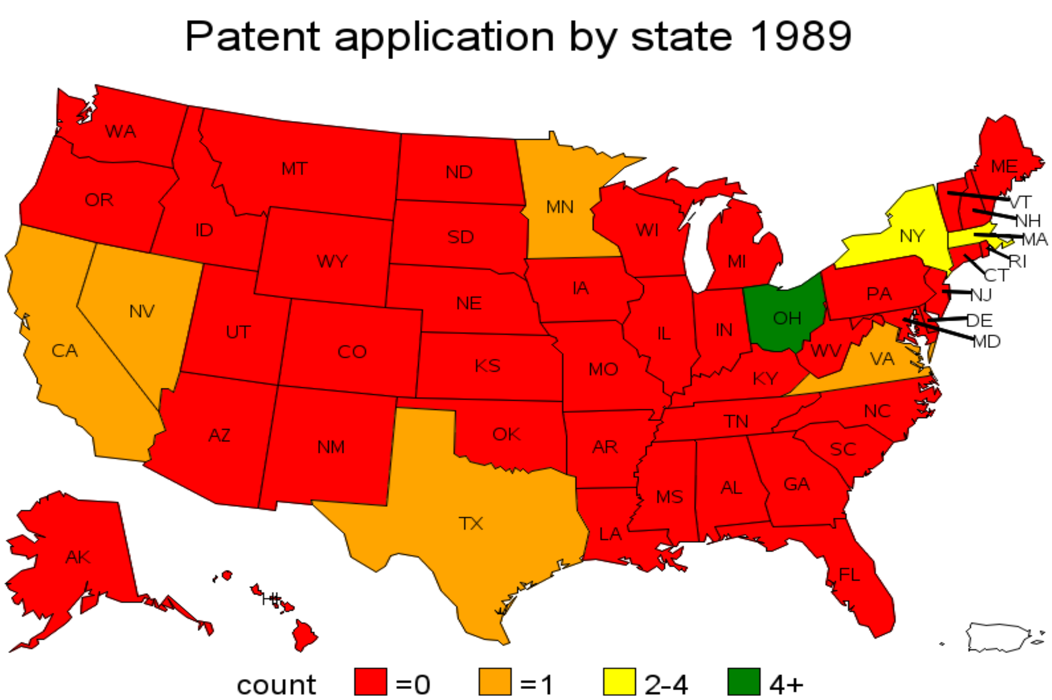
Data Source: Delphion, Author's calculation.

Figure 1.3 Patent Application by State 1984



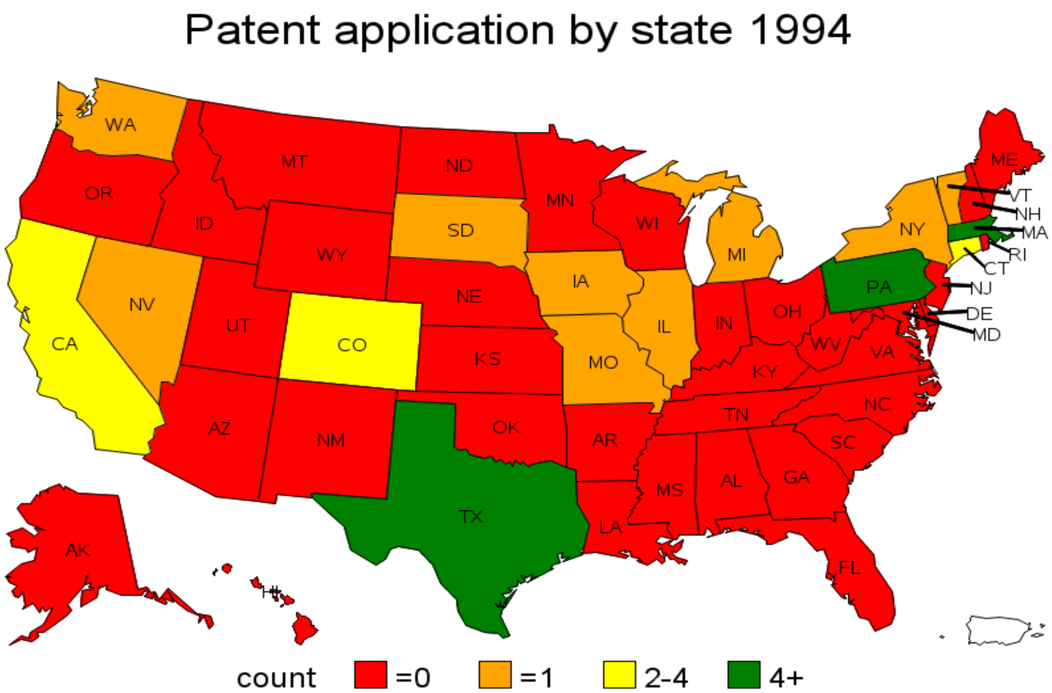
Data Source: Delphion, Authors' calculation.

Figure 1.4 Patent Application by State 1989



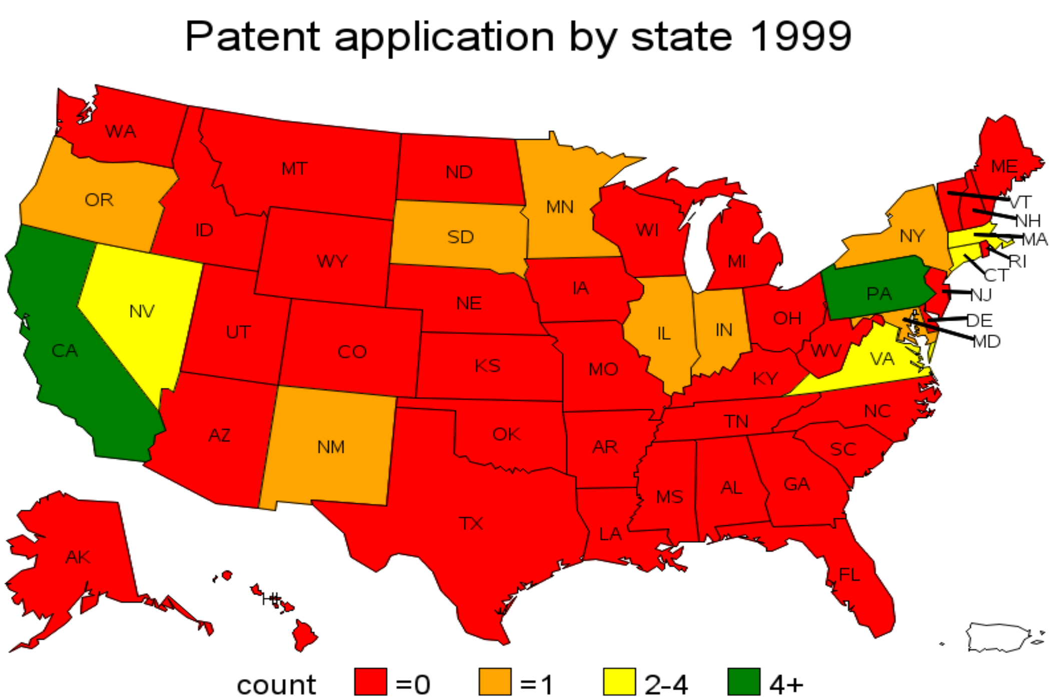
Data Source: Delphion, Authors' calculation.

Figure 1.5 Patent Application by State 1994



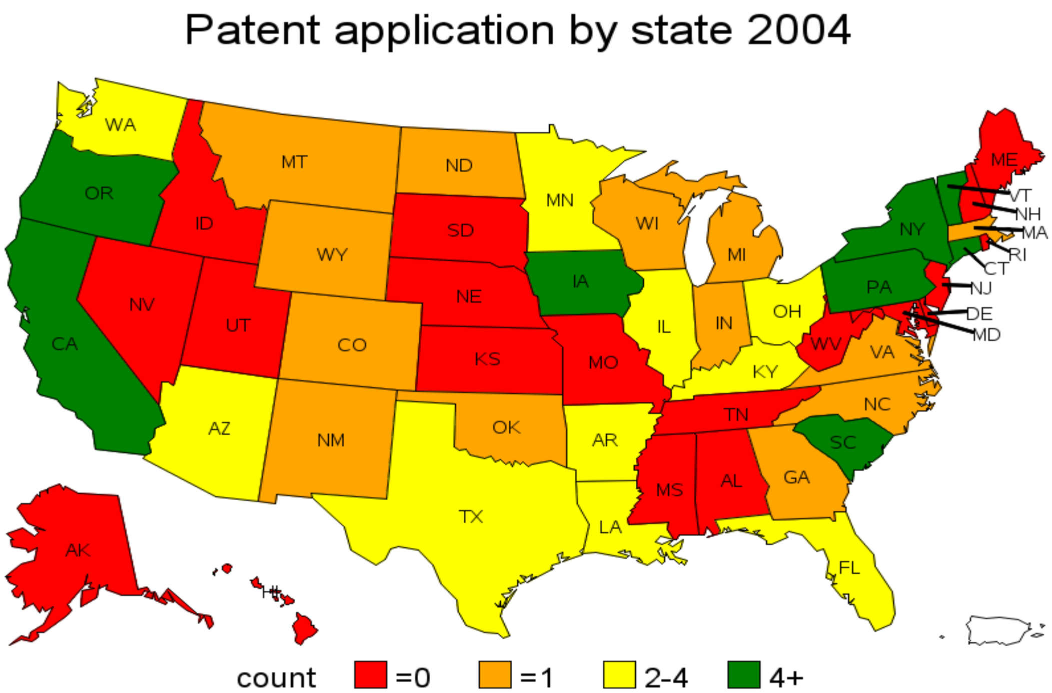
Data Source: Delphion, Author's calculation.

Figure 1.6 Patent Application by State 1999



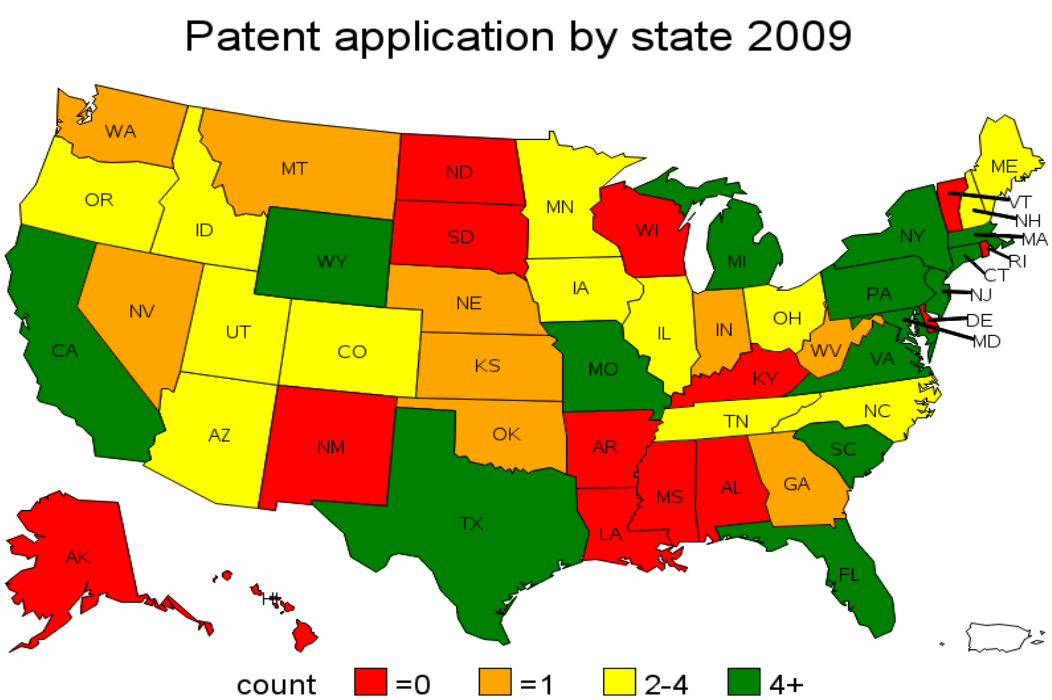
Data Source: Delphion, Author's calculation.

Figure 1.7 Patent Application by State 2004



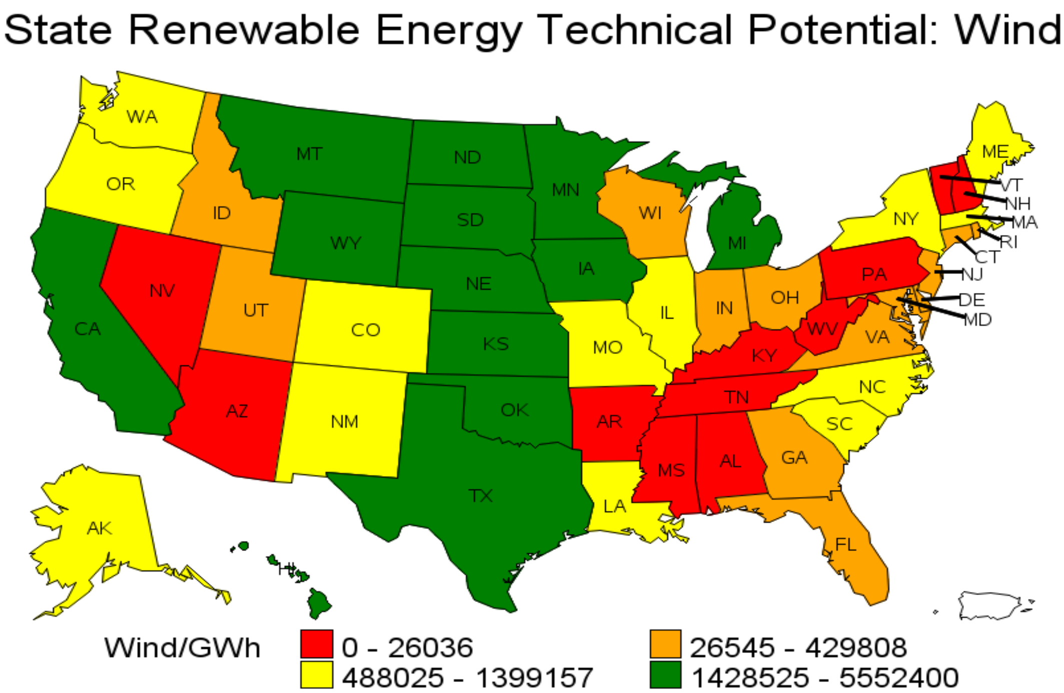
Data Source: Delphion, Author's calculation.

Figure 1.8 Patent Application by State 2009



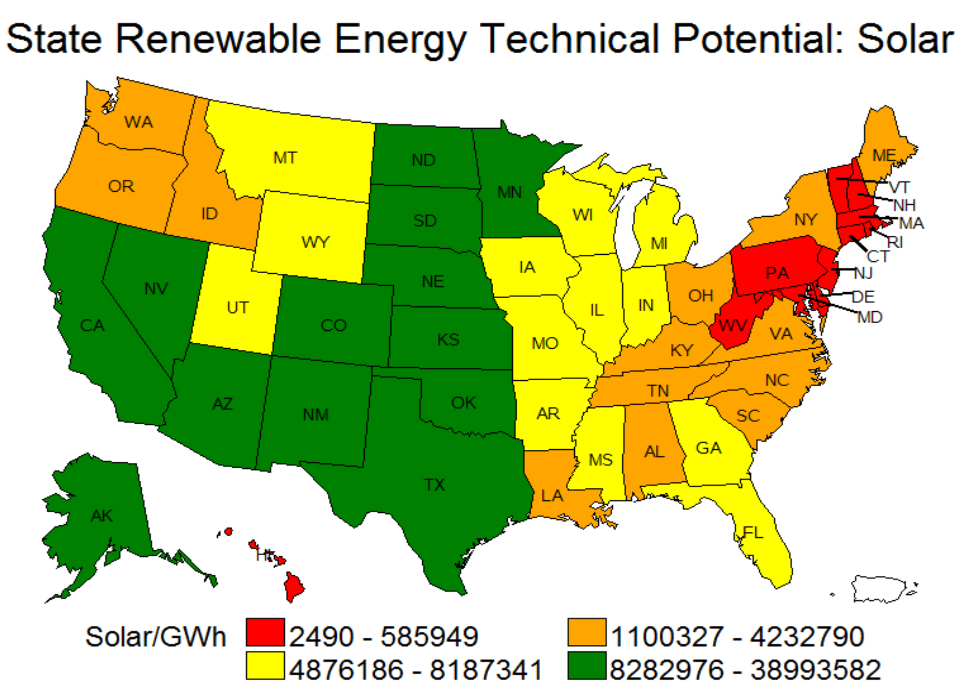
Data Source: Delphion, Author's calculation.

Figure 1.9 State Renewable Energy Technical Potential (Wind Power)



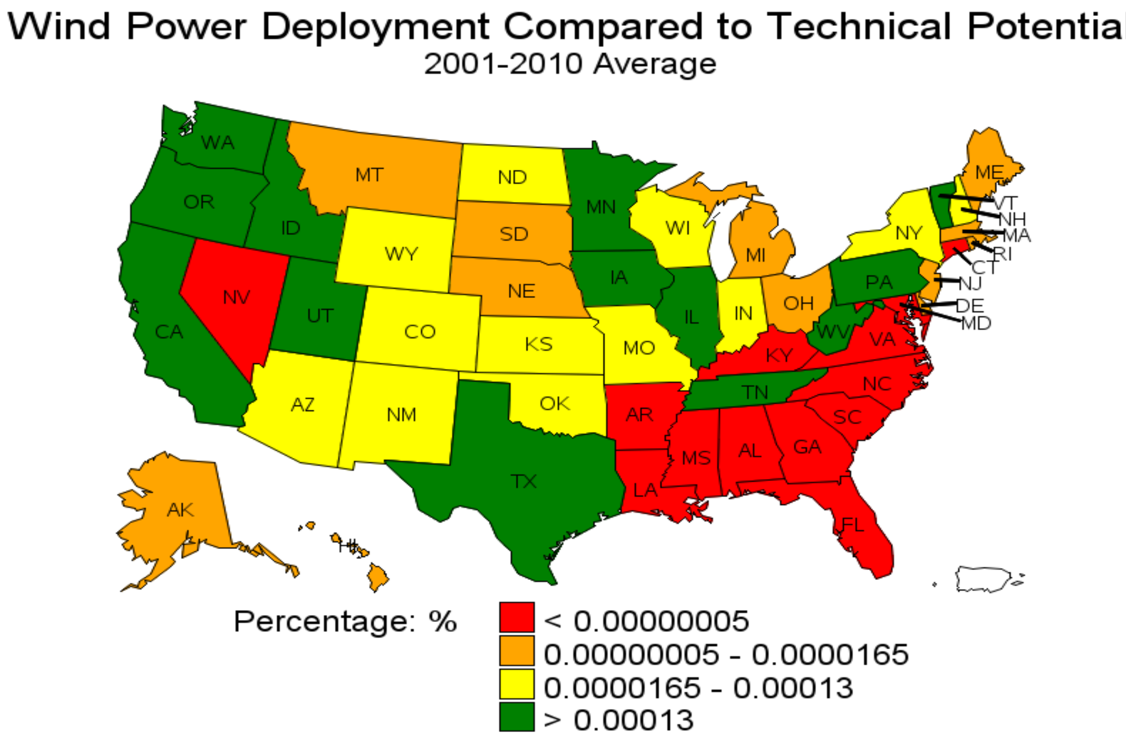
Data Source: National Renewable Energy Laboratory

Figure 1.10 State Renewable Energy Technical Potential (Solar)



Data Source: National Renewable Energy Laboratory

Figure 1.11 Wind Power Deployment Compared to Technical Potential



Data Source: National Renewable Energy Laboratory and EIA

Table 1.4 Tobit Model Regression Results with Log Population/Distance Weighting

	Random Effect		Fixed Effect	
Neighbor Tax Incentive	-0.541*	-0.806**	-0.496	-1.133*
	(0.307)	(0.397)	(0.315)	(0.620)
Neighbor Subsidy Policy	-1.584***	-0.846	-1.325***	-0.318
	(0.404)	(0.586)	(0.401)	(0.705)
Neighbor Interconnection	1.569**	0.538	1.108	-0.333
	(0.756)	(1.390)	(0.777)	(1.359)
Neighbor Net Metering	1.967***	1.651*	1.572**	0.923
	(0.600)	(0.975)	(0.665)	(0.997)
Neighbor RPS	1.535***	1.374	1.695***	0.536
	(0.579)	(0.909)	(0.336)	(1.162)
Tax Incentive Index	0.0391	0.0419	0.0481	0.0515
	(0.0500)	(0.0511)	(0.0577)	(0.0606)
Subsidy Policy Index	0.0487	0.0694	0.0472	0.0453
	(0.0769)	(0.0774)	(0.0773)	(0.0869)
Interconnection	0.173	0.111	0.231*	0.0928
	(0.136)	(0.146)	(0.122)	(0.137)
Net Metering	-0.0876	-0.0692	-0.0865	-0.00589
	(0.127)	(0.129)	(0.136)	(0.140)
RPS	-0.0413	-0.09	-0.0996	-0.144
	(0.125)	(0.123)	(0.134)	(0.128)
Log Inflation Adjusted R&D	0.112*	0.143**	-0.0383	0.0389
	(0.0632)	(0.0648)	(0.105)	(0.0938)
Log Electricity Consumption	0.372***	0.312***	0.611	0.412
	(0.107)	(0.114)	(0.584)	(0.682)
Electricity Consumption Growth	-0.0205**	-0.0214*	-0.0201*	-0.0175
	(0.0103)	(0.0113)	(0.0116)	(0.0129)
Log Electricity Price	0.224	0.386	0.349	0.643
	(0.266)	(0.288)	(0.486)	(0.515)
Population Growth	0.0836	0.0652	0.0431	0.0357
	(0.0511)	(0.0531)	(0.0789)	(0.0781)
Log Per Capita Income	1.062*	0.806	-0.169	-1.201
	(0.573)	(0.647)	(1.104)	(1.038)
Log Global Wind Capacity	-0.150**		-0.0419	
	(0.0761)		(0.0944)	
Wind Power Technical Potential	0.0639**	0.0545*		
	(0.0288)	(0.0302)		
LCV Senate Score	-0.201	-0.0503	-0.187	-0.00408
	(0.179)	(0.180)	(0.208)	(0.188)
LCV House Score	-0.165	-0.0967	-0.194	0.0749
	(0.250)	(0.261)	(0.308)	(0.352)
Time Dummy	No	Yes	No	Yes

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 1.5 Tobit Model Regression Results with Original Count Variable

	Random Effect		Fixed Effect	
Neighbor Tax Incentive	-1.941	-2.853*	-0.479	-3.775
	(1.280)	(1.672)	(3.505)	(6.471)
Neighbor Subsidy Policy	-6.477***	-3.414	-9.869*	-5.471
	(1.648)	(2.421)	(5.435)	(8.046)
Neighbor Interconnection	4.526	2.797	4.286	-4.51
	(3.133)	(5.743)	(7.865)	(11.23)
Neighbor Net Metering	7.224***	7.185*	1.744	5.22
	(2.496)	(4.064)	(4.461)	(6.481)
Neighbor RPS	7.140***	7.062*	10.93**	5.213
	(2.368)	(3.772)	(4.541)	(11.18)
Tax Incentive Index	0.428**	0.435**	2.948**	2.632**
	(0.214)	(0.218)	(1.249)	(1.278)
Subsidy Policy Index	0.326	0.437	1.007	1.267
	(0.318)	(0.320)	(0.731)	(0.909)
Interconnection	1.027*	0.902	0.754	0.483
	(0.559)	(0.604)	(1.210)	(1.548)
Net Metering	-0.23	-0.223	0.903	1.178
	(0.530)	(0.538)	(2.086)	(1.633)
RPS	-0.161	-0.318	-1.243	-1.332
	(0.509)	(0.506)	(1.444)	(1.526)
Log Inflation Adjusted R&D	0.184	0.301	-0.577	-0.0394
	(0.277)	(0.283)	(1.011)	(0.971)
Log Electricity Consumption	1.838***	1.668***	3.793	1.082
	(0.474)	(0.502)	(5.220)	(5.733)
Electricity Consumption Growth	-0.0763*	-0.0844*	-0.127*	-0.150*
	(0.0426)	(0.0472)	(0.0730)	(0.0782)
Log Electricity Price	2.710**	3.458***	7.522**	9.872***
	(1.147)	(1.245)	(3.037)	(3.128)
Population Growth	0.245	0.162	-0.143	-0.155
	(0.217)	(0.226)	(0.609)	(0.419)
Log Per Capita Income	3.383	1.74	8.949	2.305
	(2.471)	(2.778)	(7.755)	(9.670)
Log Global Wind Capacity	-0.373		-0.117	
	(0.320)		(0.521)	
Wind Power Technical Potential	0.257**	0.231*		
	(0.130)	(0.135)		
LCV Senate Score	-0.894	-0.44	-1.59	-1.175
	(0.745)	(0.753)	(1.641)	(1.812)
LCV House Score	-1.234	-1.201	-1.79	-1.34
	(1.058)	(1.103)	(1.702)	(1.770)
Time Dummy	No	Yes	No	Yes

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 1.6 Tobit Model Regression Results with Contiguity Weighting

	Random Effect		Fixed Effect	
Neighbor Tax Incentive	0.021	-0.111	-0.0917	-0.262*
	(0.0963)	(0.108)	(0.121)	(0.144)
Neighbor Subsidy Policy	-0.473***	-0.352**	-0.37	-0.285
	(0.153)	(0.160)	(0.266)	(0.228)
Neighbor Interconnection	0.606**	-0.0498	0.739***	-0.107
	(0.256)	(0.297)	(0.284)	(0.380)
Neighbor Net Metering	0.444*	0.449*	0.237	0.279
	(0.232)	(0.246)	(0.285)	(0.265)
Neighbor RPS	0.841***	0.516**	0.498*	0.171
	(0.231)	(0.240)	(0.279)	(0.238)
Tax Incentive Index	0.0693	0.0497	0.0986	0.0533
	(0.0487)	(0.0494)	(0.0761)	(0.0589)
Subsidy Policy Index	0.0109	0.0547	0.0221	0.0341
	(0.0776)	(0.0776)	(0.0792)	(0.0869)
Interconnection	0.243*	0.063	0.288**	0.0953
	(0.133)	(0.135)	(0.117)	(0.121)
Net Metering	-0.0637	-0.0318	-0.0129	0.0344
	(0.126)	(0.125)	(0.126)	(0.134)
RPS	0.103	-0.0368	0.00715	-0.0989
	(0.125)	(0.123)	(0.141)	(0.112)
Log Inflation Adjusted R&D	0.110*	0.127*	-0.0034	0.00833
	(0.0623)	(0.0647)	(0.106)	(0.101)
Log Electricity Consumption	0.374***	0.335***	0.329	0.25
	(0.104)	(0.110)	(0.651)	(0.523)
Electricity Consumption Growth	-0.0190*	-0.0203*	-0.018	-0.0141
	(0.00996)	(0.0113)	(0.0131)	(0.0113)
Log Electricity Price	0.0955	0.357	0.47	0.499
	(0.260)	(0.283)	(0.588)	(0.417)
Population Growth	0.0467	0.0691	-0.0193	0.022
	(0.0500)	(0.0531)	(0.0600)	(0.0873)
Log Per Capita Income	1.508***	0.788	1.442*	-0.662
	(0.540)	(0.632)	(0.821)	(1.285)
Log Global Wind Capacity	-0.111*		-0.0374	
	(0.0616)		(0.0708)	
Wind Power Technical Potential	0.0512*	0.0558*		
	(0.0275)	(0.0298)		
LCV Senate Score	-0.334*	-0.0492	-0.267	-0.0113
	(0.181)	(0.181)	(0.193)	(0.223)
LCV House Score	-0.31	-0.0943	-0.225	0.0514
	(0.252)	(0.261)	(0.308)	(0.342)
Time Dummy	No	Yes	No	Yes

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Log transformed count variable is used.

Table 1.7 Tobit Model Regression Results with Alternative Policy Variables

	Cumulative Policy		RPS Alternative 1		RPS Alternative 2	
Neighbor Tax Incentive	-0.199	-0.708*	-0.525*	-0.826**	-0.491	-0.839**
	(0.286)	(0.380)	(0.306)	(0.396)	(0.306)	(0.398)
Neighbor Subsidy Policy	-1.889***	-0.778	-1.482***	-0.689	-1.406***	-0.566
	(0.482)	(0.630)	(0.392)	(0.557)	(0.391)	(0.560)
Neighbor Interconnection	0.702	0.533	1.712**	0.456	1.690**	0.127
	(0.609)	(0.895)	(0.754)	(1.396)	(0.756)	(1.381)
Neighbor Net Metering	1.722***	1.143*	1.958***	1.999**	1.921***	2.218**
	(0.451)	(0.673)	(0.599)	(0.923)	(0.626)	(0.970)
Neighbor RPS	0.354	0.173	1.484**	0.666	1.380*	0.239
	(0.403)	(0.561)	(0.684)	(0.927)	(0.718)	(0.986)
Tax Incentive Index	0.0461	0.0405	0.0369	0.0423	0.0421	0.0445
	(0.0499)	(0.0498)	(0.0501)	(0.0511)	(0.0500)	(0.0511)
Subsidy Policy Index	0.0523	0.0935	0.0417	0.0633	0.0554	0.0775
	(0.0797)	(0.0791)	(0.0758)	(0.0760)	(0.0768)	(0.0769)
Interconnection	0.00189	-0.0235	0.167	0.0958	0.167	0.0847
	(0.0853)	(0.0880)	(0.136)	(0.146)	(0.136)	(0.146)
Net Metering	-0.0304	-0.012	-0.0837	-0.0607	-0.0841	-0.0445
	(0.0878)	(0.0874)	(0.126)	(0.126)	(0.127)	(0.128)
RPS	-0.069	-0.0973	0.134	0.0825	0.0183	-0.0578
	(0.0739)	(0.0733)	(0.125)	(0.123)	(0.124)	(0.123)
Time Dummy	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Tobit model with random effect is used for log transformed variable. Coefficients are reported for policy variables only.

Table 1.8 Tobit Model Regression Results with Alternative Lags in Policies

In State Policy Lags	1	1	1	2	2	2	3	3	3
Out of State Policy Lags	1	2	3	1	2	3	1	2	3
Neighbor Tax Incentive	-0.541*	-0.488	-0.416	-0.464	-0.438	-0.316	-0.453	-0.47	-0.357
	(0.307)	(0.308)	(0.319)	(0.309)	(0.311)	(0.320)	(0.311)	(0.314)	(0.324)
Neighbor Subsidy Policy	-1.584***	-1.184**	-0.0583	-1.572***	-1.190**	-0.042	-1.596***	-1.221**	-0.122
	(0.404)	(0.509)	(0.619)	(0.404)	(0.508)	(0.622)	(0.408)	(0.514)	(0.621)
Neighbor Interconnection	1.569**	0.605	0.535	1.419*	0.464	0.292	1.500**	0.617	0.419
	(0.756)	(0.762)	(0.767)	(0.752)	(0.760)	(0.768)	(0.755)	(0.757)	(0.767)
Neighbor Net Metering	1.967***	3.214***	3.910***	2.046***	3.324***	4.105***	1.977***	3.242***	4.071***
	(0.600)	(0.612)	(0.614)	(0.594)	(0.607)	(0.617)	(0.590)	(0.601)	(0.613)
Neighbor RPS	1.535***	1.223**	-0.527	1.540***	1.282**	-0.469	1.571***	1.335**	-0.324
	(0.579)	(0.594)	(0.627)	(0.580)	(0.589)	(0.625)	(0.581)	(0.591)	(0.619)
Tax Incentive Index	0.0391	0.0573	0.0843*	0.0543	0.0707	0.0926*	0.0304	0.0428	0.0592
	(0.0500)	(0.0498)	(0.0491)	(0.0516)	(0.0515)	(0.0512)	(0.0532)	(0.0531)	(0.0527)
Subsidy Policy Index	0.0487	0.0342	0.0576	0.0695	0.0649	0.083	0.0142	0.0111	0.0331
	(0.0769)	(0.0771)	(0.0777)	(0.0847)	(0.0844)	(0.0855)	(0.0925)	(0.0923)	(0.0926)
Interconnection	0.173	0.117	0.122	0.0203	-0.00595	-0.0385	-0.0186	0.00378	-0.015
	(0.136)	(0.135)	(0.136)	(0.139)	(0.138)	(0.138)	(0.142)	(0.142)	(0.142)
Net Metering	-0.0876	-0.0799	-0.0434	-0.0899	-0.112	-0.0697	-0.0223	-0.0845	-0.0736
	(0.127)	(0.126)	(0.125)	(0.129)	(0.128)	(0.127)	(0.130)	(0.131)	(0.129)
RPS	-0.0413	-0.0623	-0.0659	0.0332	-0.0143	-0.0368	0.0308	-0.00099	-0.049
	(0.125)	(0.125)	(0.125)	(0.128)	(0.128)	(0.129)	(0.133)	(0.133)	(0.133)

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Coefficients are reported for policy variables only. Log transformed count variable is used.

Table 1.9 Zero Inflated Poisson Model Regression Results

	Model 1		Model 2		Model 3		Model 4	
Variable	Count	Inflate	Count	Inflate	Count	Inflate	Count	Inflate
Previous Patent Count				-0.0670***				-0.0833
				(0.0199)				(0.0594)
Neighbor Tax Incentive	0.531	2.968***	0.476	2.532**	1.201	5.006**	1.113	4.812*
	(0.722)	(0.998)	(0.726)	(1.148)	(0.911)	(2.406)	(0.975)	(2.799)
Neighbor Subsidy Policy	-2.207***	0.956	-2.184***	1.29	-1.208	1.544	-1.109	2.428
	(0.614)	(1.438)	(0.614)	(1.235)	(1.247)	(5.275)	(1.180)	(3.850)
Neighbor Interconnection	0.339	-5.604**	0.458	-5.161**	-0.302	-4.991	-0.295	-3.941
	(1.229)	(2.562)	(1.219)	(2.422)	(2.318)	(20.71)	(2.000)	(14.70)
Neighbor Net Metering	0.11	-6.085**	0.216	-4.633*	-0.426	-10.65**	-0.183	-8.134**
	(0.871)	(3.051)	(0.845)	(2.667)	(1.320)	(5.040)	(1.280)	(4.140)
Neighbor RPS	1.672**	-0.207	1.510**	-1.961	2.339	2.559	2.241	0.248
	(0.662)	(2.297)	(0.598)	(2.365)	(1.842)	(6.924)	(2.118)	(7.179)
Tax Incentive Index	0.0137	0.00678	0.00888	0.0243	0.0502	0.0516	0.0417	0.0457
	(0.0673)	(0.169)	(0.0672)	(0.148)	(0.0695)	(0.234)	(0.0741)	(0.219)
Subsidy Policy Index	0.0443	-0.0503	0.0417	0.0356	0.0851	-0.0534	0.0909	0.123
	(0.0763)	(0.221)	(0.0781)	(0.236)	(0.0788)	(0.370)	(0.0800)	(0.388)
Interconnection	-0.0363	-0.144	-0.0426	-0.363	-0.107	0.104	-0.117	-0.105
	(0.171)	(0.448)	(0.160)	(0.331)	(0.200)	(0.888)	(0.184)	(0.598)
Net Metering	0.457**	0.464	0.455***	0.623**	0.437***	0.299	0.442***	0.476
	(0.181)	(0.403)	(0.173)	(0.315)	(0.166)	(0.803)	(0.149)	(0.489)
RPS	-0.323	-0.718*	-0.31	-0.770**	-0.419*	-1.148*	-0.396	-1.083
	(0.255)	(0.431)	(0.266)	(0.372)	(0.248)	(0.668)	(0.270)	(0.922)
Time Dummy	No		No		Yes		Yes	

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Coefficients are reported for policy variables only.

Table 1.10 Regression Results for Poisson Model with Fixed or Random Effects

	Random Effect		Fixed Effect	
Neighbor Tax Incentive	-0.531**	-0.743*	-0.580*	-1.055**
	(0.271)	(0.391)	(0.301)	(0.433)
Neighbor Subsidy Policy	-1.647***	-0.145	-1.581***	0.0393
	(0.283)	(0.475)	(0.295)	(0.494)
Neighbor Interconnection	1.802***	-0.879	1.862***	-1.196
	(0.640)	(1.024)	(0.674)	(1.071)
Neighbor Net Metering	1.656***	2.305***	1.605***	1.835*
	(0.521)	(0.893)	(0.535)	(0.941)
Neighbor RPS	1.563***	-0.431	1.638***	-0.657
	(0.413)	(0.768)	(0.433)	(0.799)
Tax Incentive Index	0.134***	0.131***	0.136**	0.123**
	(0.0481)	(0.0492)	(0.0527)	(0.0537)
Subsidy Policy Index	0.0453	0.0703	0.0525	0.0639
	(0.0599)	(0.0633)	(0.0642)	(0.0684)
Interconnection	0.0428	-0.148	0.0245	-0.202
	(0.104)	(0.120)	(0.108)	(0.125)
Net Metering	0.143	0.299**	0.178	0.359***
	(0.110)	(0.119)	(0.115)	(0.125)
RPS	-0.242***	-0.393***	-0.258***	-0.394***
	(0.0935)	(0.0982)	(0.0977)	(0.103)
Log Inflation Adjusted R&D	-0.0732	0.0202	-0.186**	-0.062
	(0.0759)	(0.0800)	(0.0908)	(0.100)
Log Electricity Consumption	0.752***	0.615***	0.934**	0.757
	(0.135)	(0.145)	(0.459)	(0.483)
Electricity Consumption Growth	-0.0304***	-0.0325***	-0.0284***	-0.0296***
	(0.00941)	(0.0105)	(0.00956)	(0.0107)
Log Electricity Price	1.235***	1.178***	1.294***	1.122***
	(0.272)	(0.300)	(0.351)	(0.377)
Population Growth	0.0896*	0.0887	0.0656	0.0807
	(0.0531)	(0.0560)	(0.0592)	(0.0633)
Log Per Capita Income	2.162***	1.304*	2.125**	-0.209
	(0.610)	(0.727)	(0.856)	(1.126)
Log Global Wind Capacity	-0.0798		-0.0809	
	(0.0788)		(0.0916)	
Wind Power Technical Potential	0.0637	0.0536		
	(0.0428)	(0.0459)		
LCV Senate Score	-0.286*	-0.0547	-0.341**	-0.116
	(0.156)	(0.162)	(0.163)	(0.169)
LCV House Score	-0.598**	-0.202	-0.767**	-0.228
	(0.282)	(0.303)	(0.312)	(0.337)
Time Dummy	No	Yes	No	Yes

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 1.11 Tobit Model Regression Results with In-State Policy Varialbes ONLY

	Random Effect		Fixed Effect	
Tax Incentive Index	0.155***	0.0752	0.200***	0.0734
	(0.0500)	(0.0494)	(0.0714)	(0.0688)
Subsidy Policy Index	0.0768	0.102	0.0767	0.0386
	(0.0798)	(0.0771)	(0.105)	(0.0738)
Interconnection	0.243*	0.0171	0.202	0.0426
	(0.135)	(0.133)	(0.148)	(0.147)
Net Metering	0.0816	0.0493	0.111	0.0583
	(0.128)	(0.124)	(0.140)	(0.145)
RPS	0.141	-0.0863	0.0458	-0.132
	(0.127)	(0.124)	(0.142)	(0.147)
Log Inflation Adjusted R&D	0.142**	0.151**	-0.00349	0.0064
	(0.0645)	(0.0644)	(0.111)	(0.104)
Log Electricity Consumption	0.260**	0.275***	0.0507	0.0561
	(0.105)	(0.106)	(0.583)	(0.726)
Electricity Consumption Growth	-0.0213**	-0.0199*	-0.018	-0.017
	(0.0101)	(0.0114)	(0.0128)	(0.0144)
Log Electricity Price	0.415	0.45	0.635	0.497
	(0.261)	(0.284)	(0.475)	(0.439)
Population Growth	0.0263	0.0751	-0.0372	0.043
	(0.0522)	(0.0534)	(0.0659)	(0.0919)
Log Per Capita Income	1.408**	0.574	1.754*	-1.026
	(0.559)	(0.631)	(0.957)	(1.344)
Log Global Wind Capacity	0.0367		0.0434	
	(0.0519)		(0.0834)	
Wind Power Technical Potential	0.0421	0.0590*		
	(0.0296)	(0.0302)		
LCV Senate Score	-0.268	0.00942	-0.169	-0.00551
	(0.185)	(0.180)	(0.180)	(0.227)
LCV House Score	-0.0341	0.0789	-0.0992	0.164
	(0.257)	(0.259)	(0.280)	(0.336)
Time Dummy	No	Yes	No	Yes

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 1.12 Probit and Ordered Probit Regression Results for Policy Variables

	RPS	Net Metering	Interconnection	Tax Incentive Index	Subsidy Policy Index
	Probit Random Effects			Ordered Probit Random Effects	
Wind Power Technical Potential	0.622	0.470*	-0.428	0.2	0.151
	(0.488)	(0.276)	(0.366)	(0.176)	(0.245)
Log Inflation Adjusted R&D	1.102	-0.512	-0.216	-0.0766	-0.117
	(1.028)	(0.443)	(0.658)	(0.108)	(0.163)
Log Electricity Consumption	-1.91	-0.948	1.628	0.158	-1.158
	(1.754)	(0.851)	(1.227)	(0.375)	(0.866)
Electricity Consumption Growth	-0.0701	-0.047	-0.0795	0.00245	0.00385
	(0.0971)	(0.0452)	(0.0611)	(0.0135)	(0.0215)
Log Electricity Price	-4.687	-13.31***	-3.531	-2.259***	0.798
	(2.964)	(2.430)	(3.469)	(0.526)	(0.808)
Population Growth	0.377	-1.417***	-0.43	0.0754	-0.362**
	(0.689)	(0.295)	(0.676)	(0.0770)	(0.148)
Log Per Capita Income	72.72***	66.64***	47.98***	3.394**	-5.227**
	(6.196)	(4.395)	(3.878)	(1.400)	(2.313)
LCV Senate Score	4.376**	1.358	0.489	0.232	0.414
	(2.189)	(0.929)	(1.198)	(0.243)	(0.408)
LCV House Score	1.317	-1.468	-2.756	0.0268	1.057**
	(2.874)	(1.768)	(2.241)	(0.352)	(0.495)
Time Dummy	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Appendix 1.1 Calculation of Marginal Effects in a Nonlinear Tobit Model

For the nonlinear Tobit Model given in section 1.3,

$$y_{it} = \log(1 + \#Patents_{it}) = \begin{cases} x_{it}\beta + \epsilon_{it} & if\ x_{it}\beta + \epsilon_{it} > 0 \\ 0 & if\ x_{it}\beta + \epsilon_{it} \le 0 \end{cases}$$

Since the marginal effect is the response in the number of patents corresponding to a unit change in a specific policy variable, first write the number of patents applications C_{it} as,

$$C_{it} \equiv \#Patents_{it} = \exp(y_{it}) - 1 = \begin{cases} \exp(x_{it}\beta + \epsilon_{it}) - 1 & if\ x_{it}\beta + \epsilon_{it} > 0 \\ 0 & if\ x_{it}\beta + \epsilon_{it} \le 0\text{'} \end{cases}$$

where $\epsilon_{it} = \mu_i + u_{it}$. Since $\mu_i \sim IID\ \mathcal{N}(0,\ \sigma_\mu^2)$ and $u_{it} \sim IID\ \mathcal{N}(0, \sigma_u^2)$, ϵ_{it}would be distributed as $IID\ \mathcal{N}(0, \sigma_u^2 + \sigma_\mu^2)$, denote $\sigma^2 = \sigma_u^2 + \sigma_\mu^2$.

Now, the expected value of the number of patent applications is calculated as,

$$E(C) = E(C|x\beta + \epsilon > 0)\Pr(x\beta + \epsilon > 0) + E(C|x\beta + \epsilon \le 0)\Pr(x\beta + \epsilon \le 0)$$

$$= E(C|x\beta + \epsilon > 0)\Pr(x\beta + \epsilon > 0)$$

$$= \Pr(\epsilon > -x\beta)\, E[(\exp(x\beta) * \exp(\epsilon) - 1)|\epsilon > -x\beta]$$

$$= F\left(\frac{x\beta}{\sigma}\right)\left[\exp(x\beta)\int_{-x\beta}^{+\infty}\exp(\epsilon)\ \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\epsilon^2}{2\sigma^2}\right)}{F(\frac{x\beta}{\sigma})}d\epsilon - 1\right]$$

$$= \exp\left(\frac{\sigma^2}{2}\right)\exp(x\beta)\, F\left(\frac{x\beta}{\sigma} + \sigma\right) - F(\frac{x\beta}{\sigma})$$

where F is the cumulative distribution function and f would be the probability density function of the standard normal distribution.

Given the fact that $\frac{\partial F(\frac{x\beta}{\sigma})}{\partial x_j} = \frac{\beta_j}{\sigma} f(\frac{x\beta}{\sigma})$ and $\frac{\partial F(\frac{x\beta}{\sigma}+\sigma)}{\partial x_j} = \frac{\beta_j}{\sigma} f(\frac{x\beta}{\sigma} + \sigma)$, the effect of a change in the j^{th} variable of x on the expected value of C could be calculated as,

$$\frac{\partial E(C)}{\partial x_j} = \beta_j \exp\left(\frac{\sigma^2}{2}\right) \exp(x\beta) \left[F\left(\frac{x\beta}{\sigma} + \sigma\right) + \frac{1}{\sigma} f\left(\frac{x\beta}{\sigma} + \sigma\right)\right] - \frac{\beta_j}{\sigma} f\left(\frac{x\beta}{\sigma}\right)$$

Chapter 2: Nonparametric Prediction for Spatially Dependent Data

2.1 Introduction

Spatial data has become important in a variety of research areas such as econometrics, urban and regional science, environmental studies, image processing and many others. For an overview of theoretical and applied econometric studies involving spatial issues, see Aenselin and Bera (1998), Case (1991), Cliff and Ord (1973), Conley (1999), Kelejian and Robinson (1993) and Lee (2004). A primary goal of studying spatial data in economics is to understand and explain spatial spillovers where factors or outcomes in one location are correlated with outcomes in nearby locations. Prediction with spatially dependent data involves predicting the un-sampled outcome in a specific spatial location with outcomes sampled at nearby locations and factors (regressors) of all locations.

Typically, dependence between spatial units is modeled as a function of spatial distance, whether the distance can be geographic or economic, analogous to the lag structure in modeling time series data. However, the challenge is that unlike the unidirectional flow of time, no natural order is exhibited in spatial data. The problem with standard spatial econometric models is that they are parametric and rely heavily on an assumed model structure; however, the true model characterizing the interaction between spatial units is unknown, see Manski (1993) and Partridge (2012). In this paper, we will apply nonparametric estimation methods to prediction with spatially dependent data simulated on a regular integer lattice on a random field $\mathbb{Z}^N, N \geq 1$. Specifically, we will compare the prediction performances of the nonparametric k-Nearest Neighbors estimator of conditional expectation proposed by Li and Tran (2009), the Nadraya-

Watson estimator suggested by Lu and Chen (2002) and P.M. Robinson (2011), and the local linear kernel estimator proposed by Hallin et al. (2004b) and Jenish (2012) when used with spatially dependent data. The prediction performance measured by root mean squared error (RMSE) is also compared to that of the maximum likelihood method assuming a linear data generating process.

There are two popular parametric model specifications in spatial econometrics. The first one is the spatial autoregressive model (SAR) used by Kelejian and Prucha (1998), which uses a weighted average of nearby values of the dependent variable as an explanatory variable: $\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\beta + \boldsymbol{u} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}(\boldsymbol{X}\beta + \boldsymbol{u})$, where $\boldsymbol{W}$ is an $n \times n$ spatial weight matrix. The second specification is the spatial error model of Anselin (1996), uses a similar structure to directly model spatial relationship in the error term: $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u}$, with $\boldsymbol{u} = \theta\boldsymbol{W}\boldsymbol{u} + \boldsymbol{e} = (\boldsymbol{I} - \theta\boldsymbol{W})^{-1}\boldsymbol{e}$. Although the two models provide convenient specifications to account for spatial relationship and may be used to estimate the relationship of neighboring dependent variable or explanatory variable on the dependent variable itself, the overuse of these models are still questionable since it is natural to expect an increase in explanatory power when adding a weighted average of nearby values of the dependent variable as explanatory variable. Actually, spatial econometric models like the SAR and spatial error model are fundamentally unidentified. We need to estimate $n(n-1)$ parameters characterizing the potential relationships among the observations in the spatial weight matrix with only n observations. The standard spatial econometric models tried to overcome this problem by assuming that the spatial weight matrix is known *a priori*, which reduces the number of parameters to be estimated from $n(n-1)$ to a single parameter (ρ or θ). According to Hansen (2004), in the nonparametric approaches we don’t assume that the true model structure is known and admit that the fitted models are approximations of the true model. In order to strike a balance between the estimation bias and variance, the nonparametric approach choose the bandwidth to minimize the overall measure of fit, usually mean squared error (MSE) or root mean squared error (RMSE).

Given that the true model is unknown and the spatial weight matrix could be specified with error, this paper investigates the prediction performances measured by RMSE of various nonparametric estimation methods when applied to spatially dependent data. The performance of these methods are also compared to that of the maximum likelihood method assuming a linear data generating process which deviates slightly from the true data generating process. In the model, dependence structure is added to cross sectional data based on the geographic locations of individual observations. An observation for each individual is a realization of a random process at a point in a Euclidean space: the random field $\mathbb{Z}^N$. The distribution of a random vector observed at a set of locations will be affected by the economic distance between these locations since closer locations will exhibit stronger dependence in their observations. A common approach to model spatial dependence is to assume that the data is normally distributed with a covariance structure decided by the distances between the sampling points. Case (1991), Kelejian and Prucha (1998) and Lee (2004) have used a spatial analogue to the auto-regressive moving-average (ARMA) model for time series data to capture spatial dependence across individuals over space. Given a realization of the random field $\mathbb{Z}^N$ at all points over an m by n rectangular grid, we first estimate the model by nonparametric methods following Li and Tran (2009), Lu and Chen (2002), P.M. Robinson (2011), Hallin et al. (2004b) and Jenish (2012). To fit the data with taable likelihood method assuming a linear data generating process, we first transform the data sample over the m by n rectangular grid into an mn dimensional vector. Based on the dependence structure of the random fields over the m by n rectangular grid, we construct an mn

by mn spatial weight matrix whose elements are functions of the pair-wise distances between the sample points with corresponding indexes over the rectangular. We then estimate $Vec(Y) = \rho WVec(X) + \epsilon$ by maximum likelihood method. The RMSE of these different estimation techniques are then calculated and compared.

The rest of this paper is organized as follows, we will briefly talk about the related literature in section 2.2. The model setup and estimation procedures are introduced in section 2.3. In section 2.4, we will implement the Monte Carlo simulations and discuss the results. In section 2.5, we will use the three nonparametric estimation methods and maximum likelihood method to predicting crop yield with precipitation for US counties within corresponding agriculture belts. Finally, we will conclude in section 2.6.

2.2 Related Literature

Tremendous effort has been expended in the modeling and estimation of spatial dependence. There is a long tradition of studying lattice data sampled over an equally spaced domain in each of the $N \geq 2$ dimensions. For example, Whittle (1954) studied stationary processes in the plane as an extension to the autoregressive moving-average process in time series, shedding light on the fact that the realization of a time series process is only influenced by its past values, while dependence extends in all directions for a spatial process. Guyon (1982) and Dahlhaus and Kunsch (1987) considered the estimation of parameters of a stationary process on a N-dimensional lattice and obtained efficient parameter estimators taking into account the "edge effect".

Due to its flexibility and ease of use, the spatial autoregressive (SAR) models of Cliff and Ord (1981) has become very popular. In a spatial autoregressive model, the spatial observations or disturbances are modeled as a linear combination of the corresponding observations or disturbances for neighboring locations. The transformation matrix captured by the spatial weight matrix is assumed to be known except for a single unknown parameter. Various estimation methods have also been proposed to estimate SAR models. For example, Lee (2004) investigated the asymptotic properties of the maximum likelihood estimator (MLE) and the quasi-maximum likelihood estimator (QMLE) for the SAR model under the normality assumption. It is shown that the spatial weight matrix plays a role in the rates of convergence of the MLE and QMLE. The MLE and QMLE may have a rate of convergence of the square root of the sample size and limiting normal distribution when each unit is influenced by only a few neighboring units. On the other hand, if each unit can be influenced by many neighbors, the estimators for some parameter components may have a slower rate of convergence and may even be inconsistent. Conley (1999) proposed a class of spatial GMM estimators based on moment conditions deduced from economic theory. The estimator is obtained by minimizing an objective function derived from the sample counterpart of the population moments. The only assumption is that the data is stationary and spatially mixing. Since there is no model specification, this GMM estimation technique is more robust in the sense that it can avoid the misspecification problem of MLE. Kelejian and Prucha (1999) suggested a generalized method of moments estimator for the parameter of the spatial lag term in the SAR model. It is shown that the estimator is consistent under some regular conditions and computationally simple relative to the MLE and QMLE.

Given the scope for parametric and structural misspecification, researcher have turned to nonparametric methods to model and estimate spatial data. Under some mixing conditions, Tran (1990) established the asymptotic normality of the kernel density estimators for random fields indexed by a set of N-dimensional integer lattice points. Hallin et al. (2004a) studied kernel density estimators for spatial processes with linear or nonlinear structures. Convergence results are obtained for either mixing or non-mixing processes. Lu and Chen (2002) and Lu and Chen (2004) investigated weak consistency as well as convergence rates of the Nadraya-Watson estimator for a spatial conditional regression model under spatial mixing.

A recent paper by Robinson (2011) established consistency and asymptotic distribution theory for the Nadaraya-Waston estimator by extending strong mixing time series to $N \geq 2$ dimensional random fields. In their nonparametric regression, they assumed a linear structure that covers both lattice linear autoregressive-moving-average model and SAR models. The model can account for both short- and long-range dependence by assuming that the dependence between two observations decreases as the distance between them increases. The asymptotic theory is of the "increase domain" variety, where the distance between two closest neighboring observations is fixed, rather than the "infill asymptotics" on a bounded domain, where the distance between two closest neighboring observations tends to 0. Li and Tran (2009) used the nearest neighbor methods to estimate the conditional expectation for multivariate data observed over the integer lattice points in the $N-$dimensional Euclidean space and have obtained the asymptotic normality of the estimator under general mixing assumptions.

Compared to the Nadaraya-Waston estimator, the local linear kernel estimator is advantageous in that it has reduced bias and has better boundary properties as pointed out by Fan (1993) and Fan and Yao (2003). Lu and Linton (2007) have considered local linear modeling in a time series context and have proved a central limit theorem for the local linear estimator under near epoch dependence, which is weaker than the strong ($\alpha-$) mixing conditions extensively studied in the time series literature[22]. Some recent studies have applied the local linear kernel

---

[22] Although a form of strong dependence analogous to long memory in time series is allowed under the framework of Robinson (2011), nonparametric estimation in the case of strong dependence in the data generating process is an interesting while under-studied area.

estimator to the spatial random field framework. Hallin et al. (2004b) proposed a local linear kernel estimator for the conditional expectation of a spatial process observed over a rectangular domain and proved the asymptotic normality of the estimator and its first derivative as the size of the rectangular domain goes to infinity at different rates in different directions under general mixing conditions. Jenish (2012) has extended the result of the local linear estimator for stationary mixing random fields to a even larger class of dependent random fields. The uniform consistency and asymptotic normality of the local linear kernel estimator are established under near-epoch dependence and the results also allow heterogeneous data-generating process and unevenly spaced data.

Baltagi and Li (2006) considered prediction in a panel data model with spatial autocorrelation in the context of the demand for liquor. They compared the prediction performance of various estimators such as OLS, fixed effects with and without spatial correlation, random-effects GLS ignoring spatial correlation, and a random-effects estimator accounting for spatial correlation. Forecast performance based on RMSE showed that estimators taking into account spatial correlation and state heterogeneity outperform other estimators. Druska and Horrace (2004) have developed an estimator for the panel data model where disturbances are spatially correlated in the cross-sectional dimension as an extension to the cross-sectional model of Kelejian and Prucha. Under the stochastic frontier framework, they have applied the approach to a panel of Indonesian rice farms where spatial correlations represent productivity shock spillovers pertinent to geographic proximity and weather.

2.3 Model Setup and Estimation Procedures

Following Li and Tran (2009), Hallin et al. (2004b) and Li and Tran (2009), let $\mathbb{Z}^N, N \geq 1$ be the integer lattice point set in the N-dimensional Euclidean space. A point $\boldsymbol{i}$ in $\mathbb{Z}^N$ is decided by its coordinate $(i_1, i_2, \dots, i_N)$ and will be referred to as a site in the space. Spatial data will be modeled as realizations of random fields, which are in essence vector stochastic processes indexed by $\boldsymbol{i} \in \mathbb{Z}^N$. In this paper, we will consider the following strictly stationary $(d+1)$-dimensional random fields[23]:

$$\{(Y_{\boldsymbol{i}}, \boldsymbol{X}_{\boldsymbol{i}}); \boldsymbol{i} \in \mathbb{Z}^N\}.$$

where $Y_{\boldsymbol{i}}$ takes values in $\mathbb{R}$, and $\boldsymbol{X}_{\boldsymbol{i}}$ takes value in $\mathbb{R}^{\boldsymbol{d}}$, and $(Y_{\boldsymbol{i}}, \boldsymbol{X}_{\boldsymbol{i}})$ are defined over some probability space $(\Omega, \mathcal{F}, P)$.

In econometrics, we are usually interested in how a vector $\boldsymbol{X}_{\boldsymbol{i}}$ of covariates will affect the response variable $Y_{\boldsymbol{i}}$ with the data exhibiting spatial dependence either in $\boldsymbol{X}_{\boldsymbol{i}}$ or $Y_{\boldsymbol{i}}$. Specifically, assuming that $Y_{\boldsymbol{i}}$ has finite expectation, the quantity of interest in a spatial regression problem might be:

$$g(\boldsymbol{x}) \equiv E[Y_{\boldsymbol{i}} | \boldsymbol{X}_{\boldsymbol{i}} = \boldsymbol{x}]$$

The structure of spatial dependence remains unspecified. Let g be a well-defined, real-valued $\boldsymbol{x}$-measurable function. In a particular case where $\boldsymbol{X}_{\boldsymbol{i}}$ itself is measurable with respect to a subset of $Y_{\boldsymbol{j}}$s, with $\boldsymbol{j}$ ranging over some neighborhood of $\boldsymbol{i}$, g is called a spatial autoregressive function. To estimate the spatial regression equation, Li and Tran (2009) have used the nearest neighbor (kNN) method to estimate the conditional expectation. Robinson (2012) and Lu and Chen (2002) used the Nadaraya-Watson kernel estimator. Hallin et al. (2004b) and Jenish (2012) have used the local linear kernel estimator in the spatial framework. This paper will use the nearest

[23] Here d is the dimension of $\boldsymbol{X}_{\boldsymbol{i}}$, means the number of covariates observed on a geographic site $\boldsymbol{i}$. Note that $\boldsymbol{i}$ in bold (same for small in bold) is a point in the N dimensional Euclidean space, which has N coordinates $(i_1, i_2, \dots, i_N)$ with $1 \le i_k \le n_k$ for $k = 1, \dots, N$. For the Monte Carlo simulation, we take $d = 1$, $N = 2$ and denote $\boldsymbol{i} \equiv (i, j)$ with $1 \le i \le m$ and $1 \le j \le n$.

neighbor method, Nadaraya-Watson estimator and local linear kernel method to prediction in the context of spatially dependent data.

Refer to a point $\boldsymbol{i} = (i_1, i_2, \ldots, i_N) \in \mathbb{Z}^N$ as a site, and let $\mathcal{S}$ and $\mathcal{S}'$ be two sets of sites. Denote the σ-field generated by the random variables $(Y_{\boldsymbol{n}}, \boldsymbol{X}_{\boldsymbol{n}})$ with $\boldsymbol{n}$ in $\mathcal{S}$ and $\mathcal{S}'$ as Borel fields, $\mathcal{B}(\mathcal{S}) = \mathcal{B}((Y_{\boldsymbol{n}}, \boldsymbol{X}_{\boldsymbol{n}}), \boldsymbol{n} \in \mathcal{S})$ and $\mathcal{B}(\mathcal{S}') = \mathcal{B}((Y_{\boldsymbol{n}}, \boldsymbol{X}_{\boldsymbol{n}}), \boldsymbol{n} \in \mathcal{S}')$, respectively. Define the distance between $\mathcal{S}$ and $\mathcal{S}'$ as

$$dist(\mathcal{S}, \mathcal{S}') = \min\{\hat{d}(\boldsymbol{i}, \boldsymbol{j}) : \boldsymbol{i} \in \mathcal{S}, \boldsymbol{j} \in \mathcal{S}'\},$$

where $\widehat{dist}(\boldsymbol{i}, \boldsymbol{j}) = \|\boldsymbol{i} - \boldsymbol{j}\| = \sqrt{(i_1 - j_1)^2 + \cdots + (i_N - j_N)^2}$ is the distance between two points $\boldsymbol{i}$ and $\boldsymbol{j}$. Assume that $(Y_{\boldsymbol{n}}, \boldsymbol{X}_{\boldsymbol{n}})$ is observed on a rectangular region $I_{\boldsymbol{n}} = \{\boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^N, 1 \le i_k \le n_k, k = 1, \ldots, N\}$, the asymptotic behavior of nonparametric estimators is often derived as $\boldsymbol{n} \to \infty$, where $\boldsymbol{n} \to \infty$ if $\min(n_1, \ldots, n_N) \to \infty$ and denote $\widehat{\boldsymbol{n}} \equiv \prod_{k=1}^{N} n_k$.

For the asymptotic normality of the k-nearest neighbor estimator and the spatial local linear estimator, define the mixing coefficient as

$$\alpha\big(\mathcal{B}(\mathcal{S}), \mathcal{B}(\mathcal{S}')\big) = sup\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(\mathcal{S}), B \in \mathcal{B}(\mathcal{S}')\}$$

for $\mathcal{S}, \mathcal{S}' \subset \mathbb{Z}^N$. $\mathcal{B}(\mathcal{S})$ and $\mathcal{B}(\mathcal{S}')$ are σ-algebra generated by $\mathcal{S}$ and $\mathcal{S}'$. Li and Tran (2009) and Hallin et al. (2004b) assumed a general mixing condition for the random field[24]. Lu and Linton (2007) and Jenish (2012) studied the local linear kernel estimator for time series data and data

[24] Hallin et al. (2004b) also obtained asymptotic results for random fields with dependence structure characterized by $\alpha\big(\mathcal{B}(\mathcal{S}), \mathcal{B}(\mathcal{S}')\big) \le \varphi(Card(\mathcal{S}), Card(\mathcal{S}'))\phi(\hat{d}(\mathcal{S}, \mathcal{S}'))$, where $Card(\mathcal{S})$ denotes the cardinality of $\mathcal{S}$ and φ is a positive valued symmetric function non-decreasing in each variable and assumed to satisfy: $\varphi(m, n) \le \min(m, n) \quad or \quad \varphi(m, n) \le C(m + n + 1)^{\ell}$ for some $\ell > 1$ and $C > 0$. The function ϕ is assumed to satisfy $\lim_{n \to \infty} m^a \sum_{j=m}^{\infty} j^{N-1} (\phi(j))^{\delta/2+\delta} = 0$ for some $a > (4 + \delta)N/(2 + \delta)$.

sampled on integer lattice points in a random field under the near epoch dependence, respectively.

To estimate a model for spatially dependent data using nonparametric methods, the random field used as the data generating process is generally assumed to be strictly stationary, with positive and continuous density function for $\boldsymbol{X_i}$, finite absolute moments of order slightly higher than 2 for $Y_{\boldsymbol{i}}$, and continuous and twice differentiable regression equation $g(\boldsymbol{x})$ at all $\boldsymbol{x}$, see for example Lu and Chen (2002) and Hallin et al. (2004b). For the kernel function $K: \mathbb{R}^d \to \mathbb{R}$ and its corresponding bandwidth used in the estimation, Hallin et al. (2004b) used the following condition. For any $\boldsymbol{c} = (c_0, \boldsymbol{c}_1^T)^T \in \mathbb{R}^{d+1}$, the absolute value of $K_{\boldsymbol{c}}(\boldsymbol{u}) \equiv (c_0 + \boldsymbol{c}_1^T\boldsymbol{u})K(\boldsymbol{u})$ is uniformly bounded and integrable. The bandwidth $b_{\boldsymbol{n}}$ tends to zero at a rate such that $\widehat{\boldsymbol{n}}{b_{\boldsymbol{n}}}^d \to \infty$ as $\boldsymbol{n} \to \infty$.

2.3.1 k-Nearest Neighbor Estimation

The k-nearest neighbor estimator of Li and Tran (2009) for $g(\boldsymbol{x})$ is the average of the k response variables in the neighborhood of $\boldsymbol{x}$ consisting of those X's that are the k closest to $\boldsymbol{x}$, and is defined as:

$$\hat{g}_n(\boldsymbol{x}, b_{\boldsymbol{n}}) = \frac{\sum_{\boldsymbol{i}\in I_{\boldsymbol{n}}} I_{\{\boldsymbol{X_i}:\, \|(\boldsymbol{X_i}-x)/b_{\boldsymbol{n}}\| \le 1\}} Y_{\boldsymbol{i}}}{\sum_{\boldsymbol{i}\in I_{\boldsymbol{n}}} I_{\{\boldsymbol{X_i}:\, \|(\boldsymbol{X_i}-x)/b_{\boldsymbol{n}}\| \le 1\}}}$$

where I is the indicator function. Define $\Gamma_{\boldsymbol{n}} \equiv \Gamma_{\boldsymbol{n}}(\boldsymbol{x})$ to be the distance between $\boldsymbol{x}$ and its k^{th} nearest neighbor among the $\boldsymbol{X_i}$s, The k-nearest neighbor estimators are defined by $\hat{g}_{\boldsymbol{n}}(\boldsymbol{x}, \Gamma_{\boldsymbol{n}})$. Li and Tran (2009) has shown that if the data generating process for $(Y_{\boldsymbol{n}}, \boldsymbol{X_n})$ satisfies some mixing conditions, then $\hat{g}_{\boldsymbol{n}}(\boldsymbol{x}, \Gamma_{\boldsymbol{n}})$ is consistent, asymptotically normal and converges to $g(\boldsymbol{x})$ at rate

$k(\boldsymbol{n})^{1/2}$, where $k(\boldsymbol{n})$ is a fixed integer sequence tending to infinity as $\widehat{\boldsymbol{n}} \to \infty$ at such a rate that $k(\boldsymbol{n})\, \widehat{\boldsymbol{n}}^{-2/(d+2)} \to 0$ and $k(\boldsymbol{n})^{-1/2}\widehat{\boldsymbol{n}}^{\theta} \to 0$ for some $\theta > 0$[25].

2.3.2 Nadraya-Watson Estimator

Lu and Chen (2002) and P.M. Robinson (2011) have proposed and studied the asymptotic behavior of the Nadaraya-Watson estimator for spatial data. Lu and Chen (2002) investigated the weak consistency and convergence rate of the Nadaraya-Watson estimator for a non-isotropic mixing spatial data process. Weak consistency of the estimator is also obtained when sample size tends to infinity at different rates along different directions in space. P.M. Robinson (2011) considered a nonparametric kernel estimator for the conditional mean of a dependent variable given the explanatory variables assuming a linear disturbance process allowing for long range or short-range dependence. They have also established sufficient conditions for consistency and asymptotic normality of the kernel regression estimator.

The Nadraya-Watson estimator of $g(\boldsymbol{x})$ suggested by Lu and Chen (2002) and P.M. Robinson (2011) is

$$\hat{g}_n(\boldsymbol{x}, b_{\boldsymbol{n}}) = \frac{\hat{v}_n(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} = \frac{\left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}\right)^{-1} \sum_{\boldsymbol{i} \in I_{\boldsymbol{n}}} Y_{\boldsymbol{i}} K\left(\frac{\boldsymbol{X_i} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right)}{\left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}\right)^{-1} \sum_{\boldsymbol{i} \in I_{\boldsymbol{n}}} K\left(\frac{\boldsymbol{X_i} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right)}$$

Where $K(*)$ is the kernel and $b_{\boldsymbol{n}}$ is a scalar bandwidth sequence such that $b_{\boldsymbol{n}} \to 0$ as $n \to \infty$.

As shown in P.M. Robinson (2011), if the sequence $\boldsymbol{X_i}$ is identically distributed with probability density function $f(\boldsymbol{x})$. Then $\hat{f}_n(\boldsymbol{x}) = \left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}\right)^{-1} \sum_{\boldsymbol{i} \in I_{\boldsymbol{n}}} K\left(\frac{\boldsymbol{X_i} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right)$ is an consistent

[25] For the asymptotic normality to hold, it is assumed that $\phi(x) = O(e^{-\xi x})$ for $\xi > 0$. The conclusion also holds if $\phi(x) = O(x^{-\mu})$ for some μ large enough.

estimator for $f(\boldsymbol{x})$ and $\hat{v}_n(\boldsymbol{x}) = \left(\hat{\boldsymbol{n}}{b_{\boldsymbol{n}}}^d\right)^{-1}\sum_{\boldsymbol{i}\in I_{\boldsymbol{n}}} Y_{\boldsymbol{i}}K\left(\frac{\boldsymbol{X_i}-\boldsymbol{x}}{b_{\boldsymbol{n}}}\right)$ consistently estimates $g(\boldsymbol{x})f(\boldsymbol{x})$, so $\hat{g}_n(\boldsymbol{x})$ is a consistent estimate of $g(\boldsymbol{x})$, that's to say $\hat{g}_n(\boldsymbol{x}) \xrightarrow{p} g(\boldsymbol{x})$. Under some additional regularity conditions for the kernel and the density function, P.M. Robinson (2011) has also shown the consistency and the asymptotic normality of $\hat{g}_n(\boldsymbol{x})$ when $\boldsymbol{X_i}$ is not identically distributed.

2.3.3 Local Linear Kernel Estimation

Hallin et al. (2004b) and Jenish (2012) have established asymptotic results for the local linear kernel estimator with spatial data. The basic idea in local linear regression involves approximating $g(\boldsymbol{x})$ in the neighborhood of $\boldsymbol{x}$ as

$$g(\boldsymbol{z}) \cong g(\boldsymbol{x}) + \left(g'(\boldsymbol{x})\right)^T(\boldsymbol{z}-\boldsymbol{x}) \equiv a_0 + {\boldsymbol{a}_1}^T(\boldsymbol{z}-\boldsymbol{x})$$

Then, $g(\boldsymbol{x})$ and $g'(\boldsymbol{x})$ can be estimated using the kernel method rather than just nonparametrically estimating the conditional expectation $g(\boldsymbol{x})$. The estimator is constructed as

$$\begin{pmatrix} g_{\boldsymbol{n}}(\boldsymbol{x}) \\ {g_{\boldsymbol{n}}}'(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} \hat{a}_0 \\ \hat{\boldsymbol{a}}_1 \end{pmatrix} \equiv \arg \min_{(a_0,a_1)\in\mathbb{R}^{d+1}} \sum_{\boldsymbol{j}\in I_{\boldsymbol{n}}} \left(Y_{\boldsymbol{j}} - a_0 - {\boldsymbol{a}_1}^T(\boldsymbol{X_j}-\boldsymbol{x})\right)^2 K\left(\frac{\boldsymbol{X_j}-\boldsymbol{x}}{b_{\boldsymbol{n}}}\right)$$

The solution to this equation is given by Hallin et al. (2004b) as

$$\begin{pmatrix} \hat{a}_0 \\ \hat{\boldsymbol{a}}_1 b_{\boldsymbol{n}} \end{pmatrix} = {U_{\boldsymbol{n}}}^{-1} V_{\boldsymbol{n}}$$

Where $\boldsymbol{V_n} \equiv \begin{pmatrix} v_{\boldsymbol{n}0} \\ v_{\boldsymbol{n}1} \end{pmatrix}$ and $\boldsymbol{U_n} \equiv \begin{pmatrix} u_{\boldsymbol{n}00} & u_{\boldsymbol{n}01} \\ u_{\boldsymbol{n}10} & u_{\boldsymbol{n}11} \end{pmatrix}$ with $\left(\frac{X_j - x}{b_n}\right)_i = 1$

$$(\boldsymbol{V_n})_i \equiv \left(\hat{\boldsymbol{n}}{b_{\boldsymbol{n}}}^d\right)^{-1} \sum_{\boldsymbol{j}\in I_{\boldsymbol{n}}} Y_{\boldsymbol{j}} \left(\frac{\boldsymbol{X_j}-\boldsymbol{x}}{b_{\boldsymbol{n}}}\right)_i K\left(\frac{\boldsymbol{X_j}-\boldsymbol{x}}{b_{\boldsymbol{n}}}\right) \qquad i = 0, \dots, d.$$

$$(\boldsymbol{U_n})_{i\ell} \equiv \left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^d\right)^{-1} \sum_{\boldsymbol{j} \in I_{\boldsymbol{n}}} \left(\frac{\boldsymbol{X_j} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right)_i \left(\frac{\boldsymbol{X_j} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right)_\ell K\left(\frac{\boldsymbol{X_j} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right) \qquad i, \ell = 0, \dots, d$$

In addition,

$$\boldsymbol{H_n} \equiv \begin{pmatrix} \hat{a}_0 - a_0 \\ \widehat{\boldsymbol{a}}_1 b_{\boldsymbol{n}} - \boldsymbol{a}_1 b_{\boldsymbol{n}} \end{pmatrix} = \begin{pmatrix} g_{\boldsymbol{n}}(\boldsymbol{x}) - g(\boldsymbol{x}) \\ \left(g_{\boldsymbol{n}}'(\boldsymbol{x}) - g'(\boldsymbol{x})\right) b_{\boldsymbol{n}} \end{pmatrix} = \boldsymbol{U_n}^{-1}\left\{\boldsymbol{V_n} - \boldsymbol{U_n}\begin{pmatrix} a_0 \\ \boldsymbol{a}_1 b_{\boldsymbol{n}} \end{pmatrix}\right\} \equiv \boldsymbol{U_n}^{-1}\boldsymbol{W_n}$$

Where $\boldsymbol{W_n} \equiv \binom{W_{\boldsymbol{n}0}}{\boldsymbol{W_{n1}}}, \boldsymbol{W}_{\boldsymbol{n}i} \equiv \left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^d\right)^{-1} \sum_{\boldsymbol{j} \in I_{\boldsymbol{n}}} Z_{\boldsymbol{j}} \left(\frac{X_j - x}{b_n}\right)_i K\left(\frac{X_j - x}{b_n}\right), \quad i = 0, \dots, d.$

and $Z_{\boldsymbol{j}} \equiv Y_{\boldsymbol{j}} - a_0 - \boldsymbol{a}_1{}^T\left(\boldsymbol{X_j} - \boldsymbol{x}\right)$. Under certain conditions, the local linear kernel estimator is consistent and asymptotically normal. For convergence speed and the asymptotic distribution of the local linear kernel estimator derived by Hallin et al. (2004b), see Appendix 2.1.

2.4 Monte Carlo Study

In this section, we report results of a Monte Carlo study using nonparametric estimation methods: the k-Nearest Neighbor Estimator of Li and Tran (2009), the Nadraya-Watson Estimator of Lu and Chen (2002) and P.M. Robinson (2011), the Local Linear Kernel Estimator of Hallin et al. (2004b) and the maximum likelihood method with a linear approximation of the data generating process[26]. For simplicity, suppose a univariate X $(d = 1)$ is observed on a site in a two dimensional space $\boldsymbol{i} \equiv (i, j) \in \mathbb{Z}^2$ $(N = 2)$ as $X_{i,j}$. We will consider the following model:

Let $\{\xi_{i,j}, (i, j) \in \mathbb{Z}^2\}$ and $\{\varepsilon_{i,j}, (i, j) \in \mathbb{Z}^2\}$ be two mutually independent i.i.d. white noise processes with Normal distribution, and

[26] All the simulation results in this section are generated with C++ code following the C++0x standard run under the Gnome3.4.2 environment based on Kernel Linux 3.11.0-15-generic on a 64-bit machine with Intel Core i7-3770 CPU @3.40GHz×8 processor.

$$Y_{i,j} = g(X_{i,j}) + \xi_{i,j} \qquad \text{with} \qquad g(x) = \frac{2}{3}\sin x + \frac{1}{3}e^x$$

$\{X_{i,j}, (i,j) \in \mathbb{Z}^2\}$ is generated by a spatial autoregression,

$$X_{i,j} = \frac{1}{8}\left(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}\right) + \varepsilon_{i,j}$$

where $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_1^2)$ and $\xi_{i,j} \sim \mathcal{N}(0, \sigma_2^2)$. Here we choose the coefficients on all of $X_{i,j}'s$ neighbors equal to $\frac{1}{8}$, so that each of $X_{i,j}'s$ neighbors have an equal influence on $X_{i,j}$. Analogous to the time series case, this data generating process is stationary since the sum of the absolute values of the four coefficients in this "autoregressive" process is less than 1. The functional form is chosen so that it is close to linear when x is close to 0, which will give MLE, which assumes a linear data generating process, the best chance to perform well. The relative magnitude of σ_1^2 and σ_2^2 will determine the relative proportion of variation in the dependent variable resulting from variation in the covariate X and the white noise disturbance term ξ.

The data are generated iteratively by the following steps. We first draw from an i.i.d. random variable $\varepsilon_{i,j}$ for each site over a rectangular grid of $(m + 200) \times (n + 200)$, and initial values for $X_{i,j}$s are set to 0. Next, $X_{i,j}$ over $\{(i,j), i = 1, \dots, 200 + m, j = 1, \dots, 200 + n\}$ are generated using the spatial autoregressive model,

$$X_{i,j} = \frac{1}{8}\left(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}\right) + \varepsilon_{i,j},$$

recursively. This process is then iterated 100 times. The first 99 steps are used as warm-up steps to achieve stationary and the data generated in these steps are discarded. The results at the final iteration step for sites (i,j) over the domain $\{(i,j) | 101 \le i \le 100 + m, 101 \le j \le 100 + n\}$

are taken as the simulated $m \times n$ sample for $X_{i,j}$. The peripheral rows and columns are also discarded in order to mitigate “border effect” so that the final observations for $X_{i,j}$ receive enough feedback from neighboring sites in all directions.

With $(X_{i,j}, Y_{i,j})$ collected on the $m \times n$ rectangular grid, the three nonparametric methods are used to fit the data and the mean squared prediction errors are calculated for the whole $m \times n$ sample. To estimate the model by maximum likelihood method, we first need to convert both X and Y sampled on the m by n rectangular grid into a vector of dimension mn. We did column vectorization for the $m \times n$ matrix of the sampled observations of X and Y. For example,

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,n-1} & X_{1,n} \\ X_{2,1} & X_{2,2} & & X_{2,n-1} & X_{2,n} \\ \vdots & & \ddots & \vdots & \\ X_{m-1,1} & X_{m-1,2} & \cdots & X_{m-1,n-1} & X_{m-1,n} \\ X_{m,1} & X_{m,2} & & X_{m,n-1} & X_{m,n} \end{bmatrix}$$

$$Vec(X) = [X_{1,1} \quad X_{2,1} \quad \cdots \quad X_{m,1} \quad X_{1,2} \quad \cdots \quad X_{m,2} \quad \cdots \quad X_{1,n} \quad \cdots \quad X_{m,n}]^T$$

The vectorization for Y is done similarly. In order to run a regression $Vec(Y) = \rho W Vec(X) + Vec(\epsilon)$, we need to generate an $mn \times mn$ spatial weight matrix W characterizing the spatial dependence in the data generating process of X. Since X is sampled from a spatial autoregressive model $X_{i,j} = \frac{1}{8}(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}) + \varepsilon_{i,j}$, the four neighbors of $X_{i,j}$ have equal influence on $X_{i,j}$. We place equal weight on each unit $(i-1,j)$, $(i,j-1)$, $(i+1,j)$ and $(i,j+1)$ to characterize the effects of unit (i,j)’s neighbors on unit (i,j). We first construct an $mn \times mn$ spatial weight matrix W_0 where how unit (i,j)’s neighbors on the rectangular grid interact with unit (i,j) will be characterized by row $p = m(j-1) + i$. Specifically, the effect of unit $(i-1,j)$, $(i,j-1)$, $(i+1,j)$ and $(i,j+1)$ on unit (i,j) will be captured by the element in row

$p = m(j-1)+i$, column $m(j-1)+i-1$, $m(j-2)+i$, $m(j-1)+i+1$ and $mj+i$, respectively. We assign one to these elements in the pth row, and the rest of the elements are left as zero.[27] We then do a row normalization to W_0 to get the spatial weight matrix W so that the pth element of a vector resulting from left multiplying W to a vector variable is the average of the elements that are neighbors of the pth element in the vector multiplied[28]. With this spatial weight matrix W, the model $Vec(Y) = \rho W Vec(X) + Vec(\epsilon)$ is estimated by the maximum likelihood method and the RMSE is calculated.

We did Monte Carlo simulations for various pairs of σ_1^2 and σ_2^2 and two pairs of (m, n) ($(m = 20, n = 30)$ and $(m = 40, n = 50)$). With data sampled from an m by n rectangular grid for X as described in the previous procedure, we randomly drew realization $\xi_{i,j}$ from a normal distribution $\mathcal{N}(0,\ \sigma_2^2)$ for each point on the rectangular grid and add to $g(X_{i,j})$ to get $Y_{i,j}$. Then we used the three nonparametric methods and the maximum likelihood method to estimate $E[Y_{\boldsymbol{i}}|\boldsymbol{X_i} = \boldsymbol{x}]$ for X and Y sampled on this m by n rectangular grid and calculated the RMSE for each of the estimation methods[29].

Given that $E[Y_{i,j}|X_{i,j}] = E[(g(X_{i,j}) + \xi_{i,j})|X_{i,j}] = g(X_{i,j})$, if $\hat{g}_{\boldsymbol{n}}(X_{i,j})$ is a consistent estimator of $g(X_{i,j})$, then the prediction error would be $Y_{i,j} - \hat{E}[Y_{i,j}|X_{i,j}] = Y_{i,j} - \hat{g}_{\boldsymbol{n}}(X_{i,j}) \xrightarrow{p} Y_{i,j} - g(X_{i,j}) = e_{i,j}$, where $e_{i,j}$ is a sample from $\xi_{i,j}$, which has a Normal distribution $\mathcal{N}$(0,

---

[27] The neighbors on the borders of the matrix is a little bit different. For elements on the four corners of the X matrix, they have only two neighbors. Elements that are on the border but not on the corners have three neighbors. For these elements, we also assign equal weights to their neighbors.

[28] The row normalization is done by left multiplying a diagonal matrix. The diagonal matrix is the inverse of the diagonal matrix whose kth diagonal element is the sum of elements in the kth row of matrix W_0.

[29] For the nonparametric local linear estimator and the Nadaraya-Watson estimator, the normal kernel is used. As to the bandwidth choice, since there is no theoretical work on the bandwidth choice for nonparametric regression with spatial data, we chose the optimal bandwidth by minimizing the mean squared error for a given set of samples on the $m \times n$ rectangular grid. For the k-Nearest Neighbors method, k is chosen to be 19.

σ_2^2). By the law of large numbers, the mean squared error (MSE) $\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \left(Y_{i,j} - \hat{E}\left[Y_{i,j}\middle|X_{i,j}\right]\right)^2 \xrightarrow{p} \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} {e_{i,j}}^2 \xrightarrow{p} \sigma_2^2$. This process is repeated for 499 times with X fixed and Y varying as a result of independent draws from $\xi_{i,j}$. In the end, we calculated the mean and standard error of the 499 calculated RMSE. The mean can be seen as an estimate of σ_2. For each pair of σ_1^2 and σ_2^2, the above simulation process is repeated 8 times. Within each of these 8 sets of simulations, X is kept fixed while Y varies as a result of the independent random draws of $\xi_{i,j}$.

To do the simulation, we first fixed (m, n) at $(20, 30)$ and chose σ_1, σ_2 from 1, 0.5, 0.1. The results for cases when σ_1 and σ_2 are equal are reported in Table 2.1. These results show that the nonparametric local linear kernel method has the best in sample prediction performance and the mean value of the RMSE for the 499 repeated simulations is very close to the true value for all cases. The performance of the Nadaraya-Watson method is also good with the mean of RMSEs close to the true value. For the k-nearest neighbors method, the mean value of the RMSEs is about one standard deviation away from the desired true value, so we cannot reject the hypothesis that the mean of the RMSEs is statistically different from the true value with confidence level of 90% or 95%. However, the performance of maximum likelihood method is not acceptable since the mean value of the RMSEs diverges from the true value significantly. This is reasonable since we have assumed a linear relationship between X and Y in the maximum likelihood estimation method while the true data generating process is nonlinear.

When σ_1 and σ_2 are not equal, the relative performances of these estimation methods still hold: the best performer is the nonparametric local linear kernel estimator, the Nadaraya-Watson estimator also performs well and the k-nearest neighbors estimator is in general not too far off the mark. However, the maximum likelihood estimator does not perform as well as the nonparametric methods. These results are reported in Table 2.2 and Table 2.3. Another observation from these results is that the performance of these four estimation methods seems to be related to the ratio of σ_1 and σ_2, ${}^{\sigma_1}/_{\sigma_2}$; All four estimation methods perform better when this ratio is relatively small. When $\sigma_1 = 0.1$ and $\sigma_2 = 1$, even the mean value of the RMSEs of the maximum likelihood method approaches the true value, and the maximum likelihood method also outperforms the k-nearest neighbors method. However it does not perform as well as the nonparametric local linear kernel method and the Nadaraya-Watson estimator.

The asymptotic theory considered here is of the "increase domain" variety, where the asymptotic properties of an estimator are derived by increasing the sample size while keeping the distance between two closest neighboring observations fixed. To check whether the relative performance of these different estimators varies with the sample size, we increased the sample size to $m = 40$ and $n = 50$ and run the same regressions of Y on X sampled on the 40 by 50 rectangular grid. The mean of the RMSEs are in general closer to the true value compared to the corresponding case with a smaller sample size and the relative performances of these different estimation methods are preserved. The simulation results are reported in Table 2.4 through Table 2.6.

2.5 An Application to Predicting Crop Yield with Precipitation

In this section, we use these estimation methods to predict agricultural crop yield with monthly average precipitation. With synthesized climate data and corn and soybean yields at US county level for the period 1982-98, Lobell and Asner (2003) found evidence of a strong but spatially explicit coupling between inter-annual climate and crop yield anomalies. From a biological perspective, at least a certain amount of water is needed for any plants to survive.

While a regular rainfall pattern is of crucial importance to healthy crops, too much or too little can be both harmful. Drought can kill crops and increase erosion, as can be seen from the impact of the U.S. drought in 2012 on crop yields. For example U.S. corn growers produced 10.8 billion bushels in 2012, 13 percent below the 2011 level[30]. At the same time, excess precipitation can be harmful to crop growth, since excess soil moisture or continued soil saturation prevents healthy root growth. Rosenzweig et. al. (2002) have simulated US corn production losses due to excess precipitation under climate change. Based on these facts, we expect that precipitation can be used to predict crop yield and that the relationship is probably not linear[31].

We will use monthly average precipitation during the growing season to forecast crop yields at the county level from 1990 to 2012 in four agriculture belts (areas) in the United States[32]. The four agriculture belts are the Barley Belt, the Corn Belt, the Soybean Belt and the Wheat Belt[33]. The Barley Belt consists of Colorado, Idaho, Minnesota, Montana, North Dakota, South Dakota, Washington and Wyoming. The Corn Belt consists of Illinois, Indiana, Iowa, Kansas, Minnesota, Nebraska, Ohio and South Dakota. The Soybean Belt consists of Illinois, Indiana, Iowa, Minnesota, Missouri Nebraska and Ohio. The Wheat Belt consists of Idaho, Minnesota, Montana, North Dakota, South Dakota and Washington. The growing season of these four crops are selected based on information about the usual planting and harvesting dates from U.S. field crops, from United States Department of Agriculture report[34]. The growing seasons for barley,

---

[30] Detailed facts can be found here: http://www.ers.usda.gov/topics/in-the-news/us-drought-2012-farm-and-food-impacts.aspx and here: http://www.nass.usda.gov/Newsroom/2013/01_11_2013.asp

[31] The advantage of using nonparametric methods, rather than simply run a least squares regression, lies in the fact that the relationship between crop yield and precipitation is not linear and we don't need to explicitly assume a functional form in nonparametric estimation methods.

[32] Due to data disclosure restrictions, the study period for wheat is from 1990 to 2008.

[33] In some cases, Grain Belt is also referred to as Wheat Belt, meaning northern Midwestern states where most of North America's grain and soybeans are grown. We choose those contiguous states with the highest barley productions as the Barley Belt.

[34] http://swat.tamu.edu/media/90113/crops-typicalplanting-harvestingdates-by-states.pdf

corn, soybean and wheat are April to August, April to September, May to September and November to July next year, respectively[35].

The crop yield data are from the United States Department of Agriculture's National Agricultural Statistical Service[36], which provides data on agricultural production, area harvested, yield and sales at the county level. We queried available yearly crop yields for all the counties of those states located in the four corresponding agriculture belts. The unit for yield data is bushel per acre. The precipitation data are from the National Oceanic and Atmospheric Administration's National Climate Data Center[37]. The National Climate Data Center provide climate data based on observations at climate stations across the country. We extracted all the available monthly precipitation data for all the stations located in states belonging to the four agriculture belts. With information on the latitude and longitude of the stations, we could identify to which state and county each of these stations belongs to. Then we average precipitation levels observed at all the stations located in a specific county in a given month and use it as the precipitation level for that county in that month. The location of this precipitation observation is taken to be the average of the latitudes and longitudes of all those stations in that county. With monthly precipitation data for each county, we take the average of monthly precipitation levels during the growing season of a crop and use it as the precipitation level for that growing season. Lastly, the precipitation data is merged with the yield data by state, county and year of observation.

With data for county level crop yield and precipitation, we fit the data with the three nonparametric methods and the maximum likelihood method. Since the spatial correlation more

---

[35] Winter wheat grows from October to June, comprising 65% of all wheat produced. Spring wheat grows from April to July. These two growing seasons are combined and the monthly average precipitation from October to July is calculated for wheat.

[36] http://www.nass.usda.gov/Data_and_Statistics/index.asp

[37] http://www.ncdc.noaa.gov/cdo-web/

likely exists in the precipitation variable, the simple spatial regression $Y = \rho WX + \epsilon$ is used to fit the crop yield and precipitation data. The spatial weight matrix is constructed as the inverse distance-measure. With latitude and longitude information for two precipitation observations, we assume that their correlation is inversely related to the distance between them and only contemporary correlation exists[38]. For precipitation observed at two points with (latitude, longitude) pairs, (ϕ_i, λ_i) and (ϕ_j, λ_j), suppose $\Delta\phi = \phi_j - \phi_i$, $\Delta\lambda = \lambda_j - \lambda_i$ and the mean latitude is labeled as $\phi_m = (\phi_j + \phi_i)/2$, the geographic distance between them is[39]:

$$dist_{ij} = R\sqrt{(\Delta\phi)^2 + (\cos(\phi_m)\Delta\lambda)^2}$$

After calculating the distance $dist_{ij}$ capturing the correlation between precipitation in county i and county j in a given year, we assign the corresponding element in the spatial weight matrix to be $\frac{1}{dist_{ij}}$. Then we do a row normalization to this matrix so that the weight matrix has the effect of taking the average of the precipitation levels in the neighboring counties.

Before coming to our main results for the prediction performance of the various estimation methods for the crop yield data, let's first take a look at the crop yield data and the precipitation data. Figure 2.1 to Figure 2.4 depict the average yields in each state over time for each of the four crops. There seems to be an increasing trend in yield for barley, corn and soybean. But the yield for wheat seems to be flat over time, or at least no trend is obvious. There was a contemporaneous increase in yield of barley, corn and soybean for all the states in 2004. The sharp decrease in yield of corn and soybean seen in most of the states in 2012 reflects the

---

[38] We assume the dependence in precipitation levels in two counties is inversely related to the geographic distance between them. The geographic distance is calculated with information on the latitudes and longitudes of climate stations in the two counties.

[39] R is the radius of the earth, which equals 6,371.009 kilometers. All latitude and longitude are denominated in unit of radians.

drought that hit America's Midwest. In Figure 2.5, we mapped the average yield over time of the four crops for each county. From the map, colors indicating high yield counties and low yield counties are geographically clustered. In Figure 2.6, we mapped the average precipitation in selected months (March, June, September, December) over time for these counties. It is obvious that precipitation is spatially correlated. Large areas of the same colors in contiguous counties mean that a county with high precipitation level is more likely to be surrounded by counties with high precipitation levels.

The estimation results for the crop yield data are reported in Table 2.7. Consistent with our simulation results, the nonparametric local linear kernel method has the best predictive performance since it has the lowest RMSE for all the four crops. The Nadaraya-Watson estimator performs well with the second lowest RMSE for the four crops. However, what's different from the simulation results is that with an inverse distance weighting, the maximum likelihood method is performing better than the K-Nearest Neighbors method, which deserves further exploration.

2.6 Conclusions

In this paper, we have compared the prediction performance of the nonparametric local linear kernel methods, the Nadaraya-Watson estimator, the k-Nearest Neighbors method, and the maximum likelihood method applied to spatially dependent data. The Monte Carlo results show that nonparametric local linear kernel methods have the best prediction performance. The Nadaraya-Watson estimation method also performs well. In general, these two nonparametric methods outperform the k-Nearest Neighbors method and the maximum likelihood method regardless of the data generating process and sample size. In most of the cases, the mean of the

RMSEs for the Nadaraya-Watson estimation method is within one standard deviation of the true value, which means that we cannot reject the hypothesis that the RMSE is significantly different from the true value. However, this depends on the data generating process and in some cases the mean of the RMSEs is significantly different from the true value. In general, the maximum likelihood does not perform well because the spatial weight matrix can only estimate linear structure while the true data generating process is nonlinear. This also gives some support to the idea of using nonparametric methods when various misspecification may exist either in the functional form or spatial weight matrix.

We have also used these methods to predict crop yield with precipitation at the county level. The results are in general consistent with the simulation results. The nonparametric local linear kernel method has the best predictive performance. The Nadaraya-Watson estimation method also performs better than the other two estimation methods. However, with an inverse distance weighting matrix, the maximum likelihood estimator outperforms the k-Nearest Neighbors method in predicting crop yield.

2.7 References

Anselin, Luc, and Raymond JGM Florax. "New directions in spatial econometrics." (1995).

Baltagi, Badi H., and Dong Li. "Prediction in the panel data model with spatial correlation: the case of liquor." Spatial Economic Analysis 1, no. 2 (2006): 175-185.

Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen. "Inference with dependent data using cluster covariance estimators." Journal of Econometrics 165, no. 2 (2011): 137-151.

Case, Anne C. "Spatial patterns in household demand." Econometrica: Journal of the Econometric Society (1991): 953-965.

Conley, Timothy G. "GMM estimation with cross sectional dependence." Journal of econometrics 92, no. 1 (1999): 1-45.

Conley, Timothy G., and Francesca Molinari. "Spatial correlation robust inference with errors in location or distance." Journal of Econometrics 140, no. 1 (2007): 76-96.

Cliff, Andrew David, and J. Keith Ord. Spatial processes: models & applications. Vol. 44. London: Pion, 1981.

De Wit, AJW D., H. L. Boogaard, and C. A. Van Diepen. "Spatial resolution of precipitation and radiation: the effect on regional crop yield forecasts." Agricultural and Forest Meteorology 135, no. 1 (2005): 156-168.

Druska, Viliam, and William C. Horrace. "Generalized moments estimation for spatial panel data: Indonesian rice farming." American Journal of Agricultural Economics 86, no. 1 (2004): 185-198.

Fan, Jianqing. "Local linear regression smoothers and their minimax efficiencies." The Annals of Statistics (1993): 196-216.

Jianqing, Fan, and Gijbels Irene. "Local polynomial modelling and its applications." (1996).

Fan, Jianqing, and Qiwei Yao. Nonlinear time series. Vol. 2, no. 3. Springer, 2002.

Gheriballah, Abdelkader, Ali Laksaci, and Rachida Rouane. "Robust nonparametric estimation for spatial regression." Journal of Statistical Planning and Inference 140, no. 7 (2010): 1656-1670.

Hallin, Marc, Zudi Lu, and Lanh Tat Tran. "Density estimation for spatial linear processes." Bernoulli (2001): 657-668.

Hallin, Marc, Zudi Lu, and Lanh T. Tran. "Kernel density estimation for spatial processes: the L_1 theory." Journal of Multivariate Analysis 88, no. 1 (2004a): 61-75.

Hallin, Marc, Zudi Lu, and Lanh T. Tran. "Local linear spatial regression." The Annals of Statistics 32, no. 6 (2004): 2469-2500.

Hansen, Bruce E. "Nonparametric conditional density estimation." Unpublished manuscript (2004).

Jenish, Nazgul. "Nonparametric spatial regression under near-epoch dependence." Journal of Econometrics 167, no. 1 (2012): 224-239.

Lee, Lung−Fei. "Asymptotic Distributions of Quasi−Maximum Likelihood Estimators for Spatial Autoregressive Models." Econometrica 72, no. 6 (2004): 1899-1925.

Lee, Lung-fei. "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models." Journal of Econometrics 137, no. 2 (2007): 489-514.

Li, Jiexiang, and Lanh Tat Tran. "Nonparametric estimation of conditional expectation." Journal of Statistical Planning and Inference 139, no. 2 (2009): 164-175.

Lu, Zudi, and Xing Chen. "Spatial kernel regression estimation: weak consistency." Statistics & probability letters 68, no. 2 (2004): 125-136.

Chen, Xing. "Spatial nonparametric regression estimation: non-isotropic case." Acta Mathematicae Applicatae Sinica 18, no. 4 (2002): 641-656.

Lu, Zudi, and Oliver Linton. "Local linear fitting under near epoch dependence." Econometric Theory 23, no. 01 (2007): 37-70.

Manski, Charles F. "Identification of endogenous social effects: The reflection problem." The review of economic studies 60, no. 3 (1993): 531-542.

McMillen, Daniel P. "Perspectives on spatial econometrics: Linear smoothing with structured models." Journal of Regional Science 52, no. 2 (2012): 192-209.

Pinkse, Joris, and Margaret E. Slade. "The future of spatial econometrics." Journal of Regional Science 50, no. 1 (2010): 103-117.

Robinson, Peter M. "Asymptotic theory for nonparametric regression with spatial data." Journal of Econometrics 165, no. 1 (2011): 5-19.

Rosenzweig, Cynthia, Francesco N. Tubiello, Richard Goldberg, Evan Mills, and Janine Bloomfield. "Increased crop damage in the US from excess precipitation under climate change." Global Environmental Change 12, no. 3 (2002): 197-202.

United States Department of Agriculture (USDA). “Usual planting and harvesting dates for U.S. field crops”. December 1997.

Table 2.1 Root Mean Squared Prediction Error for Spatially Dependent Data Sampled on a 20 by 30 Grid on a Random Field

Variance	Estimation Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sigma_1 = 1$, $\sigma_2 = 1$	Local Linear	1.00510 (0.02789)	1.00589 (0.02877)	1.00434 (0.03033)	1.00865 (0.02986)	1.02687 (0.02728)	1.00512 (0.02983)	1.00523 (0.02888)	1.00941 (0.02842)
	Nadaraya-Watson	1.00748 (0.02820)	1.01180 (0.02921)	1.00821 (0.03056)	1.08904 (0.02813)	1.04439 (0.02840)	1.00825 (0.03003)	1.00666 (0.02888)	1.01575 (0.02832)
	K-Nearest Neighbors	1.03123 (0.02956)	1.03094 (0.03017)	1.03199 (0.03218)	1.03430 (0.03150)	1.08541 (0.03002)	1.03309 (0.03140)	1.03525 (0.03096)	1.05117 (0.03032)
	Maximum Likelihood	1.39261 (0.03465)	1.36347 (0.03546)	1.44790 (0.03548)	1.46420 (0.03597)	1.47076 (0.03485)	1.40586 (0.03447)	1.45096 (0.03539)	1.47221 (0.03624)
$\sigma_1 = 0.5$, $\sigma_2 = 0.5$	Local Linear	0.50290 (0.01361)	0.50236 (0.01516)	0.50191 (0.01437)	0.50171 (0.01464)	0.50239 (0.01489)	0.50125 (0.01477)	0.50023 (0.01476)	0.50194 (0.01382)
	Nadaraya-Watson	0.50482 (0.01374)	0.50607 (0.01530)	0.50573 (0.01455)	0.50540 (0.01500)	0.50895 (0.01480)	0.50309 (0.01475)	0.50250 (0.01486)	0.50409 (0.01395)
	K-Nearest Neighbors	0.51476 (0.01470)	0.51471 (0.01625)	0.51457 (0.01514)	0.51399 (0.01578)	0.51507 (0.01556)	0.51316 (0.01545)	0.51295 (0.01569)	0.51466 (0.01458)
	Maximum Likelihood	0.70837 (0.01656)	0.68715 (0.01747)	0.70566 (0.01706)	0.71371 (0.01669)	0.71502 (0.01820)	0.70088 (0.01726)	0.68109 (0.01833)	0.70439 (0.01653)
$\sigma_1 = 0.1$, $\sigma_2 = 0.1$	Local Linear	0.10011 (0.00284)	0.10016 (0.00295)	0.10012 (0.00280)	0.10004 (0.00294)	0.10016 (0.00282)	0.10021 (0.00292)	0.10017 (0.00307)	0.10000 (0.00284)
	Nadaraya-Watson	0.10108 (0.00288)	0.10120 (0.00294)	0.10104 (0.00280)	0.10121 (0.00293)	0.10106 (0.00285)	0.10117 (0.00297)	0.10110 (0.00308)	0.10090 (0.00289)
	K-Nearest Neighbors	0.10281 (0.00307)	0.10302 (0.00314)	0.10275 (0.00298)	0.10307 (0.00306)	0.10291 (0.00305)	0.10305 (0.00311)	0.10293 (0.00329)	0.10270 (0.00304)
	Maximum Likelihood	0.14553 (0.00346)	0.14387 (0.00364)	0.14304 (0.00348)	0.14391 (0.00373)	0.14440 (0.00332)	0.14315 (0.00367)	0.14596 (0.00360)	0.14355 (0.00364)

Table 2.2 Root Mean Squared Prediction Error for Spatially Dependent Data Sampled on a 20 by 30 Grid on a Random Field

Variance	Estimation Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sigma_1 = 1$, $\sigma_2 = 0.1$	Local Linear	0.10246 (0.00310)	0.13142 (0.00244)	0.11169 (0.00715)	0.10248 (0.00295)	0.15131 (0.00251)	0.15363 (0.00833)	0.10253 (0.00295)	0.10270 (0.00304)
	Nadaraya-Watson	0.11130 (0.00421)	0.12478 (0.00273)	0.19078 (0.00463)	0.10659 (0.00317)	0.13905 (0.00242)	0.26182 (0.00468)	0.10517 (0.00315)	0.10399 (0.00309)
	K-Nearest Neighbors	0.29970 (0.00377)	0.12782 (0.00311)	0.40619 (0.00381)	0.23248 (0.00359)	0.19370 (0.00357)	0.46097 (0.00392)	0.17712 (0.00316)	0.13671 (0.00324)
	Maximum Likelihood	1.12766 (0.00434)	0.98198 (0.00422)	1.18992 (0.00416)	1.15977 (0.00389)	1.09845 (0.00391)	1.17221 (0.00402)	1.03750 (0.00403)	1.07417 (0.00402)
$\sigma_1 = 1$, $\sigma_2 = 0.5$	Local Linear	0.50945 (0.01547)	0.50580 (0.01428)	0.51238 (0.01656)	0.51044 (0.01633)	0.50437 (0.01438)	0.50566 (0.01472)	0.50415 (0.01396)	0.52491 (0.02168)
	Nadaraya-Watson	0.51455 (0.01461)	0.50963 (0.01385)	0.53063 (0.01513)	0.52895 (0.01576)	0.51082 (0.01430)	0.51094 (0.01483)	0.50965 (0.01392)	0.56648 (0.01692)
	K-Nearest Neighbors	0.56862 (0.01576)	0.53317 (0.01494)	0.57705 (0.01636)	0.58325 (0.01626)	0.58603 (0.01508)	0.52866 (0.01554)	0.55573 (0.01542)	0.65838 (0.01672)
	Maximum Likelihood	1.15187 (0.01913)	1.15449 (0.01911)	1.23932 (0.01955)	1.16858 (0.01913)	1.22399 (0.01951)	1.07753 (0.02002)	1.21286 (0.02021)	1.22401 (0.01935)
$\sigma_1 = 0.5$, $\sigma_2 = 0.1$	Local Linear	0.10178 (0.00299)	0.10073 (0.00274)	0.10072 (0.00298)	0.10340 (0.00428)	0.10071 (0.00284)	0.10103 (0.00295)	0.10108 (0.00309)	0.10174 (0.00332)
	Nadaraya-Watson	0.10768 (0.00307)	0.10181 (0.00278)	0.10201 (0.00305)	0.14227 (0.00265)	0.10309 (0.00291)	0.10372 (0.00308)	0.10179 (0.00307)	0.13473 (0.00231)
	K-Nearest Neighbors	0.11916 (0.00321)	0.10855 (0.00293)	0.10925 (0.00300)	0.13398 (0.00350)	0.11276 (0.00316)	0.11284 (0.00322)	0.10815 (0.00326)	0.10854 (0.00316)
	Maximum Likelihood	0.53908 (0.00417)	0.50510 (0.00405)	0.52185 (0.00400)	0.52053 (0.00400)	0.52277 (0.00408)	0.53351 (0.00409)	0.50132 (0.00399)	0.51686 (0.00396)

Table 2.3 Root Mean Squared Prediction Error for Spatially Dependent Data Sampled on a 20 by 30 Grid on a Random Field

Variance	Estimation Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sigma_1 = 0.1$, $\sigma_2 = 1$	Local Linear	1.00030 (0.02909)	1.00015 (0.02944)	1.00145 (0.02857)	1.00238 (0.02862)	1.00064 (0.03133)	1.00021 (0.02910)	1.00190 (0.02980)	1.00145 (0.02999)
	Nadaraya-Watson	1.00099 (0.02906)	1.00108 (0.02943)	1.00239 (0.02857)	1.00310 (0.02872)	1.00133 (0.03142)	1.00097 (0.02910)	1.00271 (0.02979)	1.00240 (0.02995)
	K-Nearest Neighbors	1.02484 (0.03084)	1.02450 (0.03136)	1.02639 (0.03032)	1.02738 (0.03027)	1.02512 (0.03317)	1.02532 (0.03094)	1.02674 (0.03218)	1.02613 (0.03135)
	Maximum Likelihood	1.00267 (0.02905)	1.00292 (0.02942)	1.00466 (0.02875)	1.00520 (0.02883)	1.00338 (0.03125)	1.00264 (0.02896)	1.00444 (0.02964)	1.00375 (0.03016)
$\sigma_1 = 0.5$, $\sigma_2 = 1$	Local Linear	1.00084 (0.02604)	1.00241 (0.02966)	1.00067 (0.02830)	1.00220 (0.02804)	1.00101 (0.02986)	1.00204 (0.03063)	0.99995 (0.02797)	1.00102 (0.02844)
	Nadaraya-Watson	1.00488 (0.02608)	1.00716 (0.03023)	1.00433 (0.02836)	1.00597 (0.02805)	1.00513 (0.02992)	1.00651 (0.03118)	1.00395 (0.02804)	1.00519 (0.02868)
	K-Nearest Neighbors	1.02560 (0.02805)	1.02788 (0.03115)	1.02541 (0.02978)	1.02614 (0.02921)	1.02608 (0.03176)	1.02675 (0.03194)	1.02441 (0.02974)	1.02644 (0.03107)
	Maximum Likelihood	1.12733 (0.01477)	1.11791 (0.01446)	1.11026 (0.01500)	1.11936 (0.01445)	1.12564 (0.01456)	1.12437 (0.01512)	1.11989 (0.01535)	1.11329 (0.01457)
$\sigma_1 = 0.1$, $\sigma_2 = 0.5$	Local Linear	0.49993 (0.01477)	0.50119 (0.01446)	0.49950 (0.01500)	0.50079 (0.01445)	0.50243 (0.01456)	0.49972 (0.01512)	0.50016 (0.01535)	0.49867 (0.01457)
	Nadaraya-Watson	0.50108 (0.01464)	0.50237 (0.01421)	0.50068 (0.01469)	0.50172 (0.01428)	0.50361 (0.01430)	0.50088 (0.01498)	0.50152 (0.01497)	0.50025 (0.01444)
	K-Nearest Neighbors	0.51244 (0.01518)	0.51360 (0.01533)	0.51184 (0.01533)	0.51295 (0.01519)	0.51511 (0.01491)	0.51231 (0.01555)	0.51270 (0.01559)	0.51112 (0.01533)
	Maximum Likelihood	0.50830 (0.01477)	0.50948 (0.01446)	0.50835 (0.01500)	0.50899 (0.01445)	0.51036 (0.01456)	0.50885 (0.01512)	0.50892 (0.01535)	0.50892 (0.01457)

Table 2.4 Root Mean Squared Prediction Error for Spatially Dependent Data Sampled on a 40 by 50 Grid on a Random Field

Variance	Estimation Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sigma_1 = 1$, $\sigma_2 = 1$	Local Linear	1.00669 (0.01504)	1.00319 (0.01637)	1.00289 (0.01599)	1.00574 (0.01578)	1.00890 (0.01697)	1.00180 (0.01504)	1.00363 (0.01563)	1.00343 (0.01609)
	Nadaraya-Watson	1.01874 (0.01614)	1.00458 (0.01634)	1.00637 (0.01581)	1.01252 (0.01591)	1.02012 (0.01632)	1.00311 (0.01508)	1.00521 (0.01556)	1.00640 (0.01588)
	K-Nearest Neighbors	1.07396 (0.01636)	1.02973 (0.01709)	1.03097 (0.01637)	1.04038 (0.01717)	1.06029 (0.01733)	1.02841 (0.01611)	1.02929 (0.01623)	1.03282 (0.01669)
	Maximum Likelihood	1.49158 (0.01922)	1.45535 (0.01960)	1.46460 (0.01959)	1.44173 (0.02058)	1.50104 (0.01906)	1.44283 (0.01856)	1.43527 (0.02014)	1.45929 (0.02028)
$\sigma_1 = 0.5$, $\sigma_2 = 0.5$	Local Linear	0.50112 (0.00775)	0.50056 (0.00806)	0.50048 (0.00776)	0.50042 (0.00773)	0.50051 (0.00784)	0.50077 (0.00841)	0.50051 (0.00794)	0.50100 (0.00805)
	Nadaraya-Watson	0.50858 (0.00767)	0.50151 (0.00807)	0.50150 (0.00781)	0.50146 (0.00783)	0.50152 (0.00785)	0.50193 (0.00848)	0.50178 (0.00797)	0.50246 (0.00813)
	K-Nearest Neighbors	0.51395 (0.00821)	0.51298 (0.00856)	0.51299 (0.00815)	0.51281 (0.00827)	0.51294 (0.00843)	0.51317 (0.00874)	0.51300 (0.00859)	0.51403 (0.00853)
	Maximum Likelihood	0.71078 (0.00957)	0.70603 (0.00966)	0.71605 (0.00955)	0.70899 (0.00950)	0.70649 (0.00979)	0.69879 (0.01021)	0.72222 (0.00983)	0.70917 (0.00973)
$\sigma_1 = 0.1$, $\sigma_2 = 0.1$	Local Linear	0.10000 (0.00162)	0.10017 (0.00154)	0.10002 (0.00159)	0.10004 (0.00152)	0.10006 (0.00150)	0.10006 (0.00154)	0.09995 (0.00165)	0.10006 (0.00161)
	Nadaraya-Watson	0.10093 (0.00162)	0.10056 (0.00155)	0.10045 (0.00159)	0.10047 (0.00154)	0.10095 (0.00149)	0.10047 (0.00157)	0.10033 (0.00165)	0.10048 (0.00164)
	K-Nearest Neighbors	0.10263 (0.00172)	0.10279 (0.00164)	0.10268 (0.00168)	0.10266 (0.00159)	0.10268 (0.00162)	0.10268 (0.00167)	0.10252 (0.00177)	0.10265 (0.00168)
	Maximum Likelihood	0.14185 (0.00201)	0.14346 (0.00199)	0.14404 (0.00196)	0.14260 (0.00194)	0.14368 (0.00192)	0.14267 (0.00192)	0.14407 (0.00194)	0.14514 (0.00205)

Table 2.5 Root Mean Squared Prediction Error for Spatially Dependent Data Sampled on a 40 by 50 Grid on a Random Field

Variance	Estimation Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sigma_1 = 1$, $\sigma_2 = 0.1$	Local Linear	0.150418 (0.00130)	0.100936 (0.00169)	0.139038 (0.00130)	0.100869 (0.00156)	0.102351 (0.00183)	0.103018 (0.00177)	0.117797 (0.00499)	0.102415 (0.00173)
	Nadaraya-Watson	0.131632 (0.00138)	0.103748 (0.00169)	0.126817 (0.00130)	0.102605 (0.00158)	0.131465 (0.00139)	0.135981 (0.00156)	0.208155 (0.00214)	0.1292 (0.00149)
	K-Nearest Neighbors	0.143182 (0.00183)	0.144898 (0.00179)	0.130574 (0.00179)	0.158741 (0.00197)	0.179604 (0.00188)	0.147421 (0.00193)	0.259991 (0.00214)	0.247754 (0.00204)
	Maximum Likelihood	1.12908 (0.00224)	1.05525 (0.00229)	1.03298 (0.00228)	1.0718 (0.00226)	1.08461 (0.00221)	1.01978 (0.00221)	1.15415 (0.00224)	1.13594 (0.00215)
$\sigma_1 = 1$, $\sigma_2 = 0.5$	Local Linear	0.502569 (0.00801)	0.502324 (0.00811)	0.503575 (0.00812)	0.503047 (0.00866)	0.505016 (0.00892)	0.501728 (0.00812)	0.623685 (0.01168)	0.502869 (0.00787)
	Nadaraya-Watson	0.506952 (0.00810)	0.506004 (0.00822)	0.508379 (0.00830)	0.508128 (0.00816)	0.517339 (0.00851)	0.50386 (0.00816)	0.68265 (0.01175)	0.511577 (0.00800)
	K-Nearest Neighbors	0.536086 (0.00836)	0.524006 (0.00863)	0.549623 (0.00907)	0.525722 (0.00887)	0.542368 (0.00908)	0.542567 (0.00871)	0.75105 (0.01046)	0.543438 (0.00830)
	Maximum Likelihood	1.24579 (0.01054)	1.15425 (0.01056)	1.2238 (0.01111)	1.2053 (0.01128)	1.20254 (0.01026)	1.26506 (0.00990)	1.29239 (0.01103)	1.20996 (0.01071)
$\sigma_1 = 0.5$, $\sigma_2 = 0.1$	Local Linear	0.100264 (0.00152)	0.100524 (0.00164)	0.100329 (0.00165)	0.100451 (0.00164)	0.100381 (0.00161)	0.100265 (0.00160)	0.100412 (0.00165)	0.100418 (0.00162)
	Nadaraya-Watson	0.10075 (0.00151)	0.102406 (0.00166)	0.100736 (0.00167)	0.101075 (0.00166)	0.101319 (0.00163)	0.100676 (0.00162)	0.100795 (0.00167)	0.101043 (0.00170)
	K-Nearest Neighbors	0.103097 (0.00159)	0.107184 (0.00177)	0.103962 (0.00173)	0.10657 (0.00174)	0.105608 (0.00176)	0.102942 (0.00167)	0.106898 (0.00181)	0.105682 (0.00174)
	Maximum Likelihood	0.518021 (0.00203)	0.514192 (0.00222)	0.512713 (0.00225)	0.517591 (0.00227)	0.50912 (0.00203)	0.507655 (0.00224)	0.519842 (0.00219)	0.526547 (0.00226)

Table 2.6 Root Mean Squared Prediction Error for Spatially Dependent Data Sampled on a 40 by 50 Grid on a Random Field

Variance	Estimation Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sigma_1 = 0.1$, $\sigma_2 = 1$	Local Linear	1.00064 (0.01589)	0.99922 (0.01565)	1.00029 (0.01620)	1.00115 (0.01592)	1.00081 (0.01666)	0.99948 (0.01644)	1.00043 (0.01604)	1.00090 (0.01523)
	Nadraya-Watson	1.00131 (0.01589)	0.99993 (0.01568)	1.00101 (0.01619)	1.00186 (0.01595)	1.00152 (0.01664)	1.00021 (0.01648)	1.00117 (0.01607)	1.00159 (0.01520)
	K-Nearest Neighbors	1.02626 (0.01683)	1.02542 (0.01661)	1.02566 (0.01717)	1.02700 (0.01693)	1.02624 (0.01766)	1.02541 (0.01736)	1.02598 (0.01694)	1.02652 (0.01607)
	Maximum Likelihood	1.00505 (0.01581)	1.00384 (0.01575)	1.00491 (0.01605)	1.00573 (0.01607)	1.00555 (0.01663)	1.00397 (0.01662)	1.00502 (0.01611)	1.00517 (0.01518)
$\sigma_1 = 0.5$, $\sigma_2 = 1$	Local Linear	1.00166 (0.01607)	1.00108 (0.01433)	1.00203 (0.01547)	1.00092 (0.01573)	1.00051 (0.01629)	1.00243 (0.01619)	1.00014 (0.01556)	1.00119 (0.01563)
	Nadraya-Watson	1.00436 (0.01620)	1.00351 (0.01438)	1.00451 (0.01550)	1.00219 (0.01578)	1.00295 (0.01642)	1.00492 (0.01632)	1.00385 (0.01553)	1.00385 (0.01573)
	K-Nearest Neighbors	1.02676 (0.01747)	1.02635 (0.01547)	1.02718 (0.01673)	1.02600 (0.01721)	1.02573 (0.01754)	1.02739 (0.01707)	1.02540 (0.01667)	1.02622 (0.01686)
	Maximum Likelihood	1.12482 (0.01759)	1.11431 (0.01634)	1.11696 (0.01605)	1.11936 (0.01719)	1.11953 (0.01716)	1.11846 (0.01776)	1.12227 (0.01665)	1.11997 (0.01768)
$\sigma_1 = 0.1$, $\sigma_2 = 0.5$	Local Linear	0.50020 (0.00738)	0.50047 (0.00779)	0.49975 (0.00788)	0.49984 (0.00786)	0.49985 (0.00740)	0.50023 (0.00829)	0.50057 (0.00809)	0.50009 (0.00791)
	Nadraya-Watson	0.50069 (0.00739)	0.50102 (0.00779)	0.50026 (0.00788)	0.50034 (0.00787)	0.50043 (0.00740)	0.50073 (0.00831)	0.50105 (0.00807)	0.50058 (0.00793)
	K-Nearest Neighbors	0.51293 (0.00806)	0.51344 (0.00823)	0.51265 (0.00840)	0.51275 (0.00847)	0.51272 (0.00797)	0.51316 (0.00875)	0.51330 (0.00854)	0.51277 (0.00827)
	Maximum Likelihood	0.50987 (0.00757)	0.51059 (0.00816)	0.50973 (0.00791)	0.51020 (0.00809)	0.51003 (0.00760)	0.51016 (0.00836)	0.51045 (0.00825)	0.51013 (0.00799)

Table 2.7 Root Mean Squared Error for Predicting Crop Yield with Precipitation

Estimation Method \ Crop	Barley	Corn	Soybean	Wheat
Local Linear	23.2103	33.3685	8.28392	20.0992
Nadaraya-Watson	23.184	33.3792	8.28634	20.2162
K-Nearest Neighbors	23.7871	34.2133	8.50096	20.4713
Maximum Likelihood	23.571	34.6139	8.42354	20.2922
Total Variation	23.6645	34.6909	8.42654	20.2924
Sample Size	4926	14690	12706	4409

Figure 2.1 State Average Barley Yield for States in the Barley Belt

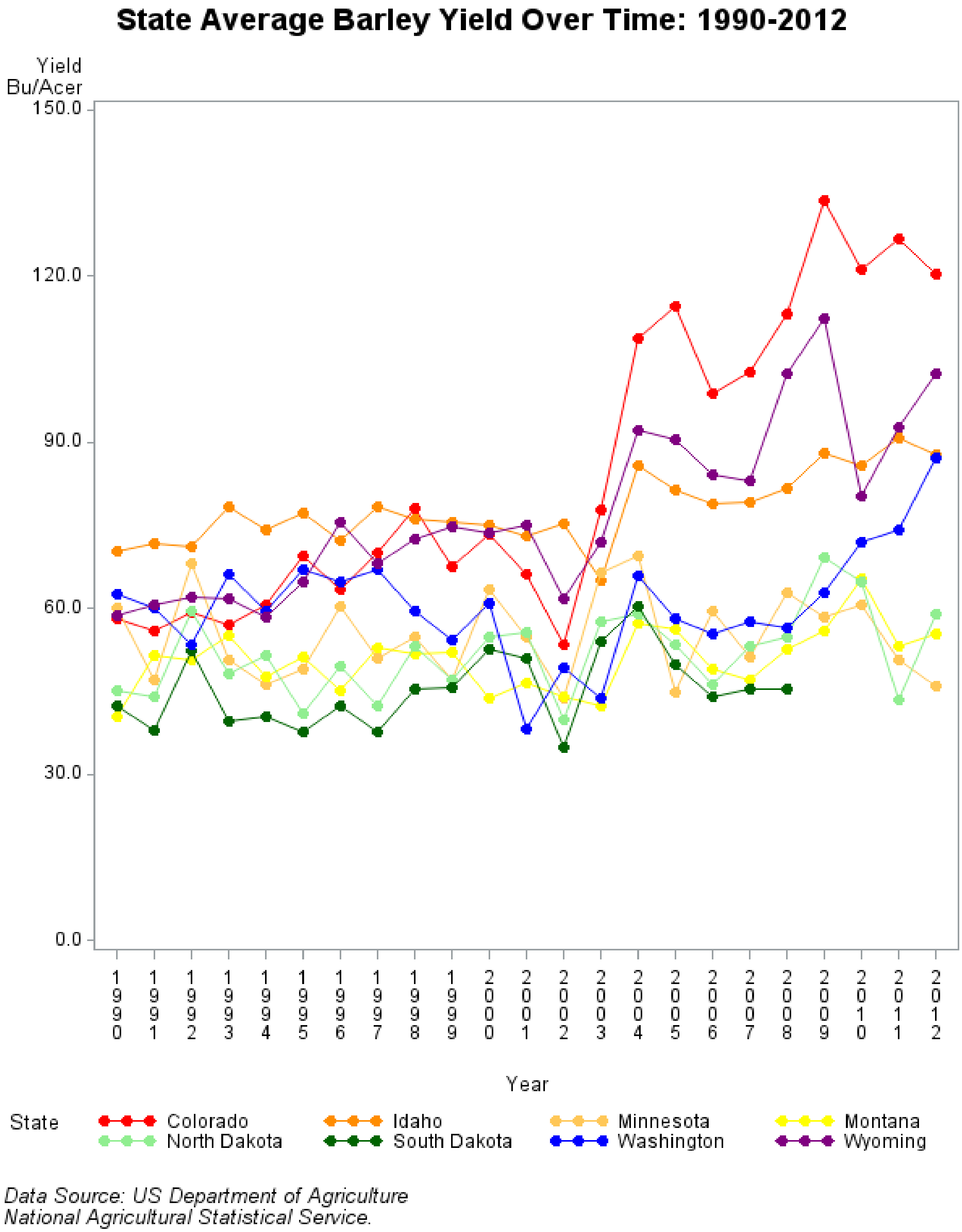


Figure 2.2 State Average Corn Yield for States in the Corn Belt

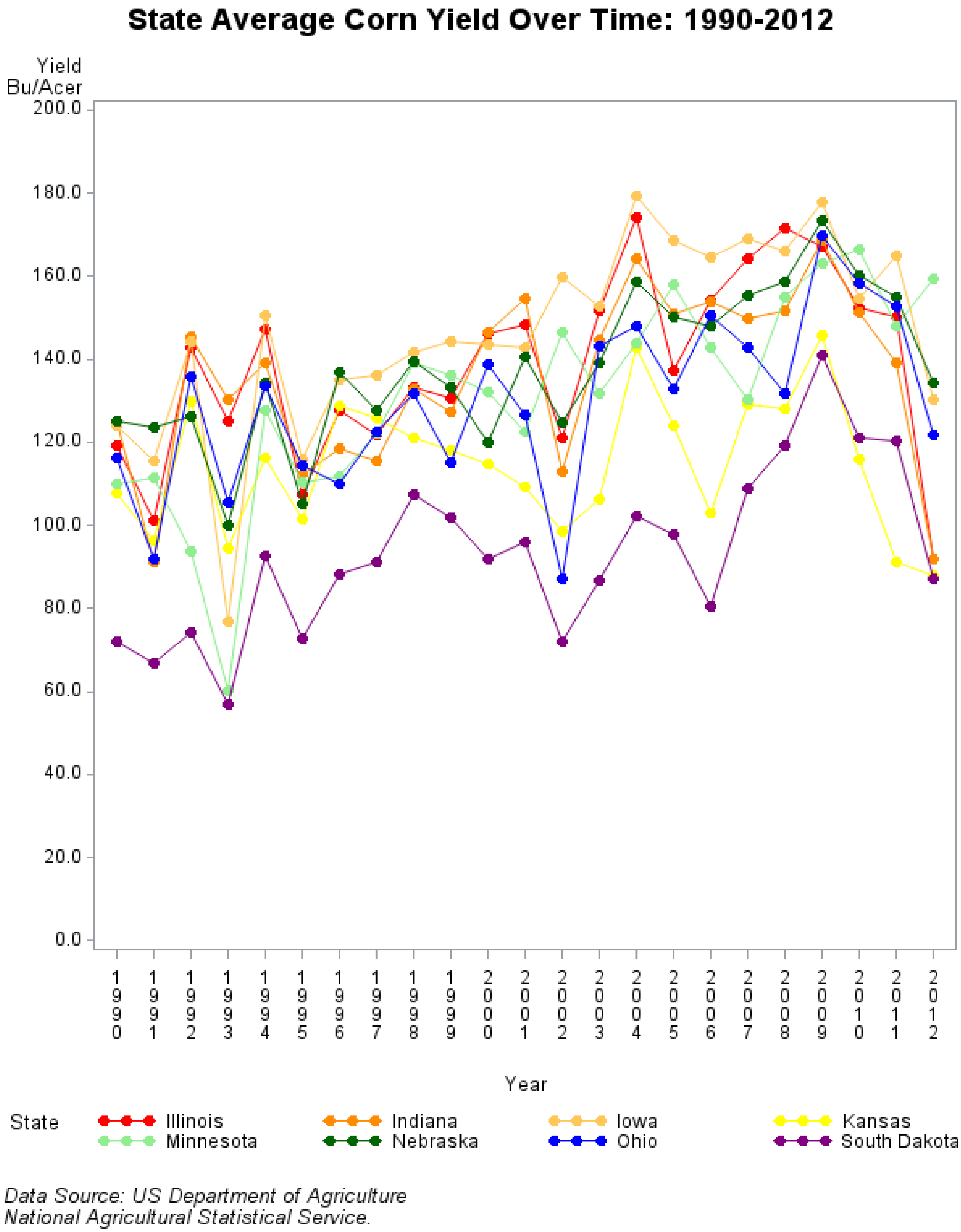


Figure 2.3 State Average Soybean Yield for States in the Soybean Belt

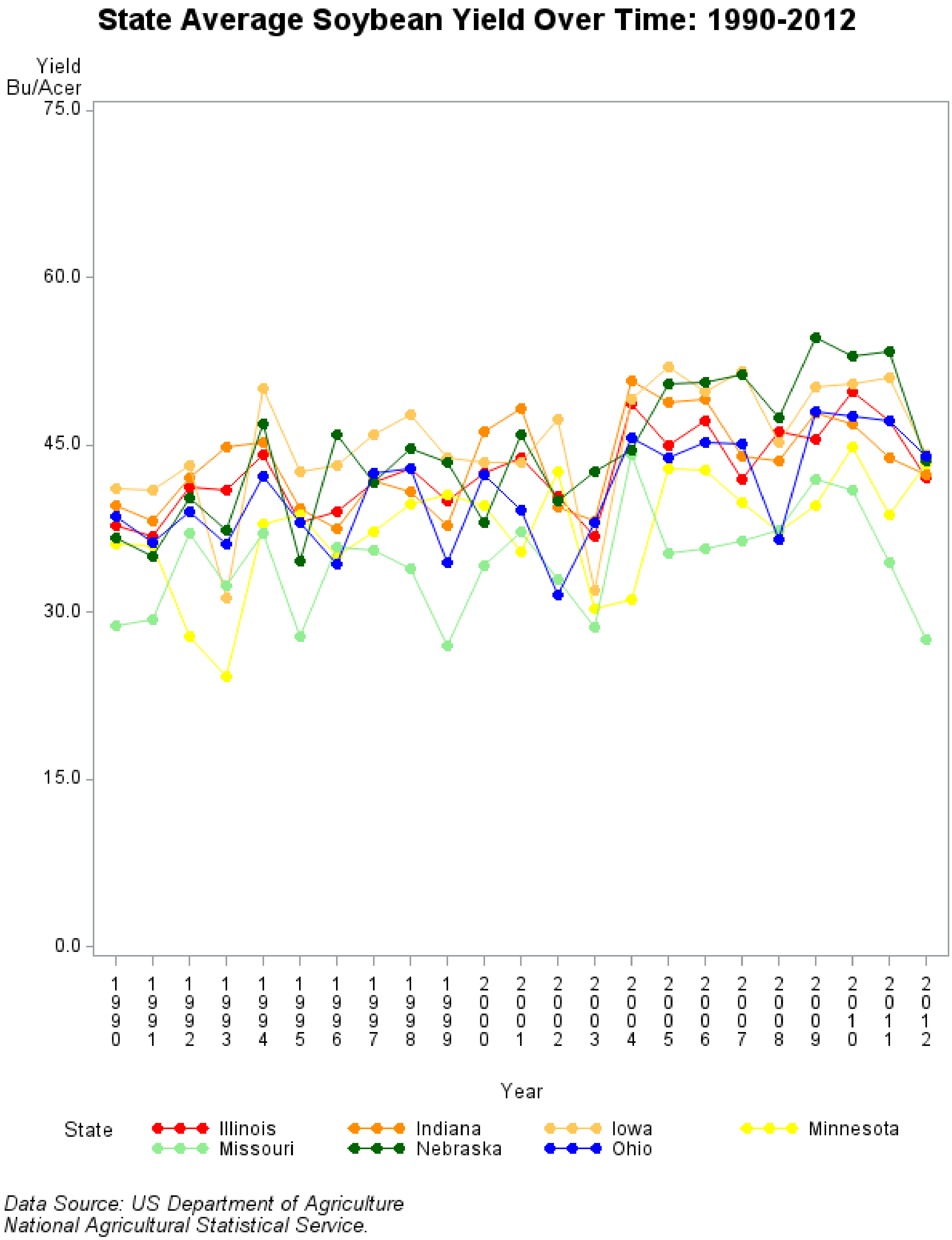


Figure 2.4 State Average Wheat Yield for States in the Wheat Belt

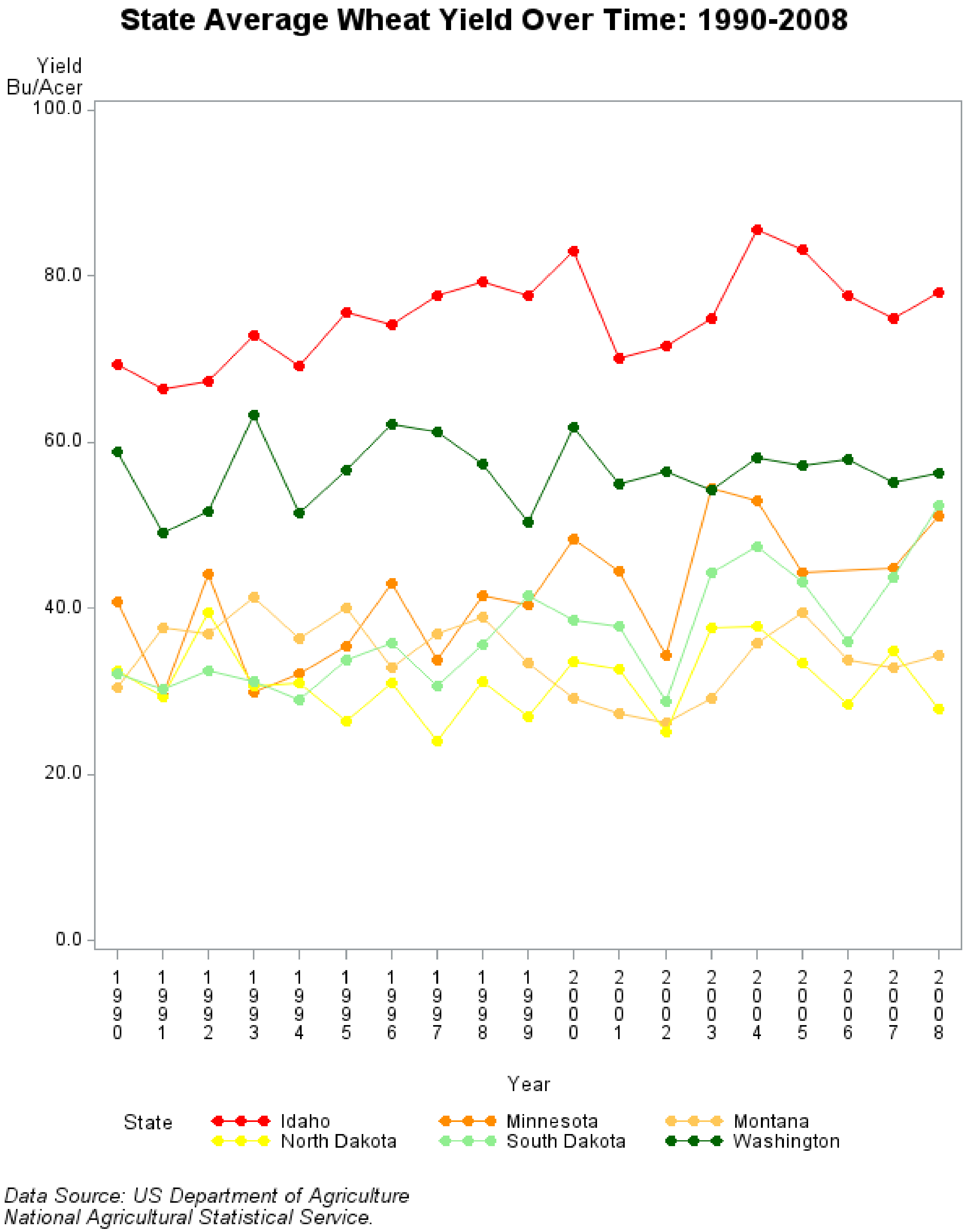


Figure 2.5 County Level Average Yield over Time

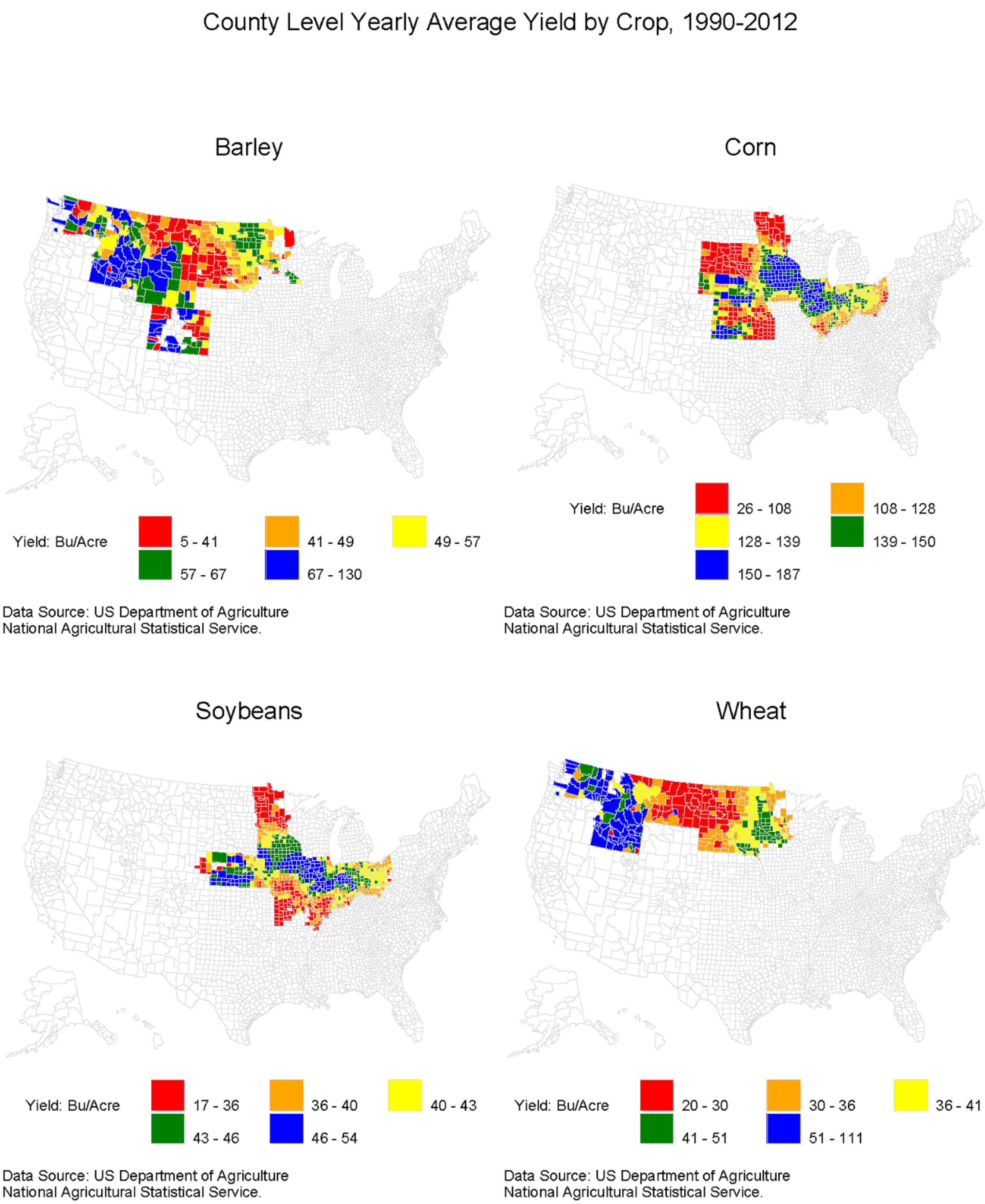
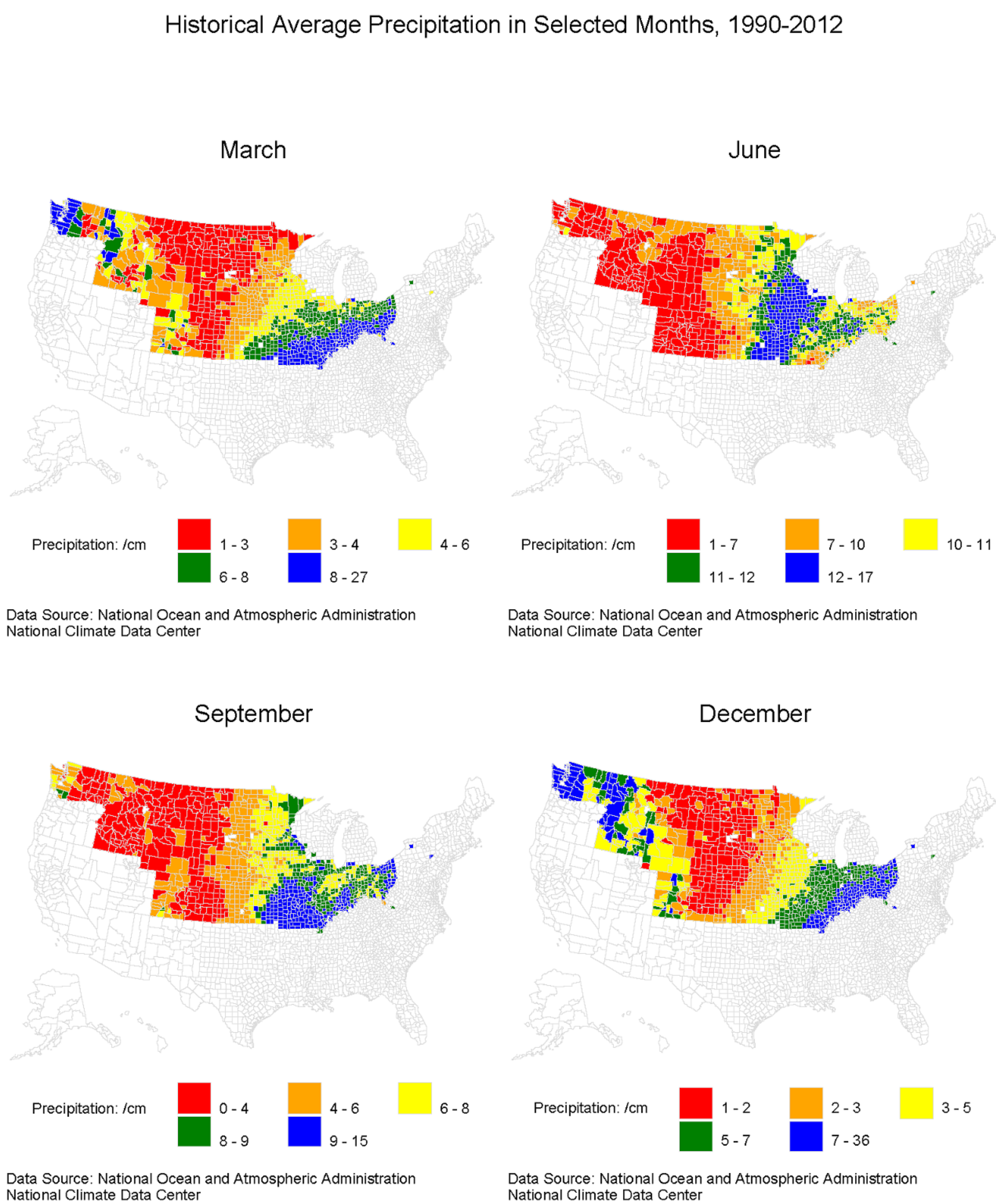


Figure 2.6 County Level Historical Average Precipitation over Time



Appendix 2.1 Asymptotic Distribution of Local Linear Kernel Estimator for Spatial Data (Hallin et al. (2004b)):

Under sufficient assumptions, let $\phi(x) = O(x^{-\mu})$ for some $\mu > 2(3+\delta)N/\delta$ and suppose that there exists a sequence of positive integers $q = q_{\boldsymbol{n}} \to \infty$ such that $q_{\boldsymbol{n}} = o((\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d})^{1/(2N)})$ and $\widehat{\boldsymbol{n}} q^{-\mu} \to 0$ as $\boldsymbol{n} \to \infty$, and the bandwidth $b_{\boldsymbol{n}}$ tends to zero in a way that $q b_{\boldsymbol{n}}{}^{\boldsymbol{\delta d}/[a(2+\delta)]} > 1$ for some $\frac{(4+\delta)N}{2+\delta} < a < \frac{\mu\delta}{2+\delta} - N$ as $\boldsymbol{n} \to \infty$[40]. Then,

$$\left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}\right)^{1/2}\left[\begin{pmatrix} g_{\boldsymbol{n}}(\boldsymbol{x}) - g(\boldsymbol{x}) \\ \left(g'_{\boldsymbol{n}}(\boldsymbol{x}) - g'(\boldsymbol{x})\right) b_{\boldsymbol{n}} \end{pmatrix} - \boldsymbol{U}^{-1}\begin{pmatrix} B_0(\boldsymbol{x}) \\ \boldsymbol{B}_1(\boldsymbol{x}) \end{pmatrix} b_{\boldsymbol{n}}{}^{2}\right] \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{U}^{-1}\boldsymbol{\Sigma}(\boldsymbol{U}^{-1})^{T}) \quad \text{as } \boldsymbol{n} \to \infty$$

Where

$$\boldsymbol{U_n} \xrightarrow{p} \boldsymbol{U} \equiv f(\boldsymbol{x})\begin{pmatrix} \int K(\boldsymbol{u}) d\boldsymbol{u} & \int \boldsymbol{u}^{\boldsymbol{T}} K(\boldsymbol{u}) d\boldsymbol{u} \\ \int \boldsymbol{u} K(\boldsymbol{u}) d\boldsymbol{u} & \int \boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}} K(\boldsymbol{u}) d\boldsymbol{u} \end{pmatrix} \quad \text{as } \boldsymbol{n} \to \infty,$$

$$\lim_{\boldsymbol{n}\to\infty} Var(\left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}\right)^{1/2} \boldsymbol{W_n}) = \boldsymbol{\Sigma} \equiv Var\left(Y_{\boldsymbol{j}} \middle| \boldsymbol{X_j} = \boldsymbol{x}\right) f(\boldsymbol{x})\begin{pmatrix} \int K^2(\boldsymbol{u}) d\boldsymbol{u} & \int \boldsymbol{u}^{\boldsymbol{T}} K^2(\boldsymbol{u}) d\boldsymbol{u} \\ \int \boldsymbol{u} K^2(\boldsymbol{u}) d\boldsymbol{u} & \int \boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}} K^2(\boldsymbol{u}) d\boldsymbol{u} \end{pmatrix},$$

$$B_0(\boldsymbol{x}) \equiv \frac{1}{2} f(\boldsymbol{x}) \sum_{i=1}^{d} \sum_{j=1}^{d} g_{ij}(\boldsymbol{x}) \int u_i\, u_j K(\boldsymbol{u}) d\boldsymbol{u},$$

---

[40] This condition for q and $b_{\boldsymbol{n}}$ are clearly explained in Hallin et al. (2004b). For two sequences of positive integer vectors, $\boldsymbol{p} = \boldsymbol{p_n} \equiv (p_1, \dots, p_N) \in \mathbb{Z}^N$ and $\boldsymbol{q} = \boldsymbol{q_n} \equiv (q, q, \dots, q) \in \mathbb{Z}^N$, with $q = q_{\boldsymbol{n}} \to \infty$ such that $p = p_{\boldsymbol{n}} \equiv \widehat{\boldsymbol{p}} = o((\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d})^{1/2})$, $q/p_k \to 0$ and $n_k/p_k \to \infty$ for all $k = 1, 2, \dots, N$, and $\widehat{\boldsymbol{n}}\phi(q) \to 0$. $b_{\boldsymbol{n}}$ tends to zero at such a rate that $q b_{\boldsymbol{n}}{}^{\delta d/[a(2+\delta)]} > 1$ and $b_{\boldsymbol{n}}{}^{-\delta d/(2+\delta)} \sum_{t=q}^{\infty} t^{N-1} [\phi(t)]^{\delta/2+\delta} \to 0 \quad as \quad \boldsymbol{n} \to \infty$.

$$\boldsymbol{B}_1(\boldsymbol{x}) \equiv \frac{1}{2} f(\boldsymbol{x}) \sum_{i=1}^{d} \sum_{j=1}^{d} g_{ij}(\boldsymbol{x}) \int u_i\, u_j \boldsymbol{u} K(\boldsymbol{u}) d\boldsymbol{u},$$

$$g_{ij}(\boldsymbol{x}) = \frac{\partial^2 g(\boldsymbol{x})}{\partial x_i \partial x_j}, i, j = 1, \dots, d, and\ \boldsymbol{u} \equiv (u_1, \dots, u_d)^{\boldsymbol{T}} \in \mathbb{R}^{\boldsymbol{d}}.$$

In addition, if the kernel $K(\cdot)$ is a symmetric density function, then the above asymptotic distribution can be reduced to:

$$\begin{pmatrix} \left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}\right)^{1/2} \left(g_{\boldsymbol{n}}(\boldsymbol{x}) - g(\boldsymbol{x}) - \boldsymbol{B}_g(\boldsymbol{x}) b_{\boldsymbol{n}}{}^2\right) \\ \left(\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d+2}\right)^{1/2} \left(g_{\boldsymbol{n}}'(\boldsymbol{x}) - g'(\boldsymbol{x})\right) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \begin{pmatrix} \sigma_0{}^2(\boldsymbol{x}) & 0 \\ 0 & \boldsymbol{\sigma}_1{}^2(\boldsymbol{x}) \end{pmatrix}\right)$$

Where

$$\boldsymbol{B}_g(\boldsymbol{x}) \equiv \frac{1}{2} \sum_{i=1}^{d} g_{ii}(\boldsymbol{x}) \int u_i{}^2\, K(\boldsymbol{u}) d\boldsymbol{u},$$

$$\sigma_0{}^2(\boldsymbol{x}) \equiv \frac{Var\left(Y_{\boldsymbol{j}} \middle| \boldsymbol{X}_{\boldsymbol{j}} = \boldsymbol{x}\right) \int K^2(\boldsymbol{u}) d\boldsymbol{u}}{f(\boldsymbol{x})}, and$$

$$\boldsymbol{\sigma}_1{}^2(\boldsymbol{x}) \equiv \frac{Var\left(Y_{\boldsymbol{j}} \middle| \boldsymbol{X}_{\boldsymbol{j}} = \boldsymbol{x}\right)}{f(\boldsymbol{x})} \times \left[\int \boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}} K(\boldsymbol{u}) d\boldsymbol{u}\right]^{-1} \left[\int \boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}} K^2(\boldsymbol{u}) d\boldsymbol{u}\right] \left[\int \boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}} K(\boldsymbol{u}) d\boldsymbol{u}\right]^{-1}.$$

To get the variance-covariance matrix of the local linear estimator, we need to estimate $f(\boldsymbol{x})$ and $Var\left(Y_{\boldsymbol{j}} \middle| \boldsymbol{X}_{\boldsymbol{j}} = \boldsymbol{x}\right) = E\left(Y_{\boldsymbol{j}}{}^{\boldsymbol{2}} \middle| \boldsymbol{X}_{\boldsymbol{j}} = \boldsymbol{x}\right) - g(\boldsymbol{x})^{\boldsymbol{2}}$ by the standard kernel method. The following estimators would be used:

$$\hat{f}(\boldsymbol{x}) = \frac{1}{\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}} \sum_{\boldsymbol{j} \in I_{\boldsymbol{n}}} K\left(\frac{\boldsymbol{X}_{\boldsymbol{j}} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right)$$

$$\hat{E}\left(Y_{\boldsymbol{j}}^{\boldsymbol{2}} \middle| \boldsymbol{X}_{\boldsymbol{j}} = \boldsymbol{x}\right) = \frac{1}{\widehat{\boldsymbol{n}} b_{\boldsymbol{n}}{}^{d}} \sum_{\boldsymbol{j} \in I_{\boldsymbol{n}}} \hat{e}_{\boldsymbol{j}}{}^{2}(\boldsymbol{x}) K\left(\frac{\boldsymbol{X}_{\boldsymbol{j}} - \boldsymbol{x}}{b_{\boldsymbol{n}}}\right) \quad with\ \hat{e}_{\boldsymbol{j}}(\boldsymbol{x}) = Y_{\boldsymbol{j}} - g_{\boldsymbol{n}}(\boldsymbol{x}).$$

Chapter 3: Modeling Extreme Price Movements in the Energy Markets

3.1 Introduction

Two of the most important lessons we can learn from the recent financial crisis is that tail events, such as a sharp fall in asset prices, could happen unexpectedly, and in an episode of deleveraging, illiquidity and flight to quality, the degree of co-movements in asset prices can spike with conventional uncorrelated asset prices - rising or falling together, see Vineer Bhansali (2008). In this paper, we use univariate and multivariate extreme value theory are to investigate the limit distributions of extreme returns, and more importantly, the correlation of extreme price movements in the energy markets. Specifically, the "exceedances over high threshold" model of Davidson and Smith (1990) is used to fit the univariate tail distribution of the return series derived from daily spot prices of WTI crude oil, New York Harbor No.2 Heating Oil, Gulf Coast Conventional Gasoline, TX Propane and LA Ultra-Low Sulfur Diesel. Next, we adopt the multivariate extreme value approach of Ledford and Tawn (1996) to estimate the bivariate tail dependence index for different pairs of returns on these energy products.

Extreme value theory has played an important methodological role in risk management for insurance and finance since the increasing complexity of the financial markets has exposed financial institutions to catastrophic losses. For a large class of distribution functions F, which may be unknown, of a random variable X, the tail distribution of X is given by the second theorem in extreme value theory, also called the Pickands-Balkema-de Haan theorem established by Balkema and de Haan (1974) and Pickands (1975). The limit tail distribution is usually approximated by a gerneralized Pareto distribution. For example, Davidson and Smith (1990) detailed estimation procedures and specification analysis on how to use the generalized Pareto distribution to fit the size and occurrence of exceedances over high thresholds for a univariate random variable. Since then, extreme value theory has been widely used to model tail-related loss distribution and to construct risk measures for univariate data.

The statistical distribution assumption for the returns on the underlying asset plays a central role in financial modeling. It is necessary to test asset pricing theories, to construct optimal portfolios, to value and hedge derivative positions, and to measure and manage financial risk, see Embrechts (2002) and Lombardi and Ravazzolo (2013). Previous research has usually assumed that the underlying asset returns follow an IID normal distribution. However, the fact that the frequency of extreme price movements in the financial markets are much higher than that implied by the normal distribution seems to contradict the normality assumption. In response, Hols and De Vries (1991) advocated the use of extreme value distribution for modelling fat-tailed exchange rate return distributions. The advantage is that the parameter characterizing the degree of tail fatness can be estimated without a pre-assumed underlying distribution. One can assess the very small probability of an exceedance over a high threshold with the estimated parameters. Longin (2005) also showed that extreme value theory can be useful to more precisely characterize the distribution of asset returns and help to choose a more suitable model where the tails of the distribution is concerned.

Since risk management is in essence the practice of allocating capital to absorb potential losses in instances of adverse market movements, risk managers are particularly concerned about extreme price movements in the underlying asset. Hence, extreme value theory has become a popular and suitable tool in risk management. To calculate value at risk (VaR) and expected shortfall, McNeil and Frey (2000) used extreme value theory to estimate the tail distribution of the disturbance term in the Garch model for current volatility. The advantage of this approach compared to other methods that ignore either the heavy tails in the disturbance term distribution or the stochastic nature of volatility is shown by back-testing with historical daily return series. Longin (2000) pointed out that computing VaR based on extreme values can cover both the normal market conditions considered by existing VaR methods and periods of turbulence concerned by stress testing since the limit distribution of extreme returns, which is usually fit by a generalized Pareto distribution, is to a large extent independent of the underlying distribution of returns. Marimoutou et. al. (2009) calculated VaR for long and short positions in the oil market applying both unconditional and conditional EVT. Their results showed that conditional EVT and filtered historical simulation approach provide major improvements over other conventional methods. Using extreme value theory, Cotter (2001) computed unconditional optimal margin levels for stock index futures traded on European exchanges.

While most of the applications of extreme value theory focus on fitting the tail distribution or calculating risk measures such as VaR or expected shortfall based on the univariate underlying asset returns; recently multivariate extreme value theory has also been used to model correlation and dependence in different assets or markets. Hartmann et. al. (2004) have constructed an extremal dependence measure for asset market linkages. Their results show that simultaneous crashes are more likely to happen in stock markets than in bond markets for the G-5 countries. At the same time, the frequency of stock-bond contagion could be comparable to that of flight to qualify from stocks into bonds. Using multivariate extreme value theory, Longin and Solnik (2001) have modeled the extreme dependence in international equity returns and their results show that correlation tend to be higher in bear markets. Poon et. al. (2004) presented a general framework for modeling the joint-tail distribution based on multivariate extreme value theory.

Their results using returns on five major stock indices have shown that the use of traditional dependence measures could lead to inaccurate portfolio risk assessment, especially when studying extreme events such as systemic risk and crisis.

In this paper, we investigate the tail distribution of the returns on spot prices of crude oil, Gasoline, Heating Oil, Propane and Diesel and study the bivariate extreme dependence between the price movements in these energy products. Special attention should be paid to this issue since recent increases in the frequency of sharp falls in energy prices, especially in oil prices, have significant impact on economic activity and have become a major concern for consumers, firms and governments. In addition, as a result of rapidly growing index investment, synchronized price booms and busts of a series of conventionally unrelated commodities have become a common phenomenon, see Tang and Xiong (2010). As an important part of commodity markets, we would also expect increased correlation between energy prices. Studying the extreme movements and extreme correlation in the energy markets is thus of crucial importance for energy portfolio optimization and risk management.

The rest of the paper is organized as follows: Section 3.2 gives out the modeling framework based on both univariate and multivariate extreme value theory. Section 3.3 deals with the econometric methodology and presents the empirical results. Section 3.4 concludes.

3.2 Models for multivariate extreme returns

The model used for joint tail estimation is the multivariate extreme value threshold model of Ledford and Tawn (1996). Analogous to the Copula theory, this multivariate extreme value threshold model decomposes the joint tail distribution into its marginal tail distributions and a dependence function characterizing the dependence between the marginal tail distributions. The limit results involve the marginal structure and the dependence structure, which we will discuss separately in the following sections.

3.2.1 Univariate Tail Distribution

We will follow the "peaks over threshold" approach of Pickands (1975) and Davison and Smith (1990) for marginal distribution of tail events. Suppose the d-dimensional random variable $(R_j; j = 1, \dots, d)$ has a joint distribution function F. Consider the unknown cumulative distribution function F_j of univariate random variable R_j, which denotes the return on the jth energy price. Extreme returns of R_j are defined as those returns in exceedance of a certain threshold u_j. In this paper, we will identify right tail exceedances as those observations of R_j greater than the threshold u_j; Similarly, left tail exceedances observations are those with the negative of R_j greater than the threshold u_j. For a given small positive value p_j, if the probability of R_j exceeding threshold u_j equals p_j, then $p_j = 1 - F_j(u_j)$. Here, we are interested in estimating the conditional distribution function F_{j,u_j} of R_j above a certain threshold u_j, which is characterized by the "conditional excess distribution function", defined as

$$F_{j,u_j}(y) = P\left(R_j - u_j \le y \mid R_j > u_j\right) = \frac{F_j\left(u_j + y\right) - F_j(u_j)}{1 - F_j(u_j)} \quad for\ 0 \le y \le R_{F_j} - u_j$$

where R_{F_j} is the right endpoint of the underlying marginal distribution F_j, it can be finite or infinite. The function F_{j,u_j} describes the distribution of the excess value over a threshold u_j, conditional on the threshold is exceeded.

Given a sequence of observations $(R_{j1}, R_{j2}, \dots, R_{jn})$ from R_j, with conditional excess distribution function F_{j,u_j}, extreme value theory will try to search for a possible non-degenerate limiting distribution for F_{j,u_j} as the threshold u_j tends to the upper point R_{F_j}. Balkema and de Haan (1974) and Pickands (1975) pointed out that for a very wide class of underlying distribution functions F_j, and large u_j, F_{j,u_j} is well approximated by the generalized Pareto distribution. That is:

$$F_{j,u_j}(y) \rightarrow G_{\xi,\sigma}(y), \quad as\, u \rightarrow \infty$$

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}}, & if \;\; \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}}, & if \;\; \xi = 0. \end{cases}$$

Here $\sigma > 0$, $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\sigma/\xi$ when $\xi < 0$. The case $\xi = 0$ corresponds to the exponential distribution with mean σ. If $\xi = 1$, $G_{k,\sigma}(y)$ is a uniform distribution on $[0,\ \sigma]$. If $\xi < 0$, $G_{\xi,\sigma}(y)$ is the Pareto distribution. The properties of the distribution tail is determined by the tail index ξ. For a given threshold, the univariate distribution tail could be characterized by three parameters: the tail probability p_j, the dispersion parameter σ, and the tail index ξ. The case of $\xi > 0$ corresponds to a power-declining tail, usually called fat-tailed distributions; The case of $\xi = 0$ corresponds to exponentially declining tail, usually called thin-tailed distributions; The case of $\xi < 0$ corresponds to distributions with no tail or finite tail distributions.

3.2.2 Modeling Multivariate Tail Dependence

To model the d-dimensional random variable $(R_j; j = 1, \ldots, d)$ with joint distribution function F, and marginal distribution functions F_j for R_j $(j = 1, \ldots, d)$, following Ledford and Tawn (1996) and Heffernan and Tawn (2004), we first do a transformation $Z_j = -1/logF_j(R_j)$ for each j, which transforms the marginal variable R_j into a unit Frechet variable Z_j with $\Pr(Z_j \leq z) = \exp(-1/z)$ for $0 < z < \infty$. Let F_* be the joint distribution function of $(Z_1, \ldots, Z_d)$, then $F_*(z_1, \ldots, z_d) = F(r_1, \ldots, r_d)$. Multivariate extreme value theory is based on the assumption that F is in the domain of attraction[41] of a multivariate extreme value distribution, which means that the normalized componentwise maxima of observations from F will have a non-degenerate multivariate limit distribution. This is equivalent to F_* being in the domain of attraction of a multivariate extreme value distribution with unit Frechet as marginal distribution, see Resnick (1987) and Tawn (1990). The joint distribution F_* is required to satisfy the following condition

$$F_*(z_1, \ldots, z_d) = \exp(-V(z_1, \ldots, z_d))$$

for $Z_j > v_j$ $(j = 1, \ldots, d)$, where v_j is the threshold for Z_j related the threshold of R_j by $v_j = -1/logF_j(u_j)$, and $V(*)$ is a function capturing the dependence structure between marginal distributions. This could also be expressed in terms of the original random variable $(R_j; j = 1, \ldots, d)$ as

$$F(r_1, \ldots, r_d) = \exp\left(-V\left(-\frac{1}{logF_1(r_1)}, \ldots, -\frac{1}{logF_d(r_d)}\right)\right)$$

when each r_j exceeds its corresponding threshold u_j.

[41] Given a sequence of i.i.d. sample $X_1, X_2, \ldots, X_n$ from distribution function f, if $\lim_{n\to\infty} P\{a_n{}^{-1}(\max(X_1, X_2, \ldots, X_n) - b_n) \leq x\} = GEV(x)$ for some non-degerate distribution function $GEV(*)$, where $a_n > 0$ and $b_n \in \mathbb{R}$ are normalizing constants that could be appropriately chosen so that $GEV(x) = GEV_\xi(x) = exp(-(1 + \xi x)^{-1/\xi})$ for all x such that $1 + \xi x > 0$, then we say that the distribution function f belongs to the domain of attration of $GEV_\xi(x)$. Actually, the extreme value distribution $GEV_\xi(x)$ can be characterized by three classes of distributions, the Gumbel, Frechet and Weibull distribution, which in turn contain distribution functions with exponential tail, polynomially decaying tail and light tail with finite right endpoint in its domain of attraction, respectively.

For the marginal distribution F_j, we could use the generalized Pareto distribution to model each marginal distribution above a threshold following Davison and Smith (1990). Making the marginal thresholds coincide with the thresholds u_j in the dependence structure gives,

$$F_{j,u_j}(r_j) = 1 - p_j \left(1 + \frac{\xi(r_j - u_j)}{\sigma}\right)_+^{-\frac{1}{\xi}}, \quad r_j \geq u_j$$

where p_j is some small probability chosen so that u_j is the $1 - p_j$ quantiles of the marginal distribution of R_j[42]. The transformed marginal variable is given by $Z_j = -\frac{1}{logF_j(R_j)}$.

For the dependence structure $V(*)$, we use the popular multivariate logistic function[43],

$$V(z_1, \ldots, z_d) = (z_1^{-\frac{1}{\gamma}} + \cdots + z_d^{-\frac{1}{\gamma}})^{\gamma}$$

with dependence parameter $\gamma \in (0, 1]$. When the marginal variables are independent, $V(z_1, \ldots, z_d) = \sum z_j^{-1}$. The multivariate distribution function can be factorized into the product of the marginal distributions as

$$F(r_1, \ldots, r_d) = \exp(-V(z_1, \ldots, z_d)) = \exp\left[-\sum_{j=1}^{d} z_j^{-1}\right] = \prod_{j=1}^{d} \exp\left(-\left(-\frac{1}{logF_j(r_j)}\right)^{-1}\right)$$

$$= \prod_{j=1}^{d} F_{j,u_j}(r_j)$$

42 $F_{j,u_j}(r_j) \equiv P(R_j \leq r_j) = P(R_j \leq u_j) + P(u_j \leq R_j \leq r_j) = 1 - p_j + P(R_j \leq r_j | R_j > u_j)P(R_j > u_j) = 1 - p_j + \left[1 - \left(1 + \frac{\xi(r_j - u_j)}{\sigma}\right)_+^{-\frac{1}{\xi}}\right] p_j = 1 - p_j \left(1 + \frac{\xi(r_j - u_j)}{\sigma}\right)_+^{-\frac{1}{\xi}}$.

43 The dependence structure $V(*)$ has to satisfy certain conditions in order to be used to represent multivariate extreme value distributions with unit Freshet margins, and logistic function is a simple example that satisfies those conditions.See for example, Pickands (1981), Joe (1990), Tawn (1990), Coles and Tawn (1991) and Ledford and Tawn (1997).

for $r_j > u_j$. In this study, we will focus on the 2-dimensional case[44], the relationship between correlation coefficient ρ for the two marginal distributions and the dependence parameter α is $\rho = 1 - \gamma^2$ as shown by Tiago de Oliveira (1973). The variables tend to be totally dependent as $\gamma \to 0$. As α increases the dependence weakens, and the variables are independent when $\gamma = 1$.

3.2.3 Maximum Likelihood Estimation

When all the component R_j exceed their corresponding thresholds, we could calculate the distribution function or likelihood by just plugging the transformed $F_{j,u_j}(r_j)$ into the dependence structure. If one marginal component occurs below the threshold, we only know that it does not exceed the threshold, but not its actual value. To construct the likelihood for the sample, we take marginal components below their respective thresholds as if they are censored at the corresponding thresholds. Ledford and Tawn (1996) derived the likelihood contribution of an observation $(r_1, \dots, r_d)$ with components $(j_1, \dots, j_m)$ exceeding their corresponding thresholds as

$$\left.\frac{\partial^m F(r_1, \dots, r_d)}{\partial r_{j_1} \dots \partial r_{j_m}}\right|_{\{r_j = \max(u_j, r_j), j=1,\dots,d\}}$$

where $F(r_1, \dots, r_d) = \exp\left(-V\left(-\frac{1}{logF_1(r_1)}, \dots, -\frac{1}{logF_d(r_d)}\right)\right)$. To construct the likelihood function with the logistic dependence structure in the 2-dimensional case, we need to consider four cases how a given sample point will contrtibute to the likelihood function according to

[44] We did this just for simplicity of computation. A multivariate model to estimate all the parameters for all the five energy products at the same time may produce more consistent parameter estimates and lead to better model fit. However, as will be shown in the following procedures of maximum likelihood estimation, due to the combinatorial explosion, the number of cases how a sample point will contribute to the likelihood function would be 2^d, where d equals 5, the number of dimensions.

whether a sample component exceeds the corresponding threshold. The space of the marginal components in the bivariate case could be partitioned into four regions

$$\{Reg_{kl}; k = I_{\{R_1>u_1\}}, l = I_{\{R_2>u_2\}}\},$$

and I is the indicator function. For convenience, write $v_j = -1/\log(1-p_j)$ as the threshold for the transformed marginal variables $Z_j = -\frac{1}{logF_j(R_j)}$. Then the likelihood contribution of a point (r_1, r_2) in region Reg_{kl} would be $L_{kl}(r_1, r_2)$, defined as

$$L_{00}(r_1, r_2) = \exp\big(-V(v_1, v_2)\big),$$

$$L_{01}(r_1, r_2) = \exp\big(-V(v_1, z_2)\big)V_2(v_1, z_2)K_2,$$

$$L_{10}(r_1, r_2) = \exp\big(-V(z_1, v_2)\big)V_1(z_1, v_2)K_1,$$

$$L_{11}(r_1, r_2) = \exp\big(-V(z_1, z_2)\big)(V_1(z_1, z_2)V_2(z_1, z_2) - V_{12}(z_1, z_2))K_1K_2,$$

where $K_j = -p_j{\sigma_j}^{-1}\left(\left(1+\frac{\xi_j(r_j-u_j)}{\sigma}\right)_+^{-\frac{1}{\xi_j}}\right)^{1+\xi_j} {z_j}^2\exp(1/z_j)$, V_j and V_{12} are the partial derivatives of V with respect to the jth component and the mixed derivative, respectively. The likelihood contribution from a typical point (y_{1t}, y_{2t}) in the model using logistic function with parameter α for the dependence structure and generalized Pareto distribution with parameters Θ: $\{\sigma_j, \xi_j, p_j: j = 1, 2\}$ for the marginal distributions can be written as

$$L_t(\alpha, \Theta) = \sum_{k,l\,\in\{0,1\}} L_{kl}(r_{1t}, r_{2t})\, I_{kl}(r_{1t}, r_{2t})$$

where $I_{kl}(r_{1t}, r_{2t}) = I_{\{(r_{1t}, r_{2t}) \in Reg_{kl}\}}$. Then the likelihood for a set of T independent points is given by $L_{(T)}(\alpha, \Theta) \equiv \prod L_t(\alpha, \Theta)$. Maximizing the likelihood function will give the estimate for (α, Θ).

3.3 Empirical Results

We will estimate both the univariate tail distribution parameters and the pairwise bivariate dependence parameter for daily return series derived from spot prices of WTI crude oil, Gulf Coast Conventional Gasoline, New York Harbor No.2 Heating Oil, TX Propane and LA Ultra-Low Sulfur Diesel. The data is from EIA's website[45]. In this section, we will first give a brief summary of the data. Then we will present the estimation results for both univariate parameter and bivariate dependence parameter. Since risk managers are usually concerned about the negative tail of the return distribution, we will first report results for the negative tail estimates. Although estimates for the positive tail are also reported.

3.3.1 Summary Statistics

The price data downloaded from EIA ranges from April 18, 1996 to May 5, 2014 with 4532 observations in total. We take the percentage changes in the price series to get the return data for each of the energy products. The summary statistics for the return series are given in Table 3.1. Both the mean and the median are very close to 0 for all the five return series. Also there are sharp increase and decrease in all the five return series with minimum returns -16.414%, -22.954%, -37.954%, -17.674% and -26.8264% and maximum returns 17.092%, 47.012%, 38.676%, 49.913% and 22.716%. These extreme observations in the energy markets necessitate the modelling of tail behavior in the energy markets. The return series for heating oil and

---

[45] http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm

propane are also slightly skewed with a skewness index -1.54598 and -2.64803, respectively. The prices and returns for the five energy products are also given in Figure 3.1 and Figure 3.2.

3.3.2 Univariate Tail Distribution Parameter Estimation

Table 3.2 and Table 3.3 gives the parameter estimates for tail exceedances fitted by the generalized Pareto distribution. The first column is the percentage quantile used to decide a threshold for the model, which means 0.5%, 1%, 2%, 3%, 4%, 5%, 7.5% and 10% of most extreme observations are chosen to estimate parameters of the limiting generalized Pareto distribution. Notice that as the percentage quantile increases, one expects smaller standard errors for the parameter estimates. This is because more observations included in parameter estimation will lead to more precise parameter estimates. However, it is not necessarily better since the increased number of observations may induce bias as more observations that do not belong to the tail are included for parameter estimation.

The maximum likelihood estimates for the distribution parameters based on left tail observations are given in the Table 3.2. Most of the tail index estimates are between 0.2 and 0.4. Since a Student's t-Distribution with k degrees of freedom has a tail index $1/_k$, this means that the return distribution in general has a tail similar to that of Student's t-Distribution with 2 to 4 degrees of freedom. What's different from others is the tail index estimates for crude oil, they are generally smaller, ranging from 0.048 to 0.2311. This means that the tail of the return distribution for crude oil is not as fat as that of the other energy products and is closer to the tail of a normal distribution[46]. As to the scale index, almost all the estimates are between 1 and 3,

[46] Although the results show that the tail of the distribution of returns on crude oil is not as fat as those on heating oil, gasoline, diesel and propane, which are refined using crude oil as a primary input, this does not imply that returns on crude oil have lower volatility than the returns on other energy products. One reason is that the volatility of returns depends not only on tail fatness, but also whether there is high peak near the center of the distribution, which cancels the effect of fat tails on volatility. In fact, the

which makes sense since they are comparable to the unconditional standard errors of the return series reported in Table 3.1, although usually slightly lower than the corresponding standard errors.

Table 3.3 gives the parameter estimates for the limiting distribution of the right tail observations. The tail index estimates are similar to those for the left tail observations for most of the energy products. Most of the tail index estimates for propane and diesel are between 0.2 and 0.4, meaning a tail fatness similar to that of a Student's t-Distribution of 2 to 5 degrees of freedom. For heating oil and gasoline, the tail index estimates are higher when thresholds are high (lower quantile percentage), with values above 0.5. Similar to the left tail observations, the tail index estimates for the crude oil also tend to be smaller. This is consistent to a previous claim that the distribution of crude oil returns may be closer to a normal distribution with regard to tail fatness.

3.3.3 Bivariate Dependence Parameter Estimation

In this section, we report estimation results of the bivariate extreme models for the return dependence between crude oil and the other four energy products. We used separate bivariate models for pairs of extreme values of crude oil/gasoline, crude oil/heating oil, crude oil/propane, crude oil/diesel. The parameter estimates for the univariate models and the unconditional correlation are used as starting values when we maximize the likelihood function for the bivariate models.

---

volatility estimates are 0.01096, 0.01090, 0.01250, 0.01010, and 0.01044 for returns on crude oil, heating oil, gasoline, diesel and propane, respectively.

Table 3.4 through Table 3.7 present the estimation results of bivariate threshold exceedance models for the four pairs of energy products. The thresholds used for the returns of the two energy products in a pair are the same each time we estimate the bivariate threshold exceedance model. For every pair of energy products, estimation results for both the negative tail exceedance (both returns lower than the threshold corresponding to a low quantile of the distribution) and the positive tail exceedance (both returns are higher than the threshold corresponding to a high quantile of the distribution) are reported in the same table. In general, the correlation coefficients calculated from the dependence index estimates are different from the unconditional correlation coefficient estimated with all the data. The implied correlation coefficients of the threshold exceedance models are in most cases higher than the unconditional and vary with the threshold chosen.

Specifically, for the bivariate threshold exceedance model for Crude Oil/Heating Oil, the correlation coefficients implied by the dependence index are 0.6774, 0.7007, 0.6849, 0.8714, 0.8702, 0.7371 and 0.8492 for negative thresholds corresponding to 1%, 2%, 3%, 4%, 5%, 7.5% and 10% quantiles, respectively. These numbers are 0.7139, 0.7570, 0.5927, 0.8710, 0.8692, 0.7115 and 0.7084 for positive thresholds corresponding to the same quantiles. All these estimates for the correlation coefficients are higher than the unconditional estimates for the correlation coefficients, which is 0.6497. Similar for Crude Oil/Gasoline, the correlation coefficients implied by the dependence index are also larger than the unconditional correlation coefficients. These results are consistent with the observed increased correlation in the financial markets during the crisis periods, as pointed out by Vineer Bhansali (2008). However, the case for Crude Oil/Propane and Crude Oil/Diesel is a little bit different, as can be seem from Table 3.6 and Table 3.7. For both the Crude Oil/Propane and Crude Oil/Diesel threshold exceedance models, the correlation coefficients is smaller than the unconditional correlation coefficients when a threshold corresponding to a low percentage quantile is used, and the correlation coefficient will increase and exceed the unconditional correlation coefficient as the percentage quantile increases. How the implied correlation coefficients changes with the thresholds used for the threshold exceedance model is demonstrated in Figure 3.3 through Figure 3.6.

3.4 Conclusions

In this paper, we have investigated the univariate tail distribution for the returns on various energy products such as Crude Oil, Gasoline, Heating Oil, Propane and Diesel. Tail index estimates for univariate threshold exceedance models show that these returns generally have fat tails similar to those of a Student's t-Distribution with 2 to 5 degrees of freedom except that the tail index estimates for the returns on Crude Oil is closer to that of a normal distribution.

We have also used the bivariate threshold exceedance model to study the extreme dependence between returns on two energy products. Parameters characterizing the extreme dependence are estimated for four pairs of energy products, crude oil/gasoline, crude oil/heating oil, crude oil/propane, crude oil/diesel. Correlation coefficients implied by the dependence index estimates show that correlations conditional on threshold exceedance are generally higher than the unconditional correlation for the correlation between crude oil/heating oil and crude oil/gasoline. However, there are some variation in the implied correlation between crude oil/propane and crude oil/diesel, whether the extreme correlation will be higher or lower than the unconditional correlation will depend on the threshold chosen.

3.5 References

Balkema, August A., and Laurens De Haan. "Residual life time at great age." The Annals of Probability (1974): 792-804.

Bhansali, Vineer. "Tail risk management." The Journal of Portfolio Management 34, no. 4 (2008): 68-75.

Bhansali, Vineer, Robert Gingrich, and Francis A. Longstaff. "Systemic credit risk: What is the market telling us?" Financial Analysts Journal (2008): 16-24.

Bollerslev, Tim, and Jeffrey M. Wooldridge. "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances." Econometric reviews 11, no. 2 (1992): 143-172.

Chan-Lau, Jorge A., Donald J. Mathieson, and James Y. Yao. "Extreme contagion in equity markets." IMF Staff Papers (2004): 386-408.

Cotter, John. "Margin exceedances for European stock index futures using extreme value theory." Journal of Banking & Finance 25, no. 8 (2001): 1475-1502.

Danielsson, Jon, and Casper G. de Vries. "Tail index and quantile estimation with very high frequency data." Journal of empirical Finance 4, no. 2 (1997): 241-257.

Davison, Anthony C., and Richard L. Smith. "Models for exceedances over high thresholds." Journal of the Royal Statistical Society. Series B (Methodological) (1990): 393-442.

Embrechts, Paul, Sidney I. Resnick, and Gennady Samorodnitsky. "Extreme value theory as a risk management tool." North American Actuarial Journal 3, no. 2 (1999): 30-41.

Embrechts, Paul, Alexander McNeil, and Daniel Straumann. "Correlation and dependence in risk management: properties and pitfalls." Risk management: value at risk and beyond (2002): 176-223.

Engle, Robert. "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models." Journal of Business & Economic Statistics 20, no. 3 (2002): 339-350.

Engle, Robert F., and Kenneth F. Kroner. "Multivariate simultaneous generalized ARCH." Econometric theory 11, no. 01 (1995): 122-150.

Frey, Rüdiger, and Alexander J. McNeil. "VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights." Journal of Banking & Finance 26, no. 7 (2002): 1317-1334.

Ganeshapillai, Gartheeban, John Guttag, and Andrew Lo. "Learning Connections in Financial Time Series." In Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 109-117. 2013.

Genest, Christian, and Louis-Paul Rivest. "Statistical inference procedures for bivariate Archimedean copulas." Journal of the American statistical Association 88, no. 423 (1993): 1034-1043.

Hartmann, Philipp, Stefan Straetmans, and Casper G. De Vries. "Asset market linkages in crisis periods." Review of Economics and Statistics 86, no. 1 (2004): 313-326.

Heffernan, Janet E., and Jonathan A. Tawn. "A conditional approach for multivariate extreme values (with discussion)." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66, no. 3 (2004): 497-546.

Hols, Martien CAB, and Casper G. De Vries. "The limiting distribution of extremal exchange rate returns." Journal of Applied Econometrics 6, no. 3 (1991): 287-302.

Joe, Harry. "Families of min-stable multivariate exponential and multivariate extreme value distributions." Statistics & probability letters 9, no. 1 (1990): 75-81.

McNeil, Alexander J., and Rüdiger Frey. "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach." Journal of empirical finance 7, no. 3 (2000): 271-300.

Kearns, Phillip, and Adrian Pagan. "Estimating the density tail index for financial time series." Review of Economics and Statistics 79, no. 2 (1997): 171-175.

Ledford, Anthony W., and Jonathan A. Tawn. "Statistics for near independence in multivariate extreme values." Biometrika 83, no. 1 (1996): 169-187.

Ledford, Anthony W., and Jonathan A. Tawn. "Modelling dependence within joint tail regions." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59, no. 2 (1997): 475-499.

Lombardi, Marco Jacopo, and Francesco Ravazzolo. "On the correlation between commodity and equity returns: implications for portfolio allocation." (2013).

Longin, Francois M. "From value at risk to stress testing: The extreme value approach." Journal of Banking & Finance 24, no. 7 (2000): 1097-1130.

Longin, Francois. "The choice of the distribution of asset returns: How extreme value theory can help?" Journal of Banking & Finance 29, no. 4 (2005): 1017-1035.

Longin, Francois, and Bruno Solnik. "Extreme correlation of international equity markets." The Journal of Finance 56, no. 2 (2001): 649-676.

Marimoutou, Velayoudoum, Bechir Raggad, and Abdelwahed Trabelsi. "Extreme value theory and value at risk: application to oil market." Energy Economics 31, no. 4 (2009): 519-530.

Okimoto, Tatsuyoshi. "New evidence of asymmetric dependence structures in international equity markets." Journal of Financial and Quantitative Analysis 43, no. 03 (2008): 787-815.

Patton, Andrew J. "Modelling asymmetric exchange rate dependence." International economic review 47, no. 2 (2006): 527-556.

Pickands III, James. "Statistical inference using extreme order statistics." The Annals of Statistics (1975): 119-131.

Pickands III, James. "Multivariate extreme value distributions." In Proceedings 43rd Session International Statistical Institute, vol. 2, pp. 859-878. 1981.

Poon, Ser-Huang, Michael Rockinger, and Jonathan Tawn. "Extreme value dependence in financial markets: Diagnostics, models, and financial implications." Review of financial studies 17, no. 2 (2004): 581-610.

Rockafellar, R. Tyrrell, and Stanislav Uryasev. "Conditional value-at-risk for general loss distributions." Journal of Banking & Finance 26, no. 7 (2002): 1443-1471.

Tang, Ke, and Wei Xiong. "Index investment and financialization of commodities". No. w16385. National Bureau of Economic Research, 2010.

Tawn, Jonathan A. "Modelling multivariate extreme value distributions." Biometrika 77, no. 2 (1990): 245-253.

Tiago de Oliveira, J. "Statistical Extremes-A Survey." Center of Applied Mathematics, Lisbon (1973).

Tse, Yiu K., and Albert K. C. Tsui. "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations." Journal of Business & Economic Statistics 20, no. 3 (2002): 351-362.

Yamai, Yasuhiro, and Toshinao Yoshiba. "Value-at-risk versus expected shortfall: A practical perspective." Journal of Banking & Finance 29, no. 4 (2005): 997-1015.

Figure 3.1 Historical Price of Energy Products from April 17, 1996 to May 5, 2014

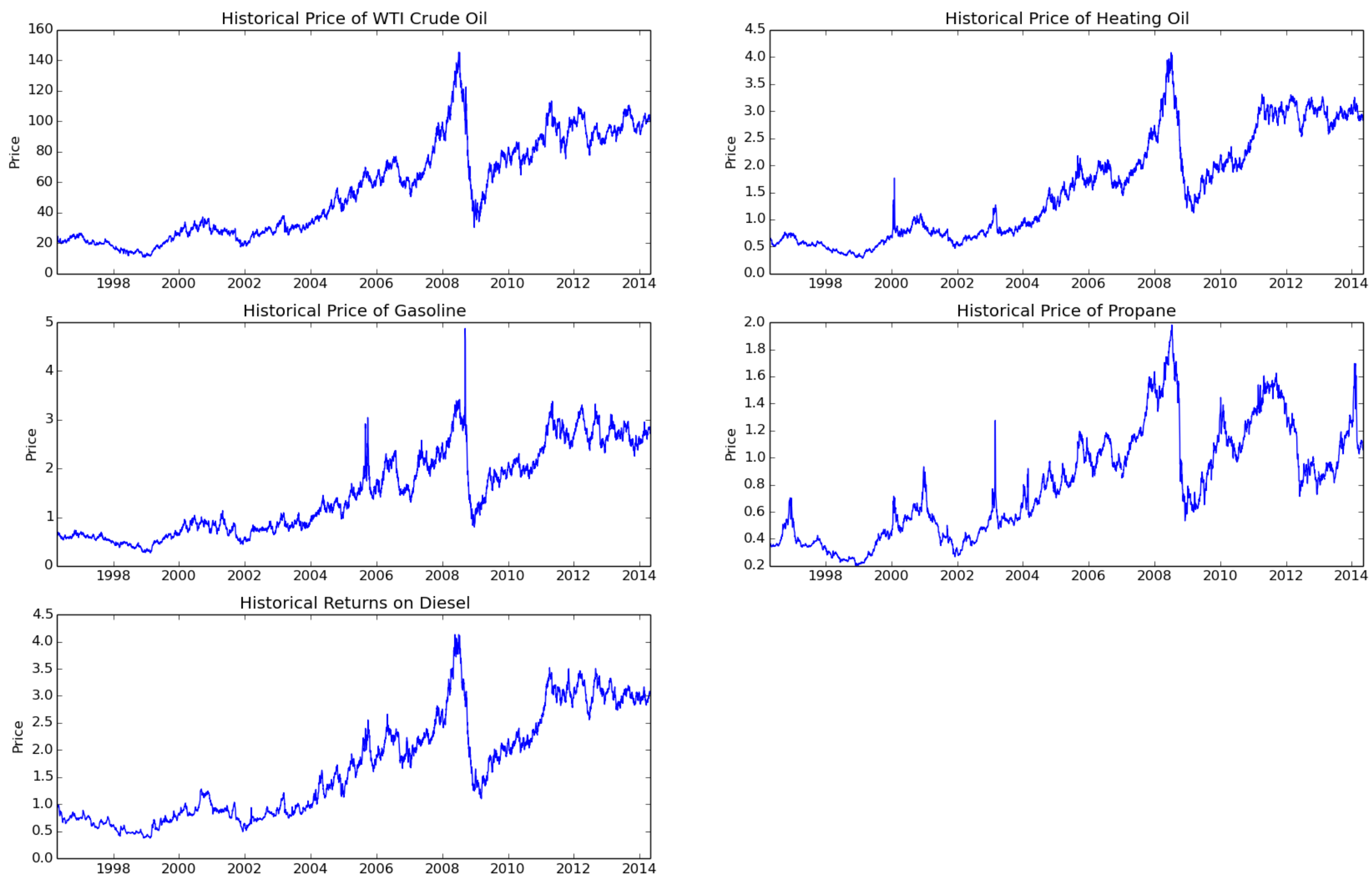


Figure 3.2 Historical Returns on Energy Products from April 18, 1996 to May 5, 2014

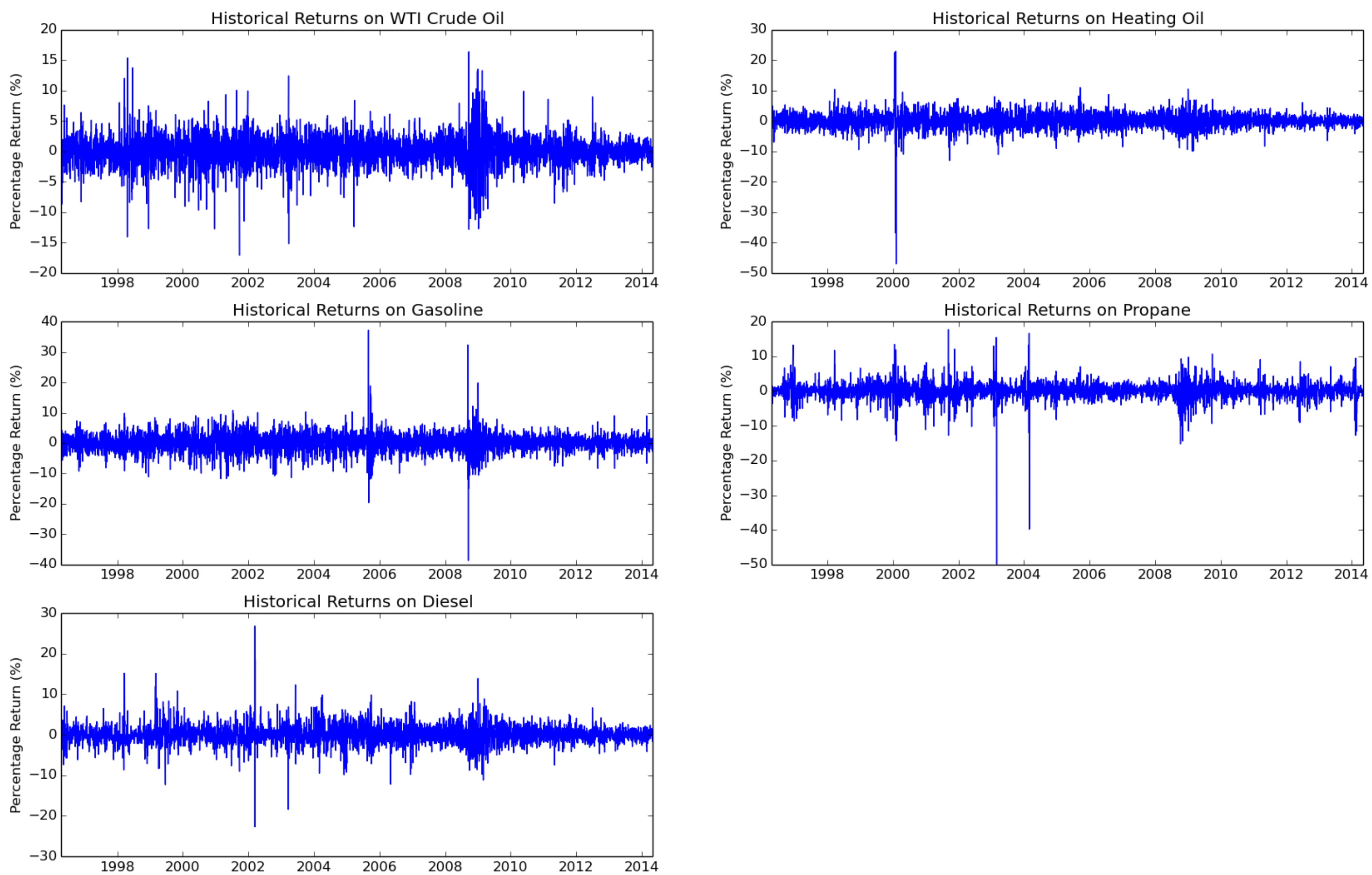


Table 3.1 Summary Statistics for Returns in Energy Prices

	Mean	Sdv	Min	Max	10% Quantile	25% Quantile	Median	75% Quantile	90% Quantile	Skewness	Kurtosis
Crude Oil	0.0308	2.4629	-16.414	17.092	-2.713	-1.3613	-0.0921	1.2370	2.7128	-0.1689	7.9238
Heating Oil	0.0326	2.5497	-22.954	47.012	-2.8259	-1.3498	0	1.2512	2.6511	-1.5460	43.5526
Gasoline	0.0332	3.0733	-37.954	38.676	-3.3959	-1.7011	0	1.552	3.4309	0.1186	17.8133
Propane	0.0246	2.4874	-17.674	49.913	-2.3894	-1.0615	0	0.9217	2.2771	-2.6480	58.1162
Diesel	0.0252	2.3251	-26.826	22.716	-0.2532	-1.1429	0	1.1217	2.3931	0.2332	14.1507

Note: Data ranges from April 18, 1996 to May 5, 2014 with 4532 observations in total.

Data downloaded from: http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm

Table 3.2 Parameter Estimates for the Left Tail Exceedances Modeled as Generalized Pareto Distribution

Threshold Quantile	Crude Oil		Heating Oil		Gasoline		Propane		Diesel	
	Tail Index	Scale Index	Tail Index	Scale Index	Tail Index	Scale Index	Tail Index	Scale Index	Tail Index	Scale Index
0.50%	0.1216	2.3277	0.6632	1.6299	0.4992	1.2968	0.3607	2.8036	0.3287	2.0238
	(1.3205)	(1.6899)	(0.0132)	(0.0841)	(0.0438)	(0.0152)	(0.1143)	(0.0045)	(0.0882)	(0.3028)
1%	0.048	2.3047	0.3689	1.9506	0.3946	1.463	0.3805	2.0942	0.2208	1.9684
	(0.2764)	(0.6941)	(0.0285)	(0.0797)	(0.0604)	(0.0786)	(0.0286)	(0.0993)	(0.0103)	(0.0699)
2%	0.1324	2.1463	0.5354	1.1805	0.1422	2.1882	0.3262	2.0438	0.2677	1.5906
	(0.1572)	(0.2778)	(0.0560)	(0.0614)	(0.0061)	(0.0411)	(0.0466)	(0.2263)	(0.0456)	(0.0959)
3%	0.0744	2.1448	0.4438	1.2371	0.0873	2.1754	0.3209	1.9583	0.2276	1.5663
	(0.0132)	(0.0401)	(0.0834)	(0.0507)	(0.0531)	(0.2002)	(0.0230)	(0.0446)	(0.0377)	(0.0608)
4%	0.1957	1.8051	0.3034	1.145	0.2212	1.702	0.3483	1.7515	0.2288	1.4981
	(0.0663)	(0.0661)	(0.0652)	(0.0233)	(0.0627)	(0.1374)	(0.0380)	(0.0364)	(0.0466)	(0.0470)
5%	0.2231	1.5407	0.3414	1.0847	0.2148	1.7166	0.3553	1.6009	0.2356	1.4519
	(0.0213)	(0.0313)	(0.0004)	(0.03650)	(0.0781)	(0.1268)	(0.0432)	(0.0557)	(0.0483)	(0.0029)
7.5%	0.2311	1.4629	0.1941	1.4099	0.2058	1.6722	0.3465	1.5597	0.174	1.4247
	(0.0108)	(0.0458)	(0.0340)	(0.1426)	(0.0729)	(0.0887)	(0.1142)	(0.1132)	(0.006)	(0.003)
10%	0.219	1.5015	0.1848	1.3107	0.2074	1.6081	0.2638	1.5266	0.0606	1.466
	(0.0762)	(0.0037)	(0.0090)	(0.0033)	(0.1172)	(0.2002)	(0.0002)	(0.0045)	(0.0488)	(0.0057)

Note: Standard Errors of parameter estimates in parenthesis.

Table 3.3 Parameter Estimates for the Right Tail Exceedances Modeled as Generalized Pareto Distribution

Threshold Quantile	Crude Oil		Heating Oil		Gasoline		Propane		Diesel	
	Tail Index	Scale Index	Tail Index	Scale Index	Tail Index	Scale Index	Tail Index	Scale Index	Tail Index	Scale Index
0.50%	0.0741	2.5886	1.1007	1.2243	0.8746	1.51	0.0703	3.2268	0.2879	2.7723
	(1.1839)	(1.8805)	(0.1350)	(0.2247)	(0.0462)	(0.2537)	(3.0127)	(5.2279)	(0.0732)	(0.3430)
1%	0.1743	2.316	0.7626	0.9156	0.5114	1.7944	0.1599	2.6409	0.5782	1.1422
	(0.6261)	(1.0987)	(0.0194)	(0.1005)	(0.0220)	(0.1256)	(0.2952)	(0.5679)	(0.0387)	(0.1092)
2%	0.3487	1.6899	0.4995	1.3765	0.4985	1.3741	0.2071	2.1325	0.3946	1.3532
	(0.2581)	(0.3196)	(0.1278)	(0.1457)	(0.0217)	(0.0165)	(0.1299)	(0.3098)	(0.0860)	(0.1825)
3%	0.2267	0.23	0.2342	1.4368	0.3904	1.3629	0.3548	1.7012	0.2327	1.7132
	(0.0448)	(0.0091)	(0.0736)	(0.1243)	(0.0302)	(0.0557)	(0.1559)	(0.1654)	(0.0215)	(0.0164)
4%	0.226	1.5053	0.2311	1.4106	0.2723	1.491	0.3483	1.5688	0.2302	1.5067
	(0.0091)	(0.0326)	(0.0845)	(0.2551)	(0.02530)	(0.0331)	(0.1117)	(0.1006)	(0.0018)	(0.0100)
5%	0.2323	1.4167	0.2134	1.3242	0.2334	1.5799	0.2511	1.531	0.2297	1.5038
	(0.0256)	(0.0475)	(0.0556)	(0.1085)	(0.0020)	(0.0042)	(0.0467)	(0.0755)	(0.0388)	(0.0559)
7.5%	0.1801	1.3945	0.2635	1.1219	0.2471	1.5459	0.2319	1.4197	0.2268	1.4687
	(0.0848)	(0.2114)	(0.0002)	(0.0069)	(0.0594)	(0.0392)	(0.0633)	(0.1201)	(0.0055)	(0.1253)
10%	0.0727	1.3928	0.1509	1.3656	0.2283	1.4941	0.2322	1.4335	0.1691	1.4084
	(0.0823)	(0.0970)	(0.0618)	(0.2497)	(0.0475)	(0.0300)	(0.0650)	(0.2522)	(0.0395)	(0.1139)

Note: Standard Errors of parameter estimates in parenthesis.

Table 3.4 Parameter Estimates for Crude Oil/Heating Oil Extreme Value Distributions

Threshold Quantile	Tail Index 1	Scale Index 1	Tail Index 2	Scale Index 2	Dependence Index
Negative Tail					
1%	0.0302	2.2666	0.308	1.89	0.568
	(0.0039)	(0.674)	(0.0157)	(0.1379)	(1.4431)
2%	0.0821	2.08	0.4692	1.1143	0.5471
	(0.0048)	(0.5189)	(0.0257)	(0.2599)	(1.4728)
3%	0.046	2.0868	0.3494	1.1422	0.5613
	(0.0006)	(0.6015)	(0.1539)	(0.3772)	(1.4386)
4%	0.1211	0.7827	0.2066	0.7453	0.3586
	(0.0624)	(2.1893)	(0.0162)	(0.9989)	(1.2694)
5%	0.1383	1.4747	0.211	0.6744	0.3603
	(0.07)	(2.2724)	(0.0592)	(1.4903)	(0.9090)
7.50%	0.149	1.4187	0.1212	0.9373	0.5127
	(0.0744)	(1.4647)	(0.0501)	(1.1567)	(0.9761)
10%	0.1418	1.0582	0.1168	1.2518	0.3883
	(0.1419)	(1.0935)	(0.929)	(0.5438)	(3.0414)
Positive Tail					
1%	0.1099	2.097	0.5419	0.7804	0.5349
	(0.0168)	(0.0785)	(0.0971)	(0.3136)	(1.4453)
2%	0.2174	0.5155	0.3251	1.202	0.493
	(0.0135)	(0.3930)	(0.0338)	(0.5194)	(1.4681)
3%	0.3584	0.1421	0.1464	1.1966	0.6382
	(0.0033)	(0.0018)	(0.0036)	(0.0562)	(0.0269)
4%	0.1421	1.4838	0.145	1.0372	0.3591
	(0.0064)	(0.8575)	(0.0165)	(0.7812)	(1.2720)
5%	0.1437	0.897	0.1319	1.1147	0.3616
	(0.0527)	(0.5066)	(0.0485)	(0.071)	(1.4802)
7.50%	0.1115	1.3961	0.1629	0.7916	0.5371
	(0.0483)	(1.2634)	(0.0775)	(0.5317)	(1.4264)
10%	0.045	1.2998	0.0933	1.3044	0.54
	(0.0152)	(0.0346)	(0.052)	(1.2126)	(0.1456)

Note: Standard Errors of parameter estimates in parenthesis.

Table 3.5 Parameter Estimates for Crude Oil/Gasoline Extreme Value Distributions

Threshold Quantile	Tail Index 1	Scale Index 1	Tail Index 2	Scale Index 2	Dependence Index
Negative Tail					
1%	0.0297	2.2734	0.3518	1.42	0.6187
	(0.0027)	(0.56)	(0.0227)	(0.2924)	(1.4425)
2%	0.0821	2.08	0.0925	2.122	0.5982
	(0.0036)	(0.4402)	(0.0047)	(0.4987)	(1.47)
3%	0.046	2.0888	0.0592	2.1194	0.6125
	(0.0065)	(0.1793)	(0.0147)	(0.5205)	(1.5138)
4%	0.121	1.658	0.1475	1.5549	0.5821
	(0.0083)	(0.4182)	(0.0101)	(0.1858)	(1.4877)
5%	0.1379	1.0477	0.1713	1.6332	0.4
	(0.0232)	(1.2638)	(0.068)	(0.2193)	(1.2078)
7.50%	0.143	1.0153	0.1359	1.5679	0.5744
	(0.042)	(0.6163)	(0.0582)	(1.1511)	(1.4713)
10%	0.1355	0.1824	0.1314	1.3882	0.5783
	(0.0642)	(0.5677)	(0.0536)	(1.4629)	(1.8888)
Positive Tail					
1%	0.1078	2.2284	0.3839	1.6643	0.5871
	(0.0154)	(0.7456)	(0.0446)	(0.4266)	(1.4627)
2%	0.2166	1.4186	0.3145	0.1043	0.5314
	(0.01)	(0.437)	(0.036)	(0.4705)	(1.4593)
3%	0.3667	0.1422	0.2446	1.1718	0.6866
	(0.004)	(0.0013)	(0.0099)	(0.0495)	(0.0316)
4%	0.1397	1.2578	0.1683	1.2669	0.5744
	(0.0146)	(0.4166)	(0.0193)	(0.0805)	(1.4323)
5%	0.1436	1.1597	0.1459	1.4291	0.573
	(0.023)	(0.3958)	(0.0343)	(0.3364)	(1.4773)
7.50%	0.1113	0.0924	0.1528	1.4987	0.4979
	(0.0253)	(0.7668)	(0.0797)	(0.213)	(1.2291)
10%	0.0449	1.3382	0.1412	1.5111	0.4886
	(0.0088)	(0.0834)	(0.0387)	(0.5954)	(0.3367)

Note: Standard Errors of parameter estimates in parenthesis.

Table 3.6 Parameter Estimates for Crude Oil/Propane Extreme Value Distributions

Threshold Quantile	Tail Index 1	Scale Index 1	Tail Index 2	Scale Index 2	Dependence Index
Negative Tail					
1%	0.0298	2.2807	0.3565	2.0702	0.8261
	(0.0017)	(0.3364)	(0.0146)	(0.2299)	(1.4576)
2%	0.0819	2.081	0.227	1.9452	0.805
	(0.0043)	(0.3093)	(0.0089)	(0.291)	(1.4873)
3%	0.0461	2.1076	0.2837	1.9211	0.8195
	(0.0014)	(0.3914)	(0.022)	(0.1174)	(1.4689)
4%	0.1217	1.7072	0.2504	1.6536	0.7891
	(0.0033)	(0.3970)	(0.0014)	(0.2697)	(0.5165)
5%	0.1427	0.9966	0.2515	1.5836	0.5179
	(0.0115)	(0.8962)	(0.0481)	(0.2779)	(1.1961)
7.50%	0.2044	1.2891	0.2261	1.5808	0.5179
	(0.0265)	(1.1377)	(0.0228)	(0.4687)	(1.3925)
10%	0.1353	1.0289	0.1631	1.5419	0.7831
	(0.0395)	(0.9746)	(0.0439)	(0.784)	(1.3302)
Positive Tail					
1%	0.1135	2.236	0.0999	2.5609	0.7981
	(0.0088)	(0.4012)	(0.0051)	(0.2707)	(1.4672)
2%	0.2218	1.6422	0.1291	2.0818	0.7848
	(0.0109)	(0.2814)	(0.0072)	(0.3579)	(1.4879)
3%	0.3667	0.1623	0.2192	1.5284	0.9597
	(0.013)	(0.0094)	(0.0145)	(0.0922)	(0.0201)
4%	0.1397	1.2689	0.2159	0.3558	0.781
	(0.0009)	(0.3471)	(0.0083)	(0.2052)	(1.4705)
5%	0.1436	0.9095	0.1857	1.4864	0.5223
	(0.017)	(0.8286)	(0.0418)	(0.7127)	(1.1081)
7.50%	0.1153	1.1284	0.1436	1.365	0.7781
	(0.0162)	(0.3616)	(0.0361)	(0.3572)	(1.5054)
10%	0.0449	0.8623	0.1436	1.0164	0.584
	(0.0067)	(0.9799)	(0.0453)	(0.4097)	(0.6445)

Note: Standard Errors of parameter estimates in parenthesis.

Table 3.7 Parameter Estimates for Crude Oil/Diesel Extreme Value Distributions

Threshold Quantile	Tail Index 1	Scale Index 1	Tail Index 2	Scale Index 2	Dependence Index
Negative Tail					
1%	0.0303	2.2671	0.1575	1.9088	0.7859
	(0.0018)	(0.3348)	(0.0023)	(0.1432)	(1.4579)
2%	0.0821	2.0801	0.2015	1.5244	0.7651
	(0.0028)	(0.3061)	(0.0032)	(0.2298)	(1.4839)
3%	0.0461	2.1076	0.1904	1.5291	0.7796
	(0.0005)	(0.3544)	(0.0214)	(0.0136)	(1.4718)
4%	0.121	1.64313	0.1548	1.3159	0.7489
	(0.007)	(0.218)	(0.0043)	(0.2751)	(1.4985)
5%	0.1379	1.1893	0.1468	1.3929	0.7421
	(0.0095)	(0.2742)	(0.0151)	(0.0361)	(1.5046)
7.50%	0.1429	1.2443	0.1076	1.2928	0.7543
	(0.0407)	(0.3991)	(0.0179)	(0.07)	(1.7534)
10%	0.1355	1.5039	0.0375	1.2879	0.5545
	(0.0285)	(0.4381)	(0.0066)	(0.1094)	(0.2388)
Positive Tail					
1%	0.1084	2.2961	0.5566	1.0784	0.7531
	(0.0079)	(0.3969)	(0.0148)	(0.1878)	(1.4655)
2%	0.2213	1.4298	0.2443	1.1787	0.7102
	(0.0077)	(0.2837)	(0.0195)	(0.3316)	(1.471)
3%	0.3664	0.1421	0.1805	1.373	0.9506
	(0.0351)	(0.0067)	(0.1199)	(0.877)	(0.2665)
4%	0.1397	1.215	0.1448	1.3997	0.7413
	(0.0041)	(0.2997)	(0.0112)	(0.1423)	(1.4676)
5%	0.1456	1.0653	0.142	1.2978	0.7391
	(0.0097)	(0.2414)	(0.0188)	(0.0029)	(1.4911)
7.50%	0.1116	1.211	0.142	1.2694	0.4934
	(0.0109)	(0.1842)	(0.0651)	(0.3413)	(1.6532)
10%	0.0462	1.3716	0.1045	1.4071	0.6671
	(0.0078)	(0.1068)	(0.0392)	(0.7607)	(0.6336)

Note: Standard Errors of parameter estimates in parenthesis.

Figure 3.3 Estimates of Extreme Correlation between Crude Oil and Heating Oil

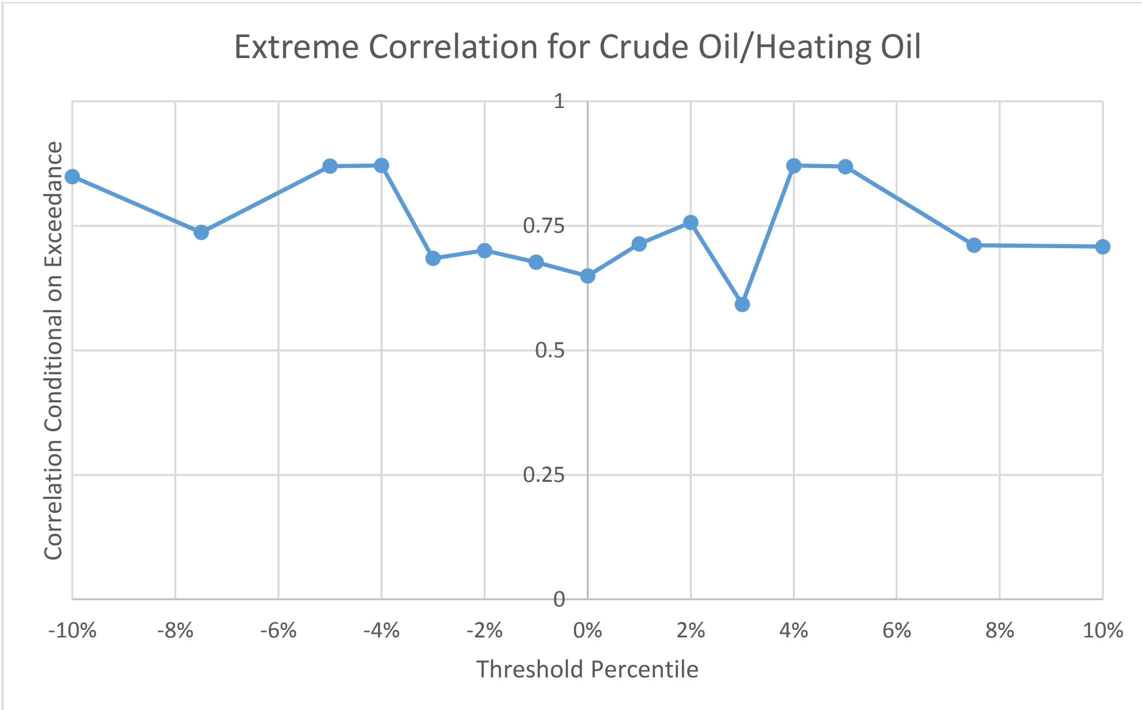


Figure 3.4 Estimates of Extreme Correlation between Crude Oil and Gasoline

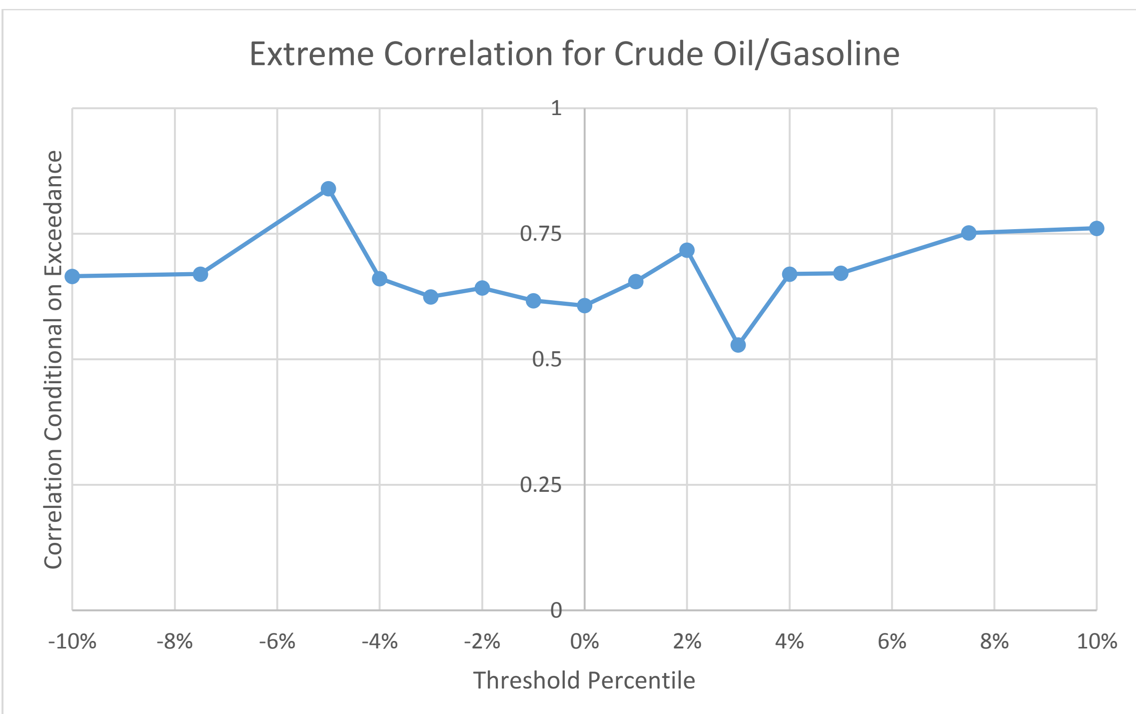


Figure 3.5 Estimates of Extreme Correlation between Crude Oil and Propane

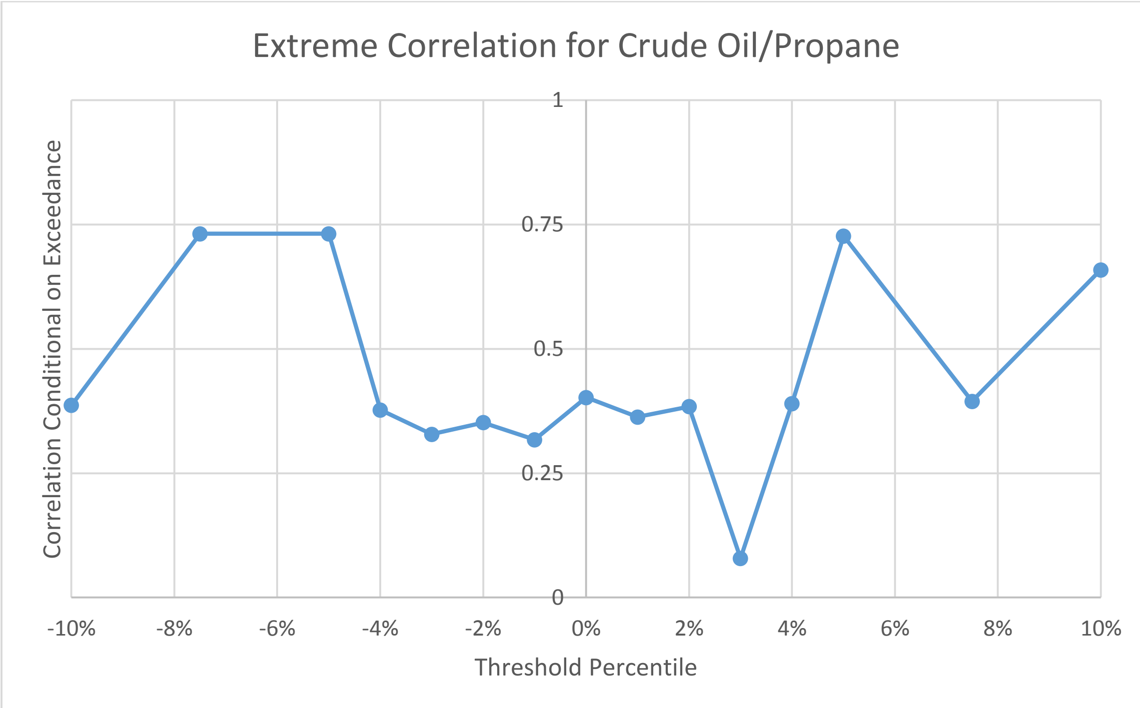
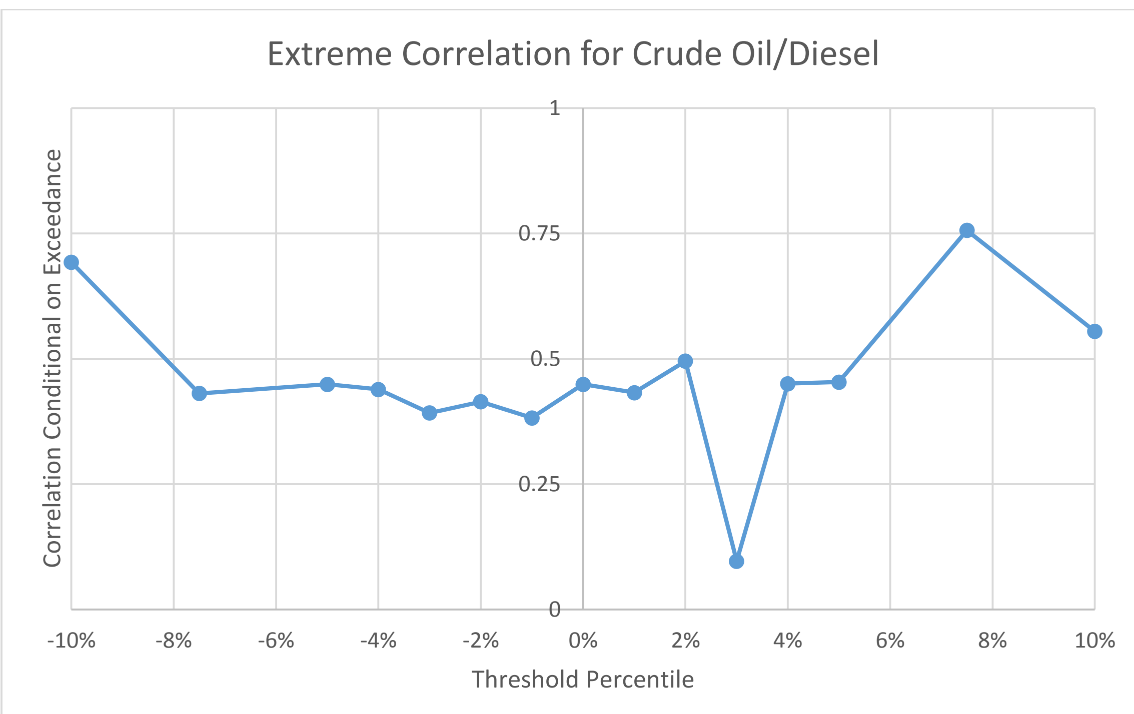


Figure 3.6 Estimates of Extreme Correlation between Crude Oil and Diesel



VITA

NAME OF AUTHOR: Chong Li

PLACE OF BIRTH: Huanggang, P.R. China

DATE OF BIRTH: September 9

EDUCATION:

M.A., Finance, Xiamen University, P.R. China, 2009

B.S., Physics, Huangzhong University of Science and Technology, P.R. China, 2006

PROFESSIONAL EXPERIENCE:

Teaching Assistant, Economics Department, Syracuse University, 2009-2010, 2011-2013

Research Assistant, Center for Policy Research, Maxwell School, Syracuse University, 2010-2011

Research Assistant, Institute for Veterans and Military Families (IVMF), Syracuse University, 2014