2019

ECONOMIC IMPACT OF ETHANOL BIOREFINERIES IN THE U.S. MIDWEST FROM 2001 TO 2015: A QUASI-EXPERIMENTAL APPROACH

Scott W. Hall
University of Kentucky, scottwhall1@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0001-5080-9586
Digital Object Identifier: https://doi.org/10.13023/etd.2019.246

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Scott W. Hall, Student

Dr. David Freshwater, Major Professor

Dr. Carl Dillon, Director of Graduate Studies

ECONOMIC IMPACT OF ETHANOL BIOREFINERIES IN THE U.S. MIDWEST FROM 2001 TO 2015: A QUASI-EXPERIMENTAL APPROACH

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture, Food and Environment at the University of Kentucky

By

Scott Wayne Hall
Pleasureville, Kentucky

Director: Dr. David Freshwater, Professor of Agricultural Economics
Lexington, Kentucky
2019

https://orcid.org/0000-0001-5080-9586

ABSTRACT OF DISSERTATION

ECONOMIC IMPACT OF ETHANOL BIOREFINERIES IN THE U.S. MIDWEST FROM 2001 TO 2015: A QUASI-EXPERIMENTAL APPROACH

The objective of this dissertation is to analyze the economic impact of newly operational ethanol biorefineries on rural counties in the U.S. Midwest region for the period 2001 to 2015 using a quasi-experimental approach. Rapid growth in the ethanol industry expanded the number of ethanol plants located in the U.S. Midwest from 54 in 2001 to 173 in 2015. Out of the counties with 119 new ethanol biorefineries, 97 counties met the general treatment criteria defined in this dissertation, but only 56 of those counties qualified for the rural treatment criteria. Counties with ethanol biorefineries that qualified for treatment were organized into a treated group based on county level data. Six counterfactual control groups (or control counties without ethanol biorefineries) were contemporaneously matched to the treated counties based on the Mahalanobis distance metric evaluated on a set of 29 selection variables. Matching occurred on two levels. In the first level, matching was performed both for the in-state level and over the entire Midwest region. In the second level, three criteria were used to select the final control groups: Mahalanobis distance metric best match, population best match, and rural-urban continuum codes (RUCC) best match.

Economic impact is evaluated based on the growth rate in real per capita earnings for the treated group over a period from one to five years after treatment relative to the control group. A difference-in-differences (DID) model is used to assess the significance of results where the dependent variable is the natural log of real per capita earnings and a set of control variables is used to capture state fixed effects, time fixed effects and spillover effects. Empirical results evaluated against a representative Midwest control group and over six regression models adjusting for various fixed effects produced, on average, one-sided significant results for average treatment on the treated (ATOT) with a (min, max) range of growth rates as (5.53%-7.63%), (10.0%-12.0%), (14.7%-19.6%), (14.5%-18.3%), and (13.3%-18.9%) from one to five years after treatment, respectively.

The minimum value of these estimates can be represented as an uncorrected average annual growth rate as 2.75%, 3.33%, 3.68%, 2.90%, and 2.22% over the respective period from one to five years after treatment. Employment levels for the treated group increased on average by 211 at the county level five years after treatment. A comparative Midwest control group lost, on average, 169 jobs over the five year period after treatment. A treated county employment multiplier calculated using the direct, indirect and induced employment impacts varied from 1.46 during the year of treatment to 7.6 five years after treatment relative to the control group. Five years after treatment, the treated group employment rate gradually increased, on average, by 2.2% which was better than either of the two counterfactual control groups used in this comparison.

Overall, the analysis presented in this dissertation does show statistically significant positive economic impacts, on average, for rural U.S. Midwest counties with newly operational ethanol biorefineries relative to control counties without an ethanol biorefinery. These results demonstrate that the Renewable Fuel Standard (RFS) contributed to positive rural economic development impacts in treated counties with the possibility of spillover effects positively affecting contiguous counties.

KEYWORDS: Rural Economic Development, Renewable Fuel Standard, Energy, Bioenergy, Ethanol Biorefineries, U.S. Midwest

Scott W. Hall

Student Signature

June 10, 2019

Date

ECONOMIC IMPACT OF ETHANOL BIOREFINERIES IN THE U.S. MIDWEST
FROM 2001 TO 2015: A QUASI-EXPERIMENTAL APPROACH

By

Scott W. Hall

Dr. David Freshwater
Director of Dissertation

Dr. Carl Dillon
Director of Graduate Studies

June 10, 2019
Date

To the memory of my mother, Mary Carole Hall, who passed away during this period of my life as I worked on my Ph.D. When I was young, she would take my brother and me on road trips which spanned across America from Washington D.C. to the Rocky Mountains and many places in between. She instilled a spirit of adventure within me and a desire to experience new things, to see new places, and to learn more about the world. I am still on that journey.

ACKNOWLEDGEMENTS

I am honored to express my deepest gratitude to my advisor and committee chair, Dr. David Freshwater, for his continuous support and encouragement during this research endeavor. Our conversations have been fruitful with ideas and have immensely contributed to a much improved finished product. I extend my appreciation to Dr. Carl Dillon, Dr. Tyler Mark, and Dr. William Hoyt by honoring me with their presence on my committee. Their experience and knowledge directed me down paths of investigation which allowed me to explore new ideas.

My appreciation is extended to many people within the Department of Agricultural Economics. I would like to thank Dr. Michael R. Reed who was the Director of Graduate Studies when I entered the program. He always provided a positive attitude which made some tough moments much easier to experience. I've had many long conversations with Dr. Carl Dillon on many subjects and shared many laughs. Janene Toelle seemed to always do the right things to keep me moving forward in the program. Rita Parsons and Karen Pulliam were always available to help with clerical or technical issues. Additionally, I want to thank the AEC family for the financial support provided by the Dr. John C. Redman Memorial Scholarship.

It would have been impossible for me to complete a Doctor of Philosophy in Agricultural Economics without the support of my friends and family. I appreciate Kevin (Oz) and Pem for their friendship and the part-time work on their house and farm which provided additional financial support for this endeavor. To my friends, Jimmy, Meredith,

Carl and Betty, I value our friendship and enjoyed our conversations during this period. To my brother, Seth, and his family, David, Dane and Mary Elizabeth, I appreciate the occasional family gatherings and special events which provided some time for relaxation and reflection. To my father, Wayne Ray Hall, from whom, I learned the value of hard work.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ADDITIONAL FILES x

CHAPTER 1. INTRODUCTION 1

1.1 Overview 1

1.2 U.S. Fuel Ethanol Industry Overview 2

1.2.1 Background on Fuel Ethanol Use 2

1.2.2 Fuel Ethanol Production Process 4

1.2.3 Corn Feedstock 6

1.2.4 Ethanol Biorefinery Expansion in U.S. Midwest 8

1.2.5 Policy Driven Fuel Ethanol Industry Expansion 9

1.2.6 Local Economic Impacts of Ethanol Biorefineries 13

1.3 Research Objective 15

1.4 Research Hypothesis 16

1.5 Research Structure 17

1.6 Tables in Chapter 1 18

1.7 Figures in Chapter 1 19

CHAPTER 2. LITERATURE REVIEW 24

2.1 Regional Economic Analysis and the Quasi-Experimental Approach 24

2.2 Quasi-Experimental Applications (Spatial Linkages) 25

2.3 Spatio-Temporal Analysis 26

2.4 Ethanol Industry Specific Literature 27

CHAPTER 3. QUASI-EXPERIMENTAL RESEARCH METHODOLOGY 29

3.1 The Rubin Causal Model 29

3.2 Quasi-Experimental Methodology 30

3.2.1 Difference-in-Differences (DID) Model 31

3.2.1.1 Identification Strategy 34

3.2.1.2 Control Variables 34

3.2.1.3 Rural Population 35
3.2.1.4 Population versus Sample Statistics (Analysis Assumption) 36
3.2.2 Treatment Definition and Treatment Event 37
3.2.3 Treated Spatial Units 38
3.2.4 Control Spatial Units 38
3.2.5 Matching Process for Control Units 38
3.2.6 Dependent Variable 40
3.2.7 Parameter of Interest: Average Treatment on the Treated (ATOT) 41
3.2.8 Spillover Effects - General Explanation 41
3.2.9 Spillover Effects – Setting the Indicator Variables 44
3.3 Mahalanobis Distance Metric (Ranking of Control Units) 46
3.4 Figures in Chapter 3 48
CHAPTER 4. DATA DESCRIPTION 51
4.1 U.S. Midwest Region 51
4.2 Spatial Units 52
4.3 Time Period Selection 52
4.4 Ethanol Biorefinery Operational Status and Location 53
4.5 Treated and Control County Data 53
4.6 Dependent Variable 54
4.7 Selection Variables 54
4.8 Missing Data 55
4.8.1 Continuity 56
4.8.2 High Uncertainty: Zero-out 56
4.8.3 Phase-in or Phase-out 56
4.9 Data Movies 57
4.10 Data Sources and Links 58
4.11 Tables in Chapter 4 60
4.12 Figures in Chapter 4 81
CHAPTER 5. EMPIRICAL RESULTS AND DISCUSSION 82
5.1 Research Hypothesis Results and Discussion 82
5.1.1 Parallel Trend Assumption and Treatment Response 82
5.1.2 Empirical Results 84
5.1.3 Residual Analysis on the Empirical Models 87
5.1.4 Robustness Check (ATOT Consistency over Control Groups) 87
5.1.5 Treated County Year-over-Year Growth 90

5.1.6 Employment Impacts ... 91
5.1.7 Employment Multiplier ... 94
5.2 Exploratory Analysis and Discussion ... 94
5.2.1 Revisit Research Hypothesis with All Eligible Treated Counties ... 94
5.2.2 Treated County Population versus Biorefinery Initial Capacity ... 95
5.3 Threats to Internal Validity ... 97
5.3.1 SUTVA I – No Interference ... 97
5.3.2 SUTVA II – No Hidden Variations of Treatments ... 99
5.3.3 Selection on Observables into Treatment ... 99
5.3.4 Population Statistics versus Sample Statistics (Assumption) ... 100
5.4 Tables in Chapter 5 ... 102
5.5 Figures in Chapter 5 ... 117
CHAPTER 6. CONCLUSIONS AND IDEAS FOR FUTURE RESEARCH ... 139
6.1 Conclusions ... 139
6.2 Ideas for Future Research ... 142
6.2.1 Explore Employment Growth in State Control Counties ... 142
6.2.2 Explore Large Capacity Ethanol Biorefineries in Rural Counties ... 142
6.2.3 Longer Term Study ... 142
6.2.4 Apply Panel Data Techniques ... 143
6.2.5 Synthetic Control Groups ... 143
6.2.6 Coarsened Exact Matching ... 143
6.2.7 Spatial Difference in Differences ... 144
APPENDICES ... 145
A 1. MAHALANOBIS DISTANCE METRIC MIDWEST ... 145
A 2. MAHALANOBIS DISTANCE METRIC IN-STATE ... 149
BIBLIOGRAPHY ... 153
VITA ... 160

LIST OF TABLES

Table 1.1 Ethanol Dry Mill: Input Feedstock and Output Co-products (RFS, 2019)..... 18
Table 4.1 U.S. Midwest States: Number of Counties, Corn Arces, & Biorefineries....... 60
Table 4.2 Midwest Best Match Treated-Control County Pairs.. 61
Table 4.3 Midwest Population Match Treated-Control County Pairs 62
Table 4.4 Midwest RUCC Match Treated-Control County Pairs..................................... 63
Table 4.5 State Best Match Treated-Control County Pairs.. 64
Table 4.6 State Population Match Treated-Control County Pairs 65
Table 4.7 State RUCC Match Treated-Control County Pairs.. 66
Table 4.8 Midwest Best Match Treated-Control County Pairs, LT25k............................ 67
Table 4.9 Midwest Population Match Treated-Control County Pairs, LT25k 68
Table 4.10 Midwest RUCC Match Treated-Control County Pairs, LT25k...................... 69
Table 4.11 State Best Match Treated-Control County Pairs, LT25k............................... 70
Table 4.12 State Population Match Treated-Control County Pairs, LT25k 71
Table 4.13 State RUCC Match Treated-Control County Pairs, LT25k............................ 72
Table 4.14 Selection Variable Categories... 73
Table 4.15 Selection Variables List (Ethanol Biorefinery, 2001) 74
Table 4.16 Data Description 75
Table 4.17 Summary Statistics (Selection Variables & Dependent Variable) 76
Table 4.18 Summary Statistics (DID Model Dependent Variable) 77
Table 4.19 Rural-Urban Continuum Codes (USDA-ERS)... 78
Table 4.20 List of Data Movies 79
Table 4.21 Data Sources and Internet Links... 80
Table 5.1 Ln(RPCE) One Year after Treatment, Treated Population < 25,000 102
Table 5.2 Ln(RPCE) Two Years after Treatment, Treated Population < 25,000........... 103
Table 5.3 Ln(RPCE) Three Years after Treatment, Treated Population < 25,000......... 104
Table 5.4 Ln(RPCE) Four Years after Treatment, Treated Population < 25,000........... 105
Table 5.5 Ln(RPCE) Five Years after Treatment, Treated Population < 25,000 106
Table 5.6 Ln(RPCE1) Robustness: Treated with Populations < 25,000......................... 107
Table 5.7 Ln(RPCE2) Robustness: Treated with Populations < 25,000......................... 108
Table 5.8 Ln(RPCE3) Robustness: Treated with Populations < 25,000......................... 109
Table 5.9 Ln(RPCE4) Robustness: Treated with Populations < 25,000......................... 110
Table 5.10 Ln(RPCE5) Robustness: Treated with Populations < 25,000 111
Table 5.11 Corrected Growth Rates Year over Year.. 112
Table 5.12 Employment Treated-Control Groups Means (Midwest)............................. 113
Table 5.13 Employment Treated-Control Groups Means (State)................................... 114
Table 5.14 Employment Multiplier (Midwest).. 115
Table 5.15 Ln(RPCE3) Robustness: All Treated Counties 116

LIST OF FIGURES

Figure 1.1 Ethanol Dry-mill Production Process Flow 19
Figure 1.2 Midwest Ethanol Biorefineries in 2001 20
Figure 1.3 Midwest Ethanol Biorefineries in 2015 21
Figure 1.4 RFS2 Volumetric Requirements 22
Figure 1.5 U.S. Fuel Ethanol Production 23
Figure 3.1 Diagram of Pre-Treatment Observations and Post-Treatment Observations .. 48
Figure 3.2 Diagram of DID Regression Analysis for One Year after Treatment 48
Figure 3.3 Diagram of DID Regression Analysis for Three Years after Treatment 48
Figure 3.4 Difference-in-Differences Estimation 49
Figure 3.5 Mahalanobis Control County Ranking by Year of Treatment 50
Figure 4.1 Flint Hills Resources, Iowa Falls, Iowa (105 mgy) 81
Figure 5.1 Parallel Tread and Treatment Response (MW-Best Match, Pop<25k) 117
Figure 5.2 Parallel Tread and Treatment Response (MW-Pop Match, Pop<25k) 118
Figure 5.3 Parallel Tread and Treatment Response (MW-RUCC, Pop<25k) 119
Figure 5.4 Parallel Tread and Treatment Response (State-Best Match, Pop<25k) 120
Figure 5.5 Parallel Tread and Treatment Response (State-Pop Match, Pop<25k) 121
Figure 5.6 Parallel Tread and Treatment Response (State-RUCC, Pop<25k) 122
Figure 5.7 Max and Min Growth Rates (Period after Treatment) 123
Figure 5.8 Max and Min Average Annual Growth Rates (Period after Treatment) 124
Figure 5.9 Normal QQ-Plot for lrpce1 DID Model Residuals 125
Figure 5.10 Normal QQ-Plot for lrpce2 DID Model Residuals 126
Figure 5.11 Normal QQ-Plot for lrpce3 DID Model Residuals 127
Figure 5.12 Normal QQ-Plot for lrpce4 DID Model Residuals 128
Figure 5.13 Normal QQ-Plot for lrpce5 DID Model Residuals 129
Figure 5.14 Employment Parallel Trend & Response (MW-Pop Match, Pop<25k) 130
Figure 5.15 Employment Parallel Trend & Response (ST-Pop Match, Pop<25k) 131
Figure 5.16 Parallel Trend and Treatment Response (Midwest-Best Match MD) 132
Figure 5.17 Parallel Trend and Treatment Response (Midwest-Population Match) 133
Figure 5.18 Parallel Trend and Treatment Response (Midwest-RUCC) 134
Figure 5.19 Parallel Trend and Treatment Response (State-Best Match MD) 135
Figure 5.20 Parallel Trend and Treatment Response (State-Population Match) 136
Figure 5.21 Parallel Trend and Treatment Response (State-RUCC) 137
Figure 5.22 Population versus Initial Capacity Diagrams 138

LIST OF ADDITIONAL FILES

Supplemental Data Midwest Corn Choropleth Movie .... MP4 3,033 KB
Supplemental Data Midwest Corn Yield Choropleth Movie .... MP4 4,943 KB
Supplemental Data Midwest Ethanol Biorefinery Locations .... MP4 1,770 KB
Supplemental Data Midwest Treated-Control Match RUCC .... MP4 2,754 KB
Supplemental Data State Treated-Control Match (Best Match) .... MP4 2,628 KB
Supplemental Data Midwest Soybean Choropleth Movie .... MP4 2,868 KB

CHAPTER 1. INTRODUCTION

1.1 Overview

The rapid expansion of the U.S. fuel ethanol industry from 2001 to 2015 created an opportunity to examine whether there were significant localized economic impacts associated with rural investment in ethanol biorefineries and their continuous operations in the U.S. Midwest. This research proposes that a quasi-experimental event took place with the recent expansion of the fuel ethanol industry and that local communities where ethanol biorefineries were located economically benefitted more so than communities without an ethanol biorefinery.

This chapter will proceed with an overview of the fuel ethanol industry. First, a brief overview of fuel ethanol use as an energy source is presented. Second, an overview of an ethanol plant's operations is presented by describing the fuel ethanol production process. Third, since corn is the primary feedstock used in ethanol production, some highlights of production technologies used to increase corn yields are discussed. Fourth, a review on ethanol industry expansion is presented. Fifth, insight into local economic impacts associate with the fuel ethanol industry is discussed. Finally, an overview of how policy has been a driving factor in the fuel ethanol industry's expansion.

This chapter will conclude with three sections which focus the research performed in this dissertation. First, the objective of the research is formalized. Second, the null and alternative hypotheses are formally stated. Finally, the overall structure of the research conducted in the remaining chapters of this dissertation is presented.

1.2 U.S. Fuel Ethanol Industry Overview

1.2.1 Background on Fuel Ethanol Use

Fuel ethanol has properties which allow it to function as both a complement and a substitute for gasoline. As a complement to gasoline, fuel ethanol is used as an additive which functions both as an oxygenate which enables more complete combustion of the fuel within an engine and as an octane enhancer to boost the octane rating of gasoline which prevents engine knock. Ethanol is classified as an oxygenate since it has one oxygen atom in its chemical structure (C_2H_5OH) which helps to facilitate more complete combustion and reduces carbon monoxide pollution. In the Clean Air Act Amendments of 1990, fuel ethanol was one of the oxygenates listed for use in the production of reformulated gasoline (RFG). The real benefit of oxygenates is the reduction in air pollutants as products of the combustion process, such as carbon monoxide (CO), nitrogen oxides (NO_X), hydrocarbons, and indirect secondary products such as ozone.

In the automotive industry, there have always been trade-offs between engine efficiency which requires higher compression ratios and the types of fuels that can be used which are capable of withstanding the higher compression ratios prior to self-ignition. This problem was solved in the early years of the automotive industry by using tetraethyl lead, a highly toxic substance, as an octane enhancer. In the 1970s, public health concerns over the use of tetraethyl lead increased public pressure on gasoline fuel refiners to produce unleaded fuels and on automotive manufacturers to produce vehicles that would run efficiently on unleaded fuels. Though the fuel ethanol industry was virtually non-existent in the 1970s, ethanol's octane rating of 113 made it a promising, non-toxic octane enhancer for gasoline. Currently, fuel ethanol is one of the most common octane boosters for gasoline. For example, a gasoline-ethanol mixture with 10% ethanol (E10) will raise gasoline's refined octane rating from about 85 to above 87 which is generally the minimum octane rating available for use in most motor vehicles. Thus, fuel ethanol provides a non-toxic solution to enhance engine efficiency.

Since fuel ethanol is an energy source, it is also a substitute for gasoline. In Brazil, Flex Fuel Vehicles (FFVs) can run on hydrous ethanol[1] (E100) or any mixture of gasoline and ethanol. One downside of ethanol is its energy content which is only about 67% of gasoline's energy content on a volumetric basis. Therefore, Brazilian consumers make fuel substitution trade-offs based on how the relative pricing of the fuels reflects the relative energy content. In the U.S., FFVs can operate with any gasoline-ethanol fuel mixture up to E85 (85% ethanol). In order to create a blended gasoline-ethanol fuel mix, the ethanol used must be anhydrous ethanol[2]. Too much water in the ethanol can cause chemical phase separation into its constituent parts (gasoline, ethanol and water) which can create engine performance issues. In the U.S., all fuel ethanol biorefineries ship denatured anhydrous ethanol to fuel blenders to insure that phase separation will not occur. Additionally, fuel ethanol's use as a substitute to gasoline is frequently used by the Renewable Fuels Association (RFA), an ethanol industry trade association, to tout historic oil import displacement through the use of fuel ethanol (RFA, 2019).

As an additional note on this topic, the terms "ethanol", "fuel ethanol", "anhydrous ethanol", and "conventional ethanol" are mostly used interchangeably in this dissertation. There can be slight differences in the meaning of the terms depending on

---

[1] Hydrous ethanol is distilled to its highest azeotropic purity of about 95% pure ethanol with water composing the remaining percentage of the mixture.

[2] Anhydrous ethanol has a water content of less than 1% by volume as specified by ASTM D4806.

context, but where appropriate, the proper terminology will be used for purposes of clarity.

1.2.2 Fuel Ethanol Production Process

Ethanol distillers, using corn as a feedstock, build their biorefineries to operate using either a dry mill or a wet mill production process. Approximately 90% of all fuel ethanol biorefineries are dry mills (RFA, 2019). A dry mill ethanol production process flow is shown in Figure 1.1. In the dry mill process, the biorefinery receives corn from local farmers which is ground into coarse flour and mixed with water to form a mash. Alpha-amylase enzymes are added into the mixture and the slurry is cooked at the proper temperature and pH level to facilitate the saccharification process which converts starches into long-chain sugars. After the first phase of saccharification is complete, gluco amylase enzymes are added to break the long-chain sugars into simple sugars which are more suitable for yeast consumption. The sugary solution is piped into fermentation tanks where yeast (saccharomyces cerevisiae) and yeast nutrients are added to ferment the mixture. For some biorefineries, carbon dioxide is a co-product from fermentation which is captured and marketed to beverage companies or used for the production of dry ice. Other biorefineries treat carbon dioxide as a by-product and release it into the atmosphere. Depending on how carbon dioxide is processed, it can greatly change the lifecycle greenhouse gas (GHG) emission profile of the resulting fuel ethanol.

After the fermentation process is complete, the resulting “Beer” with about 16% ethanol by volume is pumped into intermediate storage to facilitate a continuous flow of the Beer solution into the distillation columns. The Beer goes through a multistage distillation process which produces two main co-products: distillers grains and 190 proof ethanol. The distillers grains are further refined to extract water and corn oil. Extracted water is recycled back into the earlier stages of the process. Corn oil can be used to make biodiesel or added back into the grain solids for feed. The refined distillers grains are now wet distillers grains with solubles (WDGS) and can be immediately sold for feed. Also, the WDGS can be dried into dried distillers grains with solubles (DDGS) which has a longer shelf life for storage or can be immediately sold for feed.

The 190 proof ethanol is filtered through molecular sieves to extract most of the remaining water from the ethanol until it has less than 1% water content by volume. Then, the anhydrous ethanol is denatured by adding about 2% gasoline to make it unpalatable for human consumption. The finished denatured anhydrous ethanol (fuel ethanol) is shipped to blenders either by 8,000 to 10,000 gallon tanker trunks or by 30,000 gallon railcars. Average output yield from a dry mill process based on a bushel of corn as input is shown in Table 1.1. The ethanol industry has improved its yield efficiency over the past two decades to produce about 2.86 gallons of denatured fuel ethanol, 15.9 pounds of distillers grains (10% moisture), 0.75 pounds of corn distillers oil, and 16.5 pounds of carbon dioxide from a bushel of corn (RFA, 2019).

Since wet milling production is only used by about 10% of ethanol producers, the process will not be discussed in detail, but there are some notable differences with respect to the dry mill process that are worth discussion. In the wet milling process, capital costs are higher than in dry milling process. One advantage of the higher capital costs is that it enables greater flexibility in the mix of starch co-products produced, extraction of corn oil, and in the production of gluten feed. This dynamic mix of co-products allows wet-milling producers the short-term flexibility to optimize their profits by being more responsive to output market price fluctuations. One limiting factor is that wet mills cannot produce as much ethanol as dry mills per bushel of corn input. This occurs due to portions of the starch exiting some of the processes attached to other co-products (food starches or high fructose corn syrup).

1.2.3 Corn Feedstock

In the United States, corn is the primary feedstock used in most fuel ethanol production though many different types of starchy grains (barley, wheat, sorghum grain), sugar crops (sugar beets, sugarcane), and biomass (cellulosic and lignocellulosic) can be used to make ethanol. In most cases, starchy grains and sugar crops have greater value in other uses than to make ethanol. Biomass ethanol production is still cost prohibitive since the technological advances required to become cost competitive with corn feedstock have not yet come to fruition. Thus, corn is the most abundant starch crop available in the United States which can be efficiently utilized in the production of ethanol.

The Midwest is the ideal place to grow corn in the United States. It has rich soils, temperate climate and relatively flat terrain which facilitate large scale crop production and the use of precision agriculture equipment. Through the use of precision agriculture, genetically engineered seed corn and center pivot irrigation systems, corn yields have increased from an average 138 bushels per acre in 2001 to 168.4 bushels per acre in 2015 (USDA-NASS). In fact, Nebraska and Iowa average corn yields were 185 and 192 bushels per acre, respectively, in 2015 (USDA-NASS). Additionally, the Midwest region produced 88% (11.9 billion bushels) of the total U.S. corn production (13.6 billion bushels) in 2015 (USDA-NASS). Since ethanol production is a weight-losing production activity, industrial location theory states that corn-based ethanol plants should locate near their feedstock source in order to minimize their costs associated with transportation (Edwards, 2007). As of 2015, 87% of all U.S. operating ethanol biorefineries were located in the Midwest (RFA, 2016).

Over the past two decades, genetically engineered corn use has dramatically increased from virtually non-existent in 1996 to almost 90% market penetration in 2018 (USDA-ERS, 2018). The Bt (insect-resistant varieties) and HT (herbicide-tolerant varieties) traits were integrated into the genetic code of the corn seed to help corn producers increase productivity and reduce yield loss due to insects. More recently, genetically engineered corn products are becoming more focused on the production uses of the corn. For instance, Syngenta has introduced a genetically modified corn with higher alpha-amylase content called Enogen®. Several ethanol biorefinery operations either require their corn suppliers to grow the Enogen® corn crop as a condition of being a feedstock supplier or pay a higher premium for corn delivered to them with higher alpha-amylase content. Since ethanol producers spend several millions of dollars each year on amylase products[3] to assist in the saccharification process, this type of product is targeted at lowering the operational costs of the biorefinery.

As the corn-based fuel ethanol industry approaches the 15 billion gallon ethanol volumetric regulatory limit as defined under the Renewable Fuel Standard (RFS), the

[3] Information based on "Ethanol Profitability Spreadsheet" created by Agricultural Marketing Resource Center (AGMRC) at Iowa State University. Spreadsheet can be found at: http://www.extension.iastate.edu/agdm/energy/xls/d1-10ethanolprofitability.xlsx.

demand for corn feedstock will reach 5.357 billion bushels per year[4]. This is 39.4% of total U.S. corn production and 45% of corn production in the Midwest based on 2015 production levels. Corn is expected to be the primary feedstock for ethanol biorefineries for years to come. The biggest question facing corn producers in the relatively near future is whether the RFS will be renewed or allowed to expire in 2022.

1.2.4 Ethanol Biorefinery Expansion in U.S. Midwest

In 2001, there were 54 ethanol biorefineries operating in the U.S. Midwest as shown by the blue location dots in Figure 1.2. Ethanol plants were operating in 10 of the 12 Midwest states with North Dakota and Ohio being the exceptions. In the contiguous United States, there were 74 operational ethanol biorefineries using a variety of feedstocks. The largest capacity plants were located in the U.S. Midwest and used corn as a feedstock. Cooperatives accounted for 34 of the 54 plants in the Midwest with an average production capacity of about 28.7 million gallons of ethanol per year (mgy). The other 20 plants were not classified as cooperatives and had an average annual production capacity of about 31 million gallons which excluded the 950 million gallon operations of Archer Daniel Midlands (ADM) since those operations can range from 100 mgy to 350 mgy.

As of 2015, there were 173 ethanol biorefineries operating in the U.S. Midwest as shown in Figure 1.3. These 173 plants are about 87% of the 199 total ethanol biorefineries operating in the contiguous United States (RFA, 2016). Thus, ethanol biorefineries built since 2001 have tended to concentrate in the Midwest region primarily to be in close proximity to an abundant feedstock (corn). It is estimated that only 32 of

[4] A transformation of 2.8 gallons of ethanol per bushel of corn is assumed.

these 173 biorefineries are cooperatives or locally owned with an average annual production capacity of 62.5 mgy. The 141 Midwest plants with corporate structure other than locally owned and excluding ADM have an average annual production capacity of 73.7 mgy.

The expansion of the ethanol industry in the U.S. Midwest from 2001 to 2015 is shown in the Midwest Ethanol Biorefinery Locations movie that accompanies this dissertation. More information is provided on the movie in Section 4.9.

1.2.5 Policy Driven Fuel Ethanol Industry Expansion

The fuel ethanol industry is essentially a study of how policy is used as an instrument to drive industry growth. These policy interventions can take several forms. Several legislative acts have used tax incentives to promote the blending of fuel ethanol into the domestic gasoline fuel supply. Other policies have driven fuel ethanol demand either through bans on substitute products or directly through volumetric blending requirements. The following discussion will chronologically highlight a few of the many policies used to either directly or indirectly promote the growth of the fuel ethanol industry.

The birth of the modern fuel ethanol industry can be traced backed to the late 1970s with the passage of the Energy Tax Act of 1978 which provided an exemption from the gasoline excise tax for any gasoline blended with at least 10% ethanol (Tyner, 2008). Another boost for the ethanol industry came with the passage of the Clean Air Act Amendments in 1990 which established the oxygenate requirements for gasoline and also banned the use of lead additives in gasoline. Oxygenate requirements were established to formulate a more clean burning fuel referred to as Reformulated Gasoline (RFG) which reduced air pollution in smog prone areas (USEPA, 2018). Lead additives were given a five year phase-out period and leaded fuel sales were completely banned for on-road use on January 1, 1996 (USEPA, 1996).

In the late 1990s, another octane enhancer, methyl tertiary butyl ether (MTBE), was deemed to be a ground water and surface water contaminant. Due to the ban on leaded gasoline and the RFG requirement for smog prone areas, MTBE had become the preferred gasoline additive for refiners. Investigations into fuel leaks from underground storage tanks (USTs) showed high concentrations of MTBE in soil and water samples (USEPA, 2008). Based on concerns related to water contamination, states took legislative action to ban MTBE's use in gasoline (USEPA, 2007). Thus, the systematic elimination of a substitute product to fuel ethanol, MTBE, essentially increased the demand for fuel ethanol.

Prior to the events of September 11, 2001, the U.S. was a large importer of crude oil to support domestic energy needs. Afterwards, it was consider highly likely that individuals made rich in oil producing countries were financing international terrorism. Though this was only suspected at the time, it was eventually confirmed by the 9/11 Commission Report by the National Committee on Terrorist Attacks upon the United States (2004). In the interim, this led to discussions on energy independence and security, and sustainable energy as a means to reduce our dependence on foreign oil imports. Eventually, the Energy Policy Act of 2005 enacted the first Renewable Fuel Standard (RFS) which the U.S. Environmental Protection Agency (USEPA) referred to as RFS1 (RFS-one). In the text of the legislation describing the benefits of RFS1, the legislation stresses the importance of energy security, sustainable and renewable fuel sources, and rural economic development. Since ethanol biorefineries were already expanding operations in the U.S. Midwest at the time of this legislation, the intended effect was to drive more investment and growth into this region. The RFS1 covered the years 2006 to 2012 and annually incremented volume requirements from 4.0 billion gallons up to 7.5 billion gallons of ethanol over the seven year period of RFS1.

The RFS was revised by the Energy Independence and Security Act of 2007 (EISA) which is referred to as RFS2. RFS2 requirements have been extended through 2022 and the volumetric requirements are shown in Figure 1.4. Though RFS2 stresses the development of advanced biofuels, cellulosic, and biomass based fuels, these technologies are presently underdeveloped and in most cases are not cost competitive. Conventional ethanol is produced from corn kernel starch and is seen as less desirable than the advanced biofuels due to its less favorable greenhouse gas (GHG) lifecycle emissions profile. Under RFS2, conventional ethanol reaches it maximum volumetric requirement of 15 billion gallons by 2015. Though the legislative requirement is set in RFS2, the USEPA actually sets the true volumetric requirements in any particular calendar year for which the fuel blenders are required to comply. Also, the USEPA does not set the conventional ethanol requirement. USEPA sets the overall requirement for ethanol and advanced ethanol fuels, then the conventional ethanol requirement is determined by subtracting the advanced ethanol fuels requirement from the overall ethanol requirement. Based on the 2017 volumetric requirements released by the EPA, conventional ethanol will meet the 15 billion gallon requirement in 2017 (USEPA, 2017).

Thus, as of 2017, conventional ethanol has reached its maximum production output as defined by the RFS which can be blended in the domestic gasoline supply.

The proceeding discussion presented particular policy highlights that have influenced the production of fuel ethanol. Figure 1.5 shows the evolution of monthly and annual fuel ethanol production for the United States. Ethanol industry fuel production was fairly slow up until the early 2000s. Then, the MTBE bans and the establishment of the RFS drove industry growth up until 2010 where the industry's growth dramatically slowed. There are a few reasons for the slow-down. First, ethanol blending was approaching the 10% blend wall which is the maximum percent of ethanol in gasoline deemed safe for operation in engines built prior to 2001. Second, industry capacity build-out had out-paced volumetric requirements of the RFS. In 2010, the RFS mandated 12 billion gallons of conventional ethanol, but the total industry capacity was about 14 billion gallons and production was about 13.5 billion gallons (RFA, 2011). A final reason is that most of the conventional biorefinery capacity had been built out to meet the 15 billion gallon regulatory limit on conventional ethanol, even though the EPA established requirements were slightly lower than the regulatory levels. Plus, it would take another seven years before the EPA would allow conventional ethanol to produce at its maximum regulatory level of 15 billion gallons (USEPA, 2017).

Future efforts for industry expansion seem to be focused on increasing the amount of ethanol as a percent of the fuel mix from E10 to E15. Currently, in the U.S., most gasoline is E10. For more than two decades, E10 has been established as the blend wall which is the maximum volumetric percent of ethanol mixed with gasoline without causing engine or fuel system damage on non-FFVs. Research by the U.S. Department of Energy and U.S. EPA has determined that all vehicles built in 2001 or newer can use E15 without vehicle deterioration over its useful life span (USEPA, 2015). The Alliance of Automotive Manufacturers and the American Motorcycle Association have differing views on the use of E15 in which they claim it is harmful to engine systems in non-FFVs and may void vehicle warranty coverage. Despite the back and forth, it is not quite clear what the ethanol industry wants to achieve by pushing for E15, since a majority of ethanol biorefiners are conventional ethanol producers (corn-based) and are currently producing at their maximum legislative authorization of 15 billion gallons. Perhaps they want to become more entrenched in the marketplace prior to the potential expiration of the RFS at the end of 2022.

1.2.6 Local Economic Impacts of Ethanol Biorefineries

The best approach for explaining the local economic impacts of an ethanol biorefinery is to go through an example. Assume a 100 millon gallon per year (mgy) ethanol plant is to be constructed in a rural county. Capital costs[5] are assumed to be around $2.00 to $2.50 per gallon of capacity, but for simplicity assume the total construction costs are $200 million. A prerequisite for any potential construction site must include proximity to an existing rail line and numerous corn producers within a 30 mile radius which are capable of supplying a significant portion of the plant's feedstock demand. Most of the construction costs are expected to go to specialized firms which produce the distillery equipment, grain handling and storage facilities, large storage

[5] Iowa State University's Agricultural Marketing Resource Center (AGMRC) created a spreadsheet on ethanol biorefinery profitability which addresses capital costs and annual operating costs. Spreadsheet can be found at: https://www.agmrc.org/national-value-added-agriculture-conference/renewable-energy/ethanol/ethanol-profitability

tanks, fuel handling, grain dryers, decanter centrifuges, etc. and the crews that specialize in installing these operations. Thus, the initial construction phase is expected to have minimal local impact, since most of the resources used to build the facility are acquired from outside of the local community.

In most cases, the start of production (or “first grind” as it is generally referred to in the industry) occurs about 1 year after the start of construction. At this point, the plant employs 40 to 65 people consisting of operators, material handlers, maintenance staff, lab technicians, administrative assistants and management with an average salary in the range of $75,000 per year. Though good middle-class jobs are provided by the plant, the $4 million or more in salaries is still a minimal impact on the local community. The greatest impact on the local community is through the purchase of feedstock from local corn producers. A 100 mgy ethanol plant requires over 35.7 million bushels of corn to produce at the rated output. If a corn price of $3.50 per bushel is assumed, then the annual corn feedstock cost to the plant is about $125 million. Corn producers receive their $125 million in revenue which gets dispersed into the local economy by paying back loans at the bank, buying seed and fertilizer, buying or repairing equipment, and buying general supplies. Since it requires many corn producers to supply an ethanol plant, their aggregate purchasing in the local community generates more local benefits through multiplier effects.

Additional mechanisms which may impact the local economy are local ownership of the ethanol biorefinery and the production of WDGS and DDGS. Local ownership, such as farmer cooperatives, is more likely to reinvest some of their profits into the local community. Non-local owners are more likely to export their profits outside of the community and therefore, the local community does not benefit from the ethanol biorefinery's profits. Wet and dry distiller's grains with solubles (WDGS, DDGS) could benefit local cattle, swine, and poultry operations with a plentiful and relatively inexpensive source of feed. If these operations reduce operational costs based on using this feed source, then the operations would stay competitive and perhaps realize higher profits which would benefit the local community.

The preceding discussion has provided a brief overview related to some of the economic pathways in which an ethanol biorefinery can economically benefit a local community. Therefore, it seems reasonable that ethanol biorefineries have direct, indirect and induced economic impacts on the local economy. Thus, there should be measureable differences between counties with ethanol biorefineries and appropriate counterfactual control counties without ethanol biorefineries.

1.3 Research Objective

The objective of this research is to analyze the economic impacts of newly, continuously operational ethanol biorefineries on rural counties in the U.S. Midwest between the years 2001 to 2015 using a quasi-experimental methodology. This research will demonstrate whether rural counties with newly operational ethanol biorefineries experienced significant growth and other economic impacts relative to similarly matched control counties without an ethanol biorefinery. Additionally, one of the stated purposes of the Renewable Fuel Standard (RFS) policy was to promote rural economic development. If statistically significant positive results are achieved in this analysis, then the rural economic development objective of the RFS will be confirmed.

1.4 Research Hypothesis

The research null hypothesis is designed to present a one-sided alternative hypothesis of positive significance if the treatment effect, defined by the delta symbol (δ), has a one-sided p-value of less than 10%. A 10% threshold is selected based on a limited treatment group size (only 56 units for rural counties), of which, individual units are located over a quite varied spatial region. The research null and alternative hypotheses are formerly presented as follows:

Null hypothesis ($H_o: \delta \leq 0$): Rural counties in the U.S. Midwest with new and continuously operational ethanol biorefineries during the period 2001-2015 did not experience economic benefits relative to similar (or matched) rural counties without ethanol biorefineries.

Alternative hypothesis ($H_a: \delta > 0$): Rural counties in the U.S. Midwest with new and continuously operational ethanol biorefineries during the period 2001-2015 experienced positive economic benefits relative to similar (or matched) rural counties without ethanol biorefineries.

In this research, rural counties will be classified as all counties with populations less than 25,000. This definition is much simpler to interpret than the United States Census Bureau's definition which generally specifies rural as anyplace that is not urban (50,000 or more) or an urban cluster (at least 2,500 up to 50,000). More information will be provided on this topic in the Methodology Chapter.

1.5 Research Structure

The structure of this dissertation will proceed as follows. Chapter two presents a literature review covering quasi-experimental analysis, regional economic analysis and ethanol industry specific research.

Chapter three lays the foundation for the quasi-experimental methodology used to analyze the research hypothesis. The chapter progresses from a theoretical background on the difference-in-differences econometrics approach used for testing the research hypothesis to addressing practical issues associated with its execution.

Chapter four discusses the data used in this dissertation. The full and rural treated and control groups are presented. Descriptions and descriptive statistics are provided for the data used to match treated and control groups using the Mahalanobis distance metric. Descriptive statistics are provided for the dependent variable which can provide insight into the regression analysis.

Chapter five presents the empirical results from the quasi-experimental analysis. The hypothesis is tested against six models from one to five years after treatment using a stable Midwest control county group. One model is selected from the first stage and then run against all control groups to test the robustness of the results over control groups. Employment and employment rate means are compared against two representative control groups. An employment multiplier is calculated based on treated county employment gains after treatment. Exploratory analysis is presented on the full set of treated counties without population restrictions. Finally, an exploratory analysis is presented on the impacts of initial capacity versus population. Chapter six discusses the conclusions of the dissertation and presents ideas for future research.

1.6 Tables in Chapter 1

Table 1.1 Ethanol Dry Mill: Input Feedstock and Output Co-products (RFS, 2019)

Input	**Outputs**
1 Bushel of Corn (56 lbs)	2.86 gallons of denatured fuel ethanol
	15.9 lbs of distillers grains (10% moisture)
	0.75 lbs of corn distillers oil
	16.5 lbs of biogenic CO_2

1.7 Figures in Chapter 1

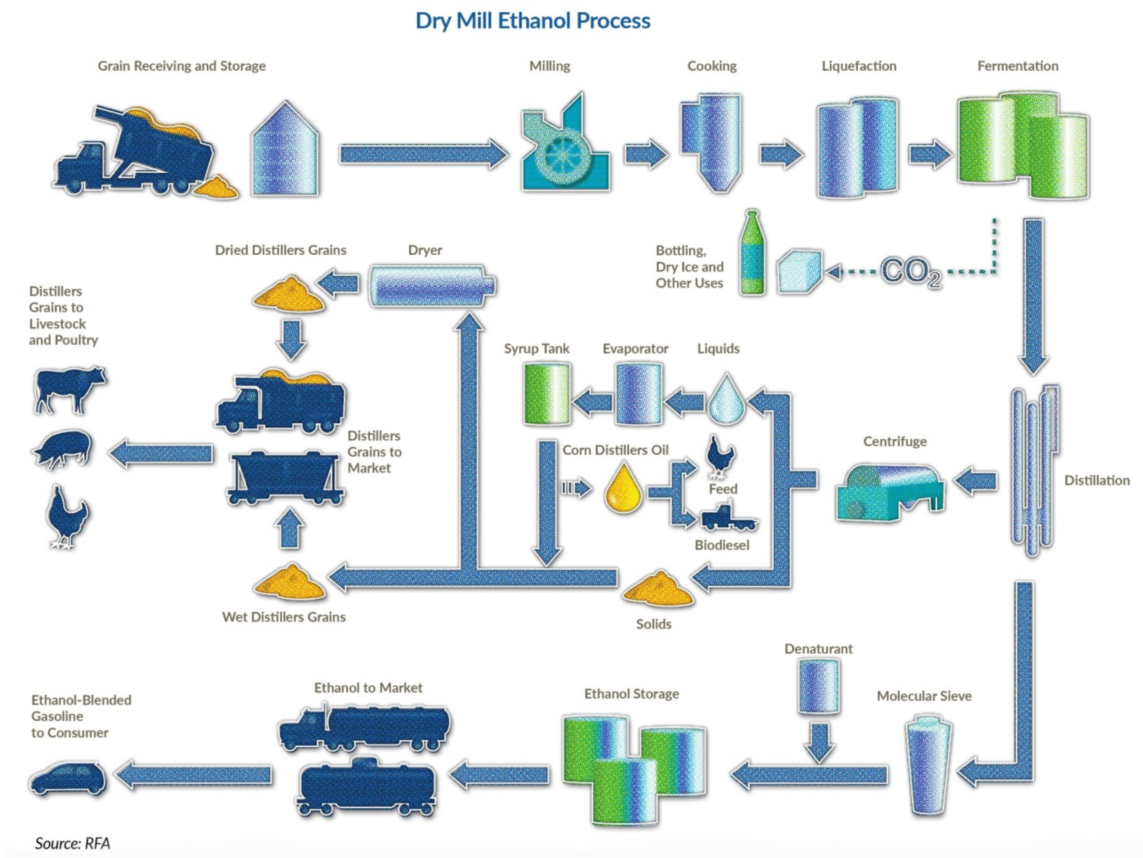

Figure 1.1 Ethanol Dry-mill Production Process Flow
[Source: Renewable Fuels Association (RFA)]

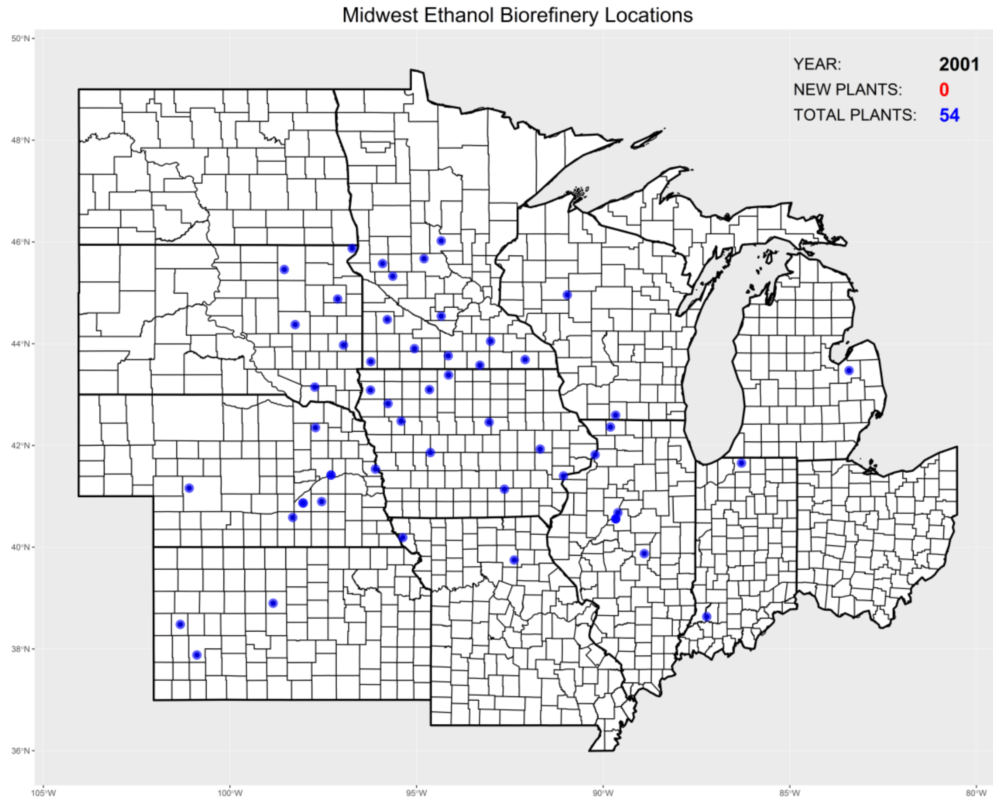

Figure 1.2 Midwest Ethanol Biorefineries in 2001

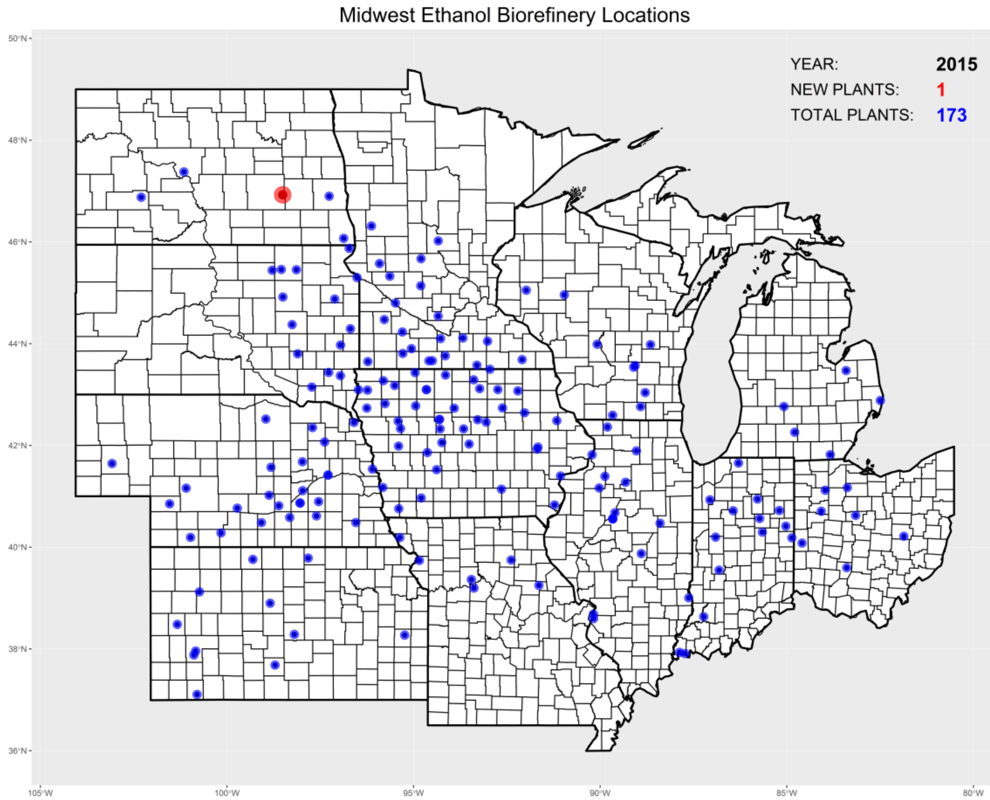

Figure 1.3 Midwest Ethanol Biorefineries in 2015

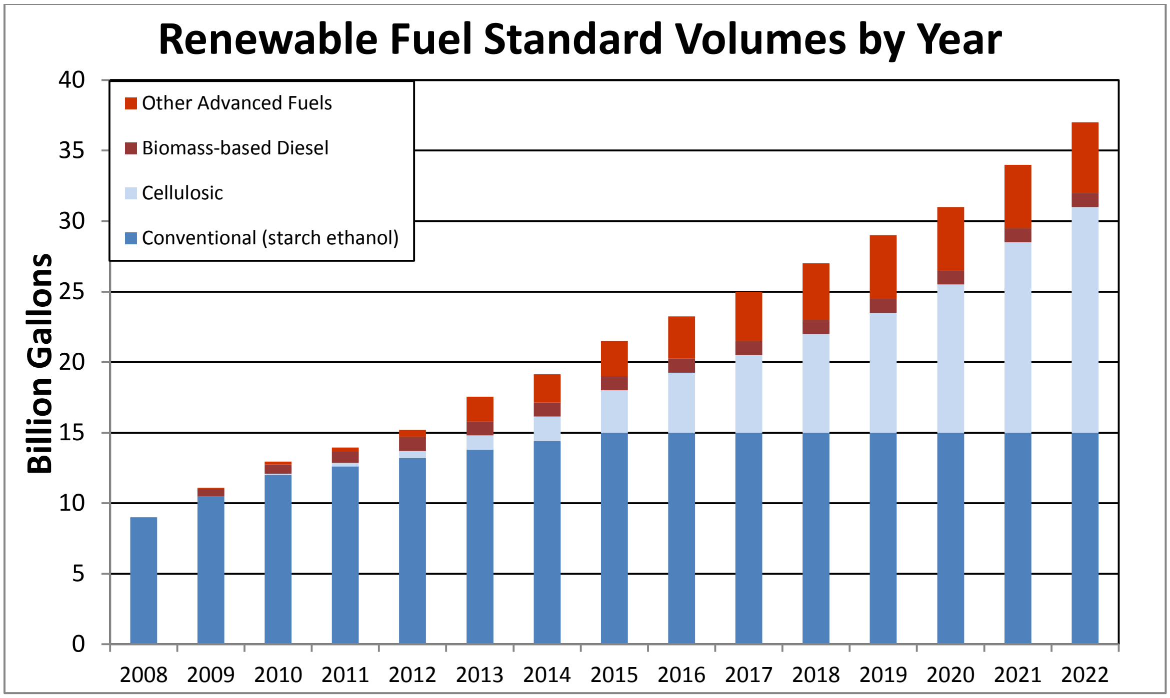

Figure 1.4 RFS2 Volumetric Requirements
(Source: EIA)

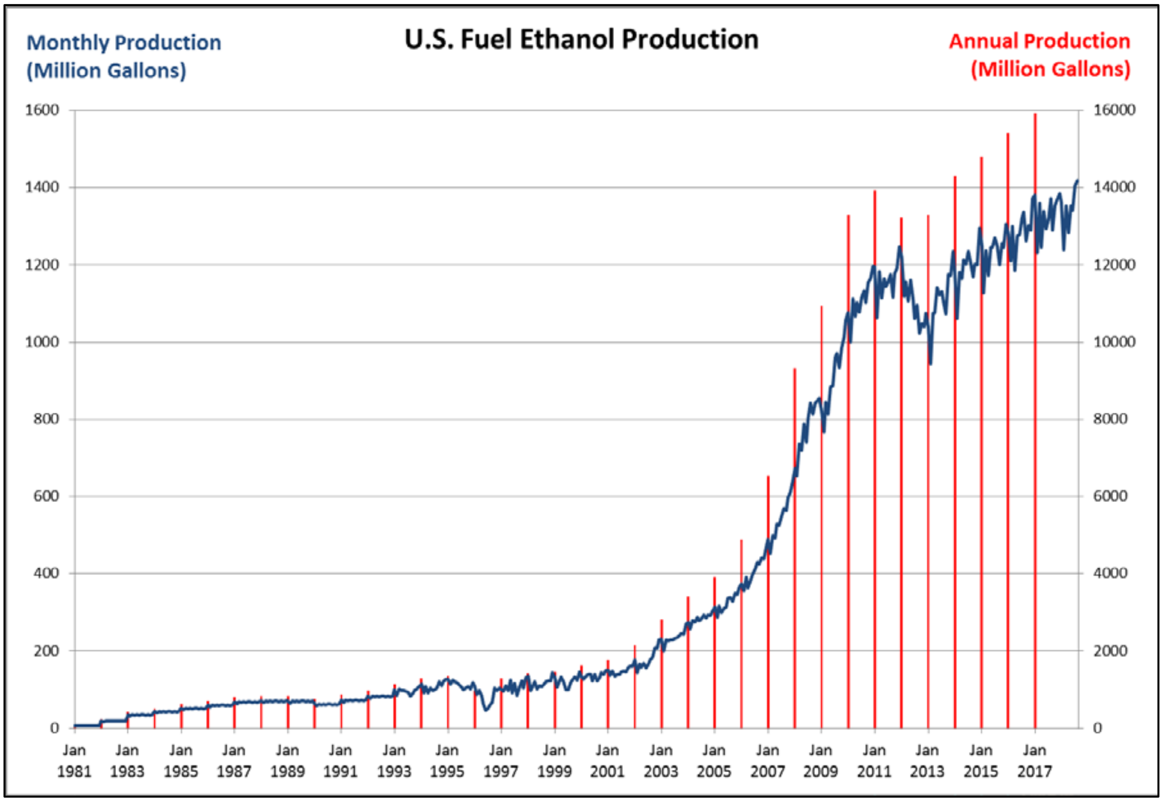

Figure 1.5 U.S. Fuel Ethanol Production

CHAPTER 2. LITERATURE REVIEW

There are four aspects of the literature relevant to this research: regional economic analysis and the quasi-experimental approach, quasi-experimental applications, recent spatio-temporal analysis, and ethanol industry specific research focused on local economic impacts. Each of these topics will be addressed in sequence.

2.1 Regional Economic Analysis and the Quasi-Experimental Approach

In Feser (2013), Campbell and Stanley (1963) are credited with the origins of the term "quasi-experiment" in an effort to build new research design approaches which would facilitate causal inference when other methodologies were not feasible. A quasi-experiment is essentially defined as a research design which has distinct similarities to traditional experiments except that the study subjects (counties are the subjects in this dissertation) are not randomly assigned to treatment and control groups (Rephann and Isserman, 1994; Feser, 2013). The nonrandom selection of the counterfactual (control group) can produce biased outcomes and therefore care must be taken in the selection process in order to find a control group that effectively represents the counterfactual of the treatment group (Feser, 2013). Though Campbell and Stanley (1963) were the leading advocates of quasi-experimental methodology, Isserman and Merrifield (1982, 1987) were among the first researchers to systematically apply the quasi-experimental approach to regional economic development studies. Their work emphasizes the application of quasi-experimental techniques to treatment events in economic geography and stresses the importance of identifying suitable control groups. In their approach (after a treatment event is identified), a set of potential control matches are scrutinized on multiple attributes (selection variables) in order to identify the best control matches suitable for a control group. Isserman and Merrified (1987) also recommend a pre-test phase prior to treatment in order to confirm that the economic variables of the control counties adequately represent the economic variables of the treatment counties. Then, the treatment effect for the treatment group is identified relative to the control group (counterfactual) either through difference methods or using a difference-in-differences approach (Feser, 2013). The selection process for control groups was further formalized in Rephann and Isserman (1994) using the Mahalanobis distance metric. In this approach, an inverse variance-covariance matrix is used to orthogonalize and apply equal weights to a set of selection variables in order to determine the best control match for a particular treatment subject (county). The Mahalanobis distance metric is the approach pursued in this research and further discussion on the quasi-experiment methodology is presented in the CHAPTER 3.

2.2 Quasi-Experimental Applications (Spatial Linkages)

In many cases, quasi-experiments have spatial dimensions. Often these spatial dimensions are suppressed because the spatial linkages are considered weak; otherwise, the spatial linkages must be explicitly modeled. Though this is demonstrated in many articles, only two will be discussed here. Card and Krueger (1994) examined employment growth at fast food restaurants to assess the impact of a 1992 minimum wage hike in New Jersey against a counterfactual eastern Pennsylvania region without a minimum wage hike. The results of the study showed “no indication” that employment was reduced in New Jersey due to the minimum wage hike. In this case, the labor markets were considered sufficiently separate; thus, spatial linkages were not modeled.

In another study using repeated sales approach to assess the impact on housing prices associated with the location of an incinerator, Kiel and McCain (1995) modeled the distance of homes from the incinerator. It was determined that during siting and construction phases of the incinerator, the distance from the incinerator did not affect housing prices. After the incinerator went operational, the distance from it became significant. In this case, the distance from the incinerator is an important characteristic and must be modeled; otherwise, the parameter results could be biased.

The importance of the preceding literature is that when there are true spatial linkages, those spatial linkages should be modeled. In this research, there are two main spatial linkages: spatial group effects which are modeled as state fixed effects and spillover effects which are due to economic activity that is not restricted to the artificial boundaries of a county. More will be discussed on these effects in CHAPTER 3.

2.3 Spatio-Temporal Analysis

There are recent efforts to improve the modeling of spatial-temporal problems by improving spatial econometric techniques. Several articles published by Jean Dubé and Diègo Legros (2012, 2014b) and Jean Dubé et al. (2014a, 2017) are expanding spatial econometric techniques, such as spatial difference-in-differences to handle more complex spatial analyses. The advantage of these techniques is that it adds structure to the spatial dimension which allows the model to be solved. Otherwise, if the spatial structure is suppressed, appropriate variables must be added to the model in order to compensate for the suppressed spatial structure. Effort was made to incorporate the spatial difference-in-difference into this research, but the complex spatio-temporal nature of this research made this extremely difficult in the short-run.

2.4 Ethanol Industry Specific Literature

There are numerous studies (Taylor and Elliot, 2012; Urbanchuk, 2007-2016) which attempt to quantify the benefits of ethanol biorefineries, but these studies are mostly based on the IMPLAN (Impact Analysis for Planning) analysis which utilizes input-output models and economic multipliers to assess local, state and national economic impacts. The results of these studies are extremely dependent on the assumptions used in the models and can produce overly inflated estimates of the economic benefits as discussed in Swenson (2007, 2008) and Low and Isserman (2009) with remedies proposed in Low and Isserman (2009). Most statistical research on the local benefits of ethanol biorefineries tend to focus on the impacts to local corn prices (McNew and Griffith, 2005; Katchova, 2009) or land values (Blomendahl, Perrin & Johnson, 2011). In McNew and Griffith (2005), corn prices were examined over the period from 2000 to 2003 and their research determined that there was a significant increase in corn prices within a 68 mile radius of an ethanol biorefinery. Katchova (2009) analyzed USDA Agricultural and Resource Management Survey (ARMS) data for 2005 and 2007 using a difference-in-differences approach between regions with and without an ethanol biorefineries. The results did not show any significant corn price differences between regions. Also, there is no indication in Katchova (2009) as to the use of any matching techniques to find a suitably matched control region; thus, there is the possibility that the results might be biased relative to a properly matched counterfactual control group. Blomendahl, Perrin & Johnson (2011) proposed that land values should reflect corn price gradients over the region, quality of land, and any transportation cost reductions associated with being near an ethanol biorefinery. Their study for 961 farmland parcels, during 2004 to 2008 in Nebraska, showed no support for ethanol biorefinery location positively affecting land values in the immediate vicinity of the plant. Thus, a clear economic impact due to ethanol biorefinery location has not been uncovered by corn price and land value analysis.

Collectively, these studies seem to show conflicting results on whether ethanol biorefineries have positive impacts on rural communities. There could be several reasons for this since each study uses different time frames and different regions in their analysis. However, the research in this dissertation spans a much longer time frame, 2001 to 2015, and covers approximately 87% of the ethanol biorefineries in the U.S. with county population limits imposed in order to isolate the effect on more rural communities. In essence, this seems be the first research which applies statistical analysis techniques at the county-level to assess whether positive economic impacts (other than corn pricing or land values) occurred in rural communities due to the initial operation of ethanol biorefineries.

CHAPTER 3. QUASI-EXPERIMENTAL RESEARCH METHODOLOGY

This chapter addresses the quasi-experimental methodology used in this dissertation. The approach pursued in this research is strongly influenced by the regional economic research of Rephann and Isserman (2004), Isserman and Merrified (1982, 1987) and Ona, Hudoyo, and Freshwater (2007). It was mentioned in Holland (1986) that there is "no causation without manipulation." In this research, numerous ethanol biorefineries were located in counties throughout the U.S. Midwest from 2001 to 2015; thus, the manipulation (i.e. treatment). Whether this manipulation had a hypothesized significant economic impact on local communities can only be solved using a carefully crafted research methodology.

3.1 The Rubin Causal Model

This is not meant to be a complete explanation of the Rubin causal model, but there are certain elements worth mentioning which can add value to the subsequent discussions. The Rubin causal model is based on the potential outcomes framework (Imbens and Rubin, 2015). For instance, if a treatment can be defined, then it is necessary to define a counterfactual to that treatment. In the context of this research, a treatment can be defined by the start of operations of the first new ethanol biorefinery in a U.S. Midwest County from 2001 to 2015. Similarly, a county without an ethanol biorefinery contemporaneously exists at the same time in the U.S. Midwest and can be classified as a no-treatment county. Since both of these events can be imagined, it is possible to propose that both events have potential outcomes. In reality, these outcomes can be measured to evaluate the response of each subject (county). A causal interpretation can be formulated by differencing the treatment measured outcome against the no treatment measured outcome.

Since this research involves non-equivalent no-treatment groups (control groups), the causal interpretation must account for the differences in the composition of these groups. This approach requires a difference-in-differences approach since it compensates for the differences between groups prior to treatment. In order for there to be a causal interpretation, the identification strategy of Section 3.2.1.1 must be satisfied.

3.2 Quasi-Experimental Methodology

A quasi-experimental research design is pursued to test the hypothesis of this research using an approach referred to by Cook and Campbell (1979) as “interrupted time series and non-equivalent no-treatment control group” design. The interrupted time series and non-equivalent control group design is diagrammed in Figure 3.1 where T and C represent the Treated and Control groups, respectively, and the X represents a point in time where the treated group is exposed to the treatment. The term “non-equivalent” refers to the fact that the treated and control groups were not established through random sampling. In the diagram, there are multiple pre-test and post-test measurements (where “O” represents observations) taken on a relevant set of variables. The multiple pre-test measurements are used for matching to a non-equivalent control group. In this research, the matching process is implemented through the Mahalanobis distance metric which is discussed in Section 3.3. The post-test measurements are used to develop a treatment response profile by evaluating the difference-in-differences (DID) regression model with pre-test and post-test data as shown in Figure 3.2 and Figure 3.3 which are examples of testing one and three years after treatment, respectively, relative to a base year (year prior to treatment). For the actual results, regression models are run on the dependent variable from one to five years after treatment.

Though the quasi-experiment follows the diagram in Figure 3.1, this is accomplished by time aligning each treated-control county pair on the treatment event X. For instance, treatment on a treated county can occur at any point in time from 2001 to 2015. In reality, it occurred at a specific time for a particular treated county. These numerous treatment events are time aligned around the treatment, X, which effectively pools the data for the regression model. Even though the data is pooled for the regression, the time information is preserved which allows for contemporaneous matching between the treated and control counties.

3.2.1 Difference-in-Differences (DID) Model

The difference-in-differences (DID) model is used to analyze the dependent economic variable over the treated and control counties in order to acquire the average treatment effect on the treated (ATOT), δ, as defined by the following equation:

$$y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 After_t + \delta(Treat_i \times After_t) + ControlVariables + e_{it} \qquad 3.1$$

where $Treat_i$ is an indicator variable that takes a value of one for a treated county and the value of zero for a control county and $After_t$ is an indicator variable that takes on the value of one in the period after the treatment has occurred and the value of zero in the period prior to treatment. Several control variables can be added to the model, such as, state fixed effects, time fixed effects and spillover fixed effects which are discussed in Section 3.2.1.2.

Taking the expectation of Equation 3.1 over the various dummy variable values produces the following result (excluding state, time, and spillover indicator variables):

$$E(y_{it}) = \begin{cases} \beta_0 & Treat = 0, After = 0 \quad [Control\ before = A] \\ \beta_0 + \beta_1 & Treat = 1, After = 0 \quad [Treatment\ before = B] \\ \beta_0 + \beta_2 & Treat = 0, After = 1 \quad [Control\ after = E] \\ \beta_0 + \beta_1 + \beta_2 + \delta & Treat = 1, After = 1 \quad [Treatment\ after = C] \end{cases}$$

which shows how the different parameters correspond to the diagram in Figure 3.4. The parameter of interest is δ and is determined as

$$\hat{\delta} = (C - E) - (B - A) = [(\beta_0 + \beta_1 + \beta_2 + \delta) - (\beta_0 + \beta_2)] - [(\beta_0 + \beta_1) - \beta_0]$$

or can be directly obtained from an DID econometric approach proposed in Equation 3.1 using an econometric analysis package, such as, STATA which automatically calculates the standard errors. Several econometric textbooks (Hill, et al., 2011; Stock & Watson, 2015) recommend the use of cluster-robust standard errors on cross-sectional subjects (counties in this case) when using panel data or DID techniques to account for any serial correlation in the data. A preliminary investigation into the use of cluster-robust standards errors for the natural log of real per capita earnings showed a reduction in the standard error in almost every case which produced higher t-statistics and lower p-values. Therefore, in this dissertation, a conservative approach to significance is pursued by utilizing Huber-White robust standard errors which produced lower t-statistics and higher p-values.

The final step is to test the hypothesis using a standard one-sided t-test which is defined as:

$$t = \frac{\hat{\delta}}{se(\hat{\delta})} \qquad 3.2$$

where $\hat{\delta}$ is the DID parameter estimate for the mean economic benefit between a treated county and an untreated county, and $se(\hat{\delta})$ is the Huber-White robust standard error of the economic benefit parameter. As stated previously, the null and alternative hypotheses are:

Null hypothesis $\quad H_o: \delta \leq 0$

Alternative hypothesis $\quad H_a: \delta > 0$

If there is growth in the treated counties that significantly exceeds the growth in the control counties (counterfactual), then the null hypothesis will be rejected in favor of the alternative hypothesis. In this case, the counties with biorefineries have significantly higher economic growth rates or performance levels than the control counties and it can be inferred that this growth is due to the presence of the biorefineries (ceteris paribus). If the test fails to reject the null hypothesis, then there is no significant difference of economic growth rates/performance levels between treated and control counties. Hence, it could be inferred that the economic benefits of biorefinery location are more widespread than just for the county where the biorefinery is located. Furthermore, the spatial-temporal effect of local economic benefits due to a biorefinery's location may diffuse into neighboring communities much more rapidly than can be captured by the phantom boundaries of a county and time scale of data (annual) used in this study.

3.2.1.1 Identification Strategy

In general, there are two steps in the identification strategy for a DID analysis. The first step is to define an appropriate counterfactual control group that is representative of the treated group in the absence of treatment. Essentially, the second step is a confirmation of the first step, since it establishes that a parallel trend must exist between the treated and control groups prior to treatment. Without the parallel trend, no meaningful results can be obtained from the DID analysis. Further, the parallel trend prior to treatment can be projected as an offset from the control group trend after treatment to establish the parallel trend assumption. The parallel trend assumption is the expected result of the treated group in the absence of treatment. If the treated group's actual trend after treatment noticeably deviates from the parallel trend assumption, then there is a treatment response in the treated group which can be quantified using the DID econometric model. A representation of this model is effectively shown in Figure 3.4 where the dashed line represents the parallel trend assumption. Thus, using an appropriate counterfactual control group and establishing the parallel trend prior to treatment are generally sufficient conditions for identification as long as there are no other factors that can have contemporaneous asymmetrical economic impacts on the treated and control groups. All contemporaneous symmetrical impacts will cancel out due the contemporaneous differencing in the DID analysis.

3.2.1.2 Control Variables

Some factors that may present contemporaneous asymmetrical economic impacts are state fixed effect, time fixed effects and spillover effects. Controlling for these variables minimizes bias in the treatment effect estimate. State fixed effects are represented by an indicator variable for each state in the regression except for the reference state (usually Iowa) to avoid multicollinearity. The intent of the state fixed effects is to capture differences between states (laws, ethanol incentives, etc.) which may bias the parameter of interest. Time fixed effects are a set of indicator variables for all time periods in the regression with the exception of the reference period (usually the earliest period) which captures differences over time. Spillover indicator variables are used to capture the economic spillover effects associate with scenarios where a contiguous treated county may affect the baseline level of economic activity measured in either another treated county or a counterfactual control county. All of these control variables are used to correct for spatio-temporal events which may cause biases in the parameter of interest.

3.2.1.3 Rural Population

In the rural-urban continuum codes (RUCC_2013) data file from USDA-ERS, the 2010 populations are listed for each county. These population values are used for treated-control county matching and to filter the treated data set using this population variable (Population_2010). Since rural county populations go through very slow change, this seemed to be a reasonable approach and greatly simplified the population matching process. The six treated-control group sets matched each treated-control county pair into a 4-tuple set based on the treated county's population in 2010. The 4-tuple set (treated before, treated after, control before, control after) was defined by the variable

“Pop_match”; thus, the treated-control groups could be population filtered on this variable and keep the treated-control matched pairs together.

Based on 2010 population statistics, the populations in the 97 qualified treated counties ranged from 2,695 residents to 270,056 residents. Obviously, a county with a population of 270,056 is not a rural county. In order to analyze only rural counties, a maximum population size for a treated county must be determined. Since there is a trade-off between population level in the treated counties and having sufficient data capable of producing any significant statistical results, the population break point for rural populations was set to be less than 25,000 for any treated county. Thus, less than 25,000 defines the rural population criteria for treated counties in this research for which 56 treated counties meet this criterion (or 56 4-tuple treated-control matched sets for each control group).

3.2.1.4 Population versus Sample Statistics (Analysis Assumption)

Often, real world quasi-experiments use the total population of data rather than a random sample from the population to assess the impacts associated with the experiment. Most econometric models are designed to use sample data from a population and then use the parameter estimates from the regression to make inferences about the population. In this research, the data evaluated in the regression model is assumed to be equivalent to sample data even though it may actually be the complete population data set. There are two ways to think about this. In the first thought process, the mean of any population data set is known with 100 percent certainty with zero standard error. Thus, to assume the population data set is essentially a sample data set is just a small leap since it still has the same mean, but now it has a mean distribution based on the total number of observations and uses the sample standard error for the mean distribution. In some sense, this is a conservative approach to a population analysis since the mean has a distribution. The second thought process is more conceptual and perhaps less palatable. It is possible to assume that the data observed in this research is merely a sample from a super population which exists in a multi-universe. In each universe, the ethanol biorefinery treatment took place in that universe's U.S. Midwest region and the one selected for this research is just a sample from that super population of those events. Though this second explanation is a bit more farfetched, the result is the same as the first explanation.

The importance of this assumption is that it allows for the use of models to evaluate the parameter of interest. Without the model, controlling for state fixed effects, time fixed effects and spillover effects would be much more difficult.

3.2.2 Treatment Definition and Treatment Event

Treatment is defined as the first new corn-based ethanol biorefinery to start continuous operations in a U.S. Midwest county between the years 2001 to 2015. The term "first" is important since it signifies the change in the county status from a county without an operational ethanol biorefinery to a county with an operational ethanol biorefinery. Also, it is expected that the "first" ethanol biorefinery in a county will have a much more dramatic impact on the local economy than a "second" biorefinery or a capacity expansion phase, since these latter events are more likely to create greater spillover effects which makes it more difficult to isolate the economic impact associated with an ethanol biorefinery. The “start” of production or “first grind” is important since it defines when that particular treatment was applied to a particular county.

When a treatment occurs in a particular county, it defines a location (county) and it occurs at a particular time (year). The term “treatment event” will be used to identify the location of the ethanol biorefinery (county) and the start date of production (year) for the county exposed to the treatment. Non-corn based ethanol biorefineries are excluded from consideration since these types of plants experience erratic operations, tend to go out of business, and their initial capacities are generally much smaller than corn-based ethanol biorefineries.

3.2.3 Treated Spatial Units

Treated spatial units are limited to the 1055 counties of the twelve state region referred to as the U.S. Midwest. A treated spatial unit is a county that has been exposed to the treatment as defined by the treatment event.

3.2.4 Control Spatial Units

Control spatial units are limited to the 1055 counties of the twelve state region referred to as the U.S. Midwest, but excludes all counties with an existing or future ethanol biorefinery through 2015.

3.2.5 Matching Process for Control Units

A two stage matching process is used to develop six sets of control counties. In the first stage of matching, each potential control county is contemporaneously ranked against each treated county based on the Mahalanobis distance metric results. The Mahalanobis metrics are calculated for two matching regions: in state matching and Midwest region matching. Each regional matching scheme was implemented into an R program to produce the ranked results for each treated county. In the second stage, a treated-control county match was determined based on three criteria: Mahalanobis distance metric best match (Best Match), population best match, and rural-urban continuum codes (RUCC) best match. The best way to describe this process is through an example.

This example starts with the ranked results from the Mahalanobis distance calculations as shown in Figure 3.5. The matching is all potential control counties (column 1) in the U.S. Midwest against all treated counties whose production started in 2003. The treated counties in 2003 are the four counties shown in columns 2 through 5 (Kearney, NE; Brookings, SD; Turner, SD; Winnebago, WI). For this example, Turner, SD is the treated county which will be matched to control counties based on the three criteria (Best Match, Population Match, and RUCC match). Currently, the Turner County column is ranked based on Mahalanobis distance calculations. When Turner County is ranked matched, its own result is zero as expected. To identify a best match result, it will be selected from the first blank line in the ProdYear column. The ProdYear column represents the first year of production for a biorefinery located in that county. A blank line in ProdYear represents that there is no ethanol biorefinery in that county. Thus, Hutchinson County, SD is the best match control county over the entire Midwest region to the treated Turner County, SD. For a best population match, check for a blank line in ProdYear column and then select the closest population match from the

Population_2010 column. In this case, Hutchinson County, SD is the best population match to Turner County, SD, but the actual selection used in this case was McCook County, SD due to the major differences in RUCC levels. Thus, under some unique conditions, more information is used to make the decision about a control county selection than just the initial criteria. In a similar fashion for RUCC, check for a blank line in ProdYear and then select the best RUCC match. McCook County, SD is the best RUCC match since it has the lowest Mahalanobis distance metric, a blank line in ProdYear, and a RUCC match to Turner County, SD.

Essentially, in-state treated-control county matches are found using the same process. The main difference is that the Mahalanobis matching program uses only in-state based data for matching a particular in-state treated county with all potential in-state control counties for each time period from 2001 to 2015 to facilitate contemporaneous matching.

3.2.6 Dependent Variable

The dependent variable used in the DID model is the natural log of real per capita earnings at the county level. Real per capita earnings represents an aggregate measure of direct, indirect and induced earnings effects that occur at the county level. For a treated county, these aggregate earnings effects associated with the presence of an ethanol biorefinery are expected to be substantial and occur by the mechanisms discussed in Section 1.2.6. In a control county, there should not be any earnings effects associated with an ethanol biorefinery and therefore should present a stable reference to compare the treated county's treatment response against.

Since earnings are in general inherently right-skewed, the natural log of real per capita earnings is used as the dependent variable to insure the error terms are more normally distributed. The log dependent variable also means the coefficients on the indicator variables in the DID model represent growth rates. Other variables, such as, employment and employment rate are used to compare treated county versus control county employment impacts on average, but are not used as dependent variables in a regression model.

3.2.7 Parameter of Interest: Average Treatment on the Treated (ATOT)

The coefficient of interest in the DID model is the average treatment on the treated (ATOT or δ). There is one main reason why the coefficient of interest should be called the ATOT as opposed to the average treatment effect (ATE). Since this is a real-world quasi-experiment, no random sampling was performed to establish the treated and control groups. This means that the control group is potentially non-equivalent to the treated group as mentioned previously. This non-equivalence could mean that a treated county exposed to the treatment could have a different response than a control county exposed to the treatment. Since there is a possible difference between those two potential outcomes, the appropriate way to refer to the parameter of interest for the treated county group is average treatment on the treated (ATOT).

3.2.8 Spillover Effects - General Explanation

Spillover effects represent the situation where the economic benefits of an ethanol biorefinery cannot be isolated to just the county where the plant resides. These economic benefits most likely occur through increased corn production and associated multiplier effects in contiguous counties relative to the treated county. Spillover effects affect both treated and control counties. Indicator variables are used to capture these effects and are based on a set of conditions for the treated and control counties.

A treated county has four specific situations which affect the assignment of spillover indicator variables. First, when a treatment event occurs in a treated county and there are no contiguous treated counties, then no spillover indicator variables are required because this is not an identifiable situation (spillover effects cannot be quantified by the information provided in the DID model). Second, when a treatment event occurs in a treated county and there is/are previous contiguous treated counties, then an indicator variable must be set in the before period (prior to treatment) and in the after period (post-treatment). The prior treatment spillover indicator variable captures the possible economic spillovers from the previously treated counties into the newly treated county and is identifiable based on the relative difference between this situation and the situation with no contiguous treated counties. The post treatment spillover indicator variable captures the spillover changes that occur based on the new plant's operations in the new treated county. Third, there is the situation where two contiguous counties are treated simultaneously with new ethanol biorefineries. In this case, a single spillover indicator variable is appropriate in the post treatment period. The fourth situation is similar to the third situation, but involves evaluating the DID model for two to five years after treatment. For instance, if the treated county under analysis has a contiguous county that undergoes a treatment event a year later, then in the two years after treatment analysis a spillover indicator variable is used to capture these spillover effects. Thus, this fourth approach is used to capture spillover effects based on DID model evaluation for two to five years after treatment and is conditional on the time after treatment.

For a control county, there are three situations which require explanation. First, when there are no contiguous treated counties prior to treatment or when there is a contiguous treated county prior to treatment, no spillover effect indicator variables are set for the control county. It is obvious that no spillover indicator is required when there is no contiguous treated county. In the case of a contiguous treated county, there are no differential events that will allow for identification of this spillover effect. Additionally, these control counties were assigned based on Mahalanobis best matches over several criteria in the pre-treatment period which means these counties were good matches whether there were spillovers effects or not. Second, a contiguous treated county occurs simultaneously with the control county. For this case, a spillover indicator variable is set in the after period. The third case involves evaluating the DID model for two to five years after treatment and has the same explanation as for the fourth situation for the treated counties.

What do the spillover effects indicator variables do? For treated counties, the spillover effect indicator variables essentially compensate for the lost economic benefit that spills over into contiguous counties. In control counties, the spillover effect indicator variables attempt to capture the increased economic activity in the control county related to the spillover from a contiguous treated county which overinflates the county's baseline level in the post treatment period. Thus, using spillover indicator variables should statistically approximate the true value of ATOT that would occur in the absence of spillover effects.

3.2.9 Spillover Effects – Setting the Indicator Variables

In order to set the indicator variables for spatial spillovers, the contiguity between all Midwest counties must be established. This is accomplished through a series of steps starting with GeoDa software. In GeoDa, a contiguity file is generated for the Midwest county level ERSI shapefile. The output file is in text format and is based on the unique id of each county unit (FIPS) and the contiguity method (Rook or Queen) used to generate the GeoDa .gal file. The gal file is organized as follows (note: all line information is separated using a single space): the first line contains header information; the second line identifies a particular county in the U.S. Midwest using its FIPS id and then is followed by a number which identifies the number of contiguous counties for that particular county; the third line lists the FIPS ids for all contiguous counties relative to the county listed on the preceding line; and finally, the contiguous information for all other counties in the U.S. Midwest are encoded on separate lines and in the same manner as described for the second and third lines. Thus, all 1055 counties and their contiguous neighbors in the Midwest are listed in the manner just described.

The next step involves converting the gal file into a contiguity matrix. An R program was written to perform this operation by using a Midwest counties file with all counties listed in ascending order by their FIPS id along with their county and state names. This file is used to order the rows and columns of the contiguity matrix in ascending order of the FIPS codes for the counties, since the gal file may not be ordered on the FIPS ids. The contiguity matrix is initialized as an all zero matrix; then, the gal file is read into the program and all spatially contiguous counties are identified with a value of one. Output from the R program is a contiguity matrix of zeros and ones identifying contiguous counties.

Since contiguity to a biorefinery county is a dynamic event in this analysis due to industry expansion from 2001 to 2015, it is important to capture the changes in contiguity to biorefineries over time in order to properly identify the spatial spillover indicator variables. This is accomplished by a another R program which reads in the biorefinery location master file and the contiguity matrix; then, converts this information into a biorefinery contiguity over time data structure for all Midwest counties. The biorefinery location file contains information on existing ethanol biorefineries (prior to 2001) and the start of production years for all new ethanol biorefineries established in the period 2001 to 2015. By identifying all biorefineries that exist at a particular point in time, for example 2001, then the columns of the contiguity matrix associated with these unique biorefinery counties are extracted to form a matrix of all Midwest counties (rows) against all existing contiguous biorefinery counties (columns) for the year 2001. Then, all rows are summed. The result is a single column vector of all Midwest counties (rows) against the number of contiguous counties with a biorefinery (column) in 2001. This process is repeated for all subsequent years to produce a dynamic contiguity to biorefinery county data structure which can be analyzed to assign spatial spillover indicator variables.

There are three unique spatial spillover indicator variables defined for a treated county. If there is an existing contiguous treated county prior to treatment, then a spillover indicator is set in the period before the treatment and another spillover indicator variable is set in the period after the treatment for the reasons explained in Section 3.2.8. If a change in the number of contiguous treated counties is detected for any period of time between the base year and any period up to five years after treatment depending on the DID model being analyzed, then another spillover indicator variable is set to capture these effects for the reasons explained in Section 3.2.8.

Control counties have only one unique spatial spillover indicator variable. If a change in the number of contiguous treated counties is detected for any period of time between the base year and any period up to five years after treatment depending on the DID model being analyzed, then a spillover indicator variable is set to capture these effects for the reasons explained in Section 3.2.8.

3.3 Mahalanobis Distance Metric (Ranking of Control Units)

The Mahalanobis distance metric is used to find the minimum distance match between a treated county and a set of potential control counties. The Mahalanobis distance metric is defined as

$$d^2(\boldsymbol{x}_T, \boldsymbol{x}_c) = (\boldsymbol{x}_T - \boldsymbol{x}_c)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_T - \boldsymbol{x}_c) \qquad 3.3$$

where $d^2(\boldsymbol{x}_T, \boldsymbol{x}_c)$ is the scalar distance metric between the selection variable vectors $\boldsymbol{x}_T$ and $\boldsymbol{x}_c$ for the treated county (T) and potential control county (c), respectively, and $\boldsymbol{\Sigma}^{-1}$ is the inverse variance-covariance matrix for the selection variables over the range of county data (U.S. Midwest county data for region matching and each state county data for in-state matching). A discussion on the selection variables used in this dissertation is presented in Section 4.7. For each treatment year (from 2001 to 2015), there is a unique set of selection variables used to contemporaneously match each treated county with a set of potential control counties based on a five year pre-treatment period prior to treatment. The Mahalanobis metric is applied between a treated county in the five year period prior to treatment against all potential control counties. Then, a search algorithm is used to find the minimum distance closest matched county suitable as a control county. This matching process is repeated for each treated county until each treated county is matched with a control county.

For the calculation of the Mahalanobis variance-covariance matrix, all counties with populations greater than 250,000 people were eliminated. Large population areas can increase the calculated variances and covariances which effectively reduces all calculated Mahalanobis distance metrics between treated and potential control counties. Excluding large population areas served the purpose of creating better matches between treated and potential control counties with similar characteristics while creating extreme mismatches between treated counties and large population potential control counties.

3.4 Figures in Chapter 3

Groups	t-5	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4	t+5	Time
T	$O_{t,t-5}$	$O_{t,t-4}$	$O_{t,t-3}$	$O_{t,t-2}$	$O_{t,t-1}$	X	$O_{t,t+1}$	$O_{t,t+2}$	$O_{t,t+3}$	$O_{t,t+4}$	$O_{t,t+5}$	
C	$O_{c,t-5}$	$O_{c,t-4}$	$O_{c,t-3}$	$O_{c,t-2}$	$O_{c,t-1}$		$O_{c,t+1}$	$O_{c,t+2}$	$O_{c,t+3}$	$O_{c,t+4}$	$O_{c,t+5}$	

Figure 3.1 Diagram of Pre-Treatment Observations and Post-Treatment Observations

Groups	t-5	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4	t+5	Time
T	•	•	•	•	$O_{t,t-1}$	X	$O_{t,t+1}$	•	•	•	•	
C	•	•	•	•	$O_{c,t-1}$		$O_{c,t+1}$	•	•	•	•	

Figure 3.2 Diagram of DID Regression Analysis for One Year after Treatment

Groups	t-5	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4	t+5	Time
T	•	•	•	•	$O_{t,t-1}$	X	•	•	$O_{t,t+3}$	•	•	
C	•	•	•	•	$O_{c,t-1}$		•	•	$O_{c,t+3}$	•	•	

Figure 3.3 Diagram of DID Regression Analysis for Three Years after Treatment

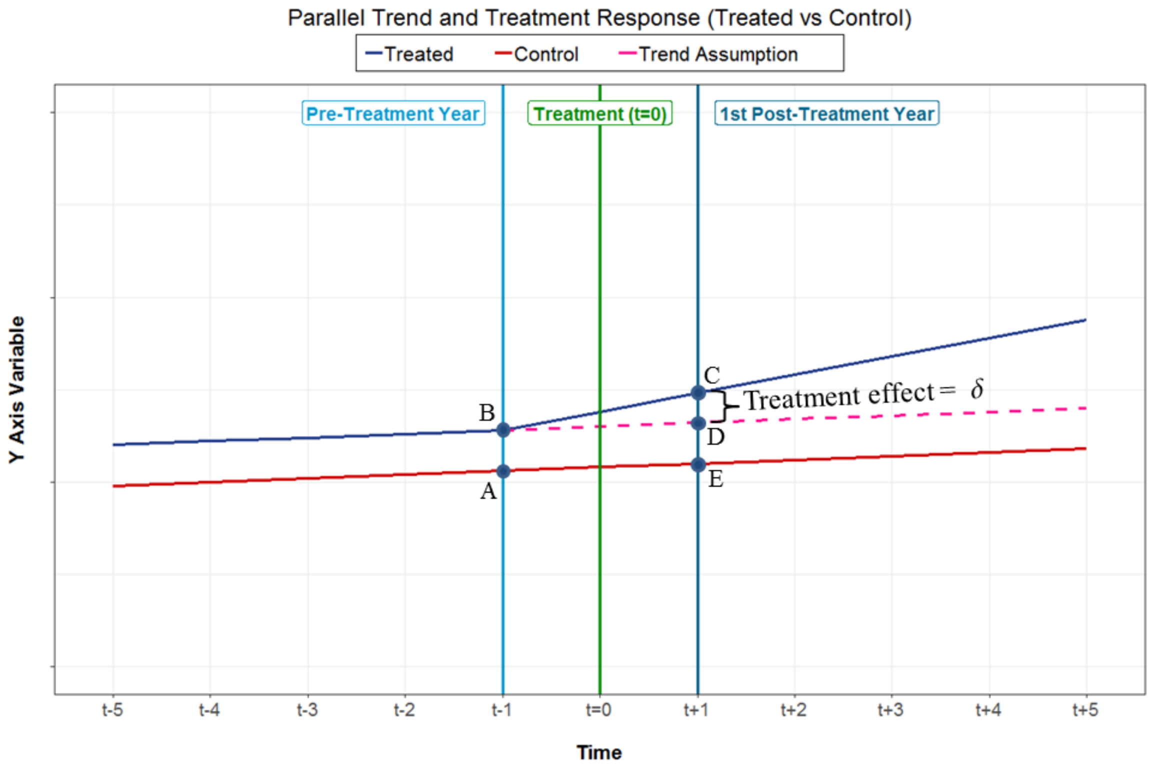

Figure 3.4 Difference-in-Differences Estimation

	Kearney_NE	Brookings_SD	Turner_SD	Winnebago_WI	FIPS	State	County_Name	Population_2010	RUCC_2013	Description	ProdYear
Turner_SD	11.17884999	6.482080742	0	8.55604867	46125	SD	Turner County	8,347	3	Metro - Counti	2003
Lake_SD	11.04601331	4.853708603	3.815769513	7.206490708	46079	SD	Lake County	11,200	6	Nonmetro - Url	2000
Hutchinson_SD	12.6673082	6.37838567	4.66428306	9.093478491	46067	SD	Hutchinson Count	7,343	8	Nonmetro - Co	
Moody_SD	10.04010713	5.403020707	4.984330241	7.849511006	46101	SD	Moody County	6,486	8	Nonmetro - Co	
Audubon_IA	9.287959359	7.325172009	5.018419293	7.371482418	19009	IA	Audubon County	6,119	8	Nonmetro - Co	
McCook_SD	11.49911783	6.614464403	5.104426161	8.83914994	46087	SD	McCook County	5,618	3	Metro - Counti	
Cass_IA	8.903547379	5.162863544	5.15855327	6.88483947	19029	IA	Cass County	13,956	6	Nonmetro - Url	
Butler_IA	9.063956935	5.983018997	5.202185411	7.49861104	19023	IA	Butler County	14,867	8	Nonmetro - Co	2008
Montgomery_IA	9.055319436	6.091084708	5.38537013	6.739290629	19137	IA	Montgomery Cou	10,740	6	Nonmetro - Url	
Madison_IA	9.003961487	5.729134126	5.39142279	7.249987586	19121	IA	Madison County	15,679	2	Metro - Counti	
Worth_IA	8.61197939	5.995849098	5.402493135	6.989810381	19195	IA	Worth County	7,598	9	Nonmetro - Co	2004
Hamlin_SD	10.22304514	4.193167431	5.429761627	7.662094152	46057	SD	Hamlin County	5,903	9	Nonmetro - Co	
Adair_IA	9.412811012	4.788985257	5.435123814	6.182598769	19001	IA	Adair County	7,682	8	Nonmetro - Co	
Guthrie_IA	9.757556992	4.78922329	5.446005927	6.633376014	19077	IA	Guthrie County	10,954	2	Metro - Counti	2000
Pipestone_MN	10.24664481	5.503259446	5.557023898	8.014503944	27117	MN	Pipestone County	9,596	6	Nonmetro - Url	
Bon Homme_SD	10.49088907	4.750124661	5.615845024	7.01716391	46009	SD	Bon Homme Coun	7,070	9	Nonmetro - Co	2000
Lac qui Parle_MN	10.46682434	5.112635886	5.618854965	7.492979598	27073	MN	Lac qui Parle Cour	7,259	9	Nonmetro - Co	

Figure 3.5 Mahalanobis Control County Ranking by Year of Treatment

CHAPTER 4. DATA DESCRIPTION

This chapter provides details on the data acquired and used in the quasi-experimental research design. The logic behind region and time period selection is presented. Treated-control county matches are provided for the full match set of counties (97) and the rural matched set of counties (56) for all control groups. Summary statistics are provided for the data used in the Mahalanobis distance calculations and for the dependent variable used in the difference-in-differences (DID) econometric models. Some of the data for the selection variables and for the treated-control county matching process has been compiled into movies to accompanying this dissertation which demonstrates the dynamic nature of the data used in this analysis.

4.1 U.S. Midwest Region

The U.S. Midwest region is selected for this analysis for two main reasons. First, 87% of operating ethanol biorefineries are located within this region. Second, about 88% of the U.S. corn crop is grown in this region which is the primary feedstock for ethanol production. Thus, if there is a significant economic impact associated with ethanol biorefinery location, then it is likely to be revealed through an analysis of the Midwest region. The U.S. Midwest is a twelve state region consisting of the following states: Illinois, Iowa, Indiana, Kansas, Michigan, Missouri, Minnesota, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. Table 4.1 shows the number of counties in each state, planted acres of corn in 2015, and number of ethanol biorefineries in each state.

4.2 Spatial Units

The spatial units of interest are the 1054 counties as shown by state in Table 4.1 and one independent city in Missouri (St. Louis) located in the U.S. Midwest for a total of 1055 spatial units. These spatial units are mostly referred to as counties throughout this dissertation.

4.3 Time Period Selection

The time period selected for this research spans from 2001 to 2015 for three main reasons. First, this is a period of rapid expansion in the U.S. ethanol industry with a large increase in the number of biorefineries in the U.S. Midwest (54 to 173) along with a dramatic increase in total industry production capacity (2 billion gallons per year to 15.7 billion gallons per year). Second, by 2015, the RFS had reached its maximum regulatory volumetric requirement of 15 Billion gallons per year from conventional sources (corn). Even though the maximum regulatory requirement is reached for conventional ethanol, this does not mean that the EPA will necessarily set the requirement at that level since advanced ethanol biofuels have priority over conventional ethanol. The actual conventional ethanol level in 2015 was 14.05 billion gallons. Third, since 2001, the Renewable Fuels Association (RFA) has published annual data on ethanol biorefinery operation, construction, and initial capacity. The RFA's *Ethanol Industry Outlook* publication is probably the most reliable source for tracking the ethanol industry's progress over this period and can be used as a starting point for tracking ethanol biorefinery initial start dates and operations.

4.4 Ethanol Biorefinery Operational Status and Location

Using RFA's *Ethanol Industry Outlook* publication from 2001 to 2016, the annual lists of all ethanol biorefineries under construction or in operation were compiled into an ethanol biorefinery data set. From this data set, each biorefinery is tracked over time to view its transition from being under construction to going operational which provides the first indication of a production start date (year). Secondary confirmation of a production start date is conducted through information searches, such as, on the company's website, News releases, industry magazines, and satellite images from Google Earth. To confirm a plant's continuous operations, a similar search is pursued. Often, if a plant goes off-line, there are news articles or industry magazine articles which will refer to this event. If there is not sufficient evidence that a plant had continuous operations or a well-defined start date, it is eliminated from the treated county list.

Finding the location of the plant is quite simple. By searching on the ethanol biorefinery's name, city location and state location in Google, search results are generated which provide a detailed street address. Using the street address in Google Earth, the ethanol biorefinery can be viewed in satellites images and geolocated using the latitude and longitudinal information which can be extracted from Google Earth for mapping purposes. A Google Earth image of a 105 mgy ethanol biorefinery located in Iowa Falls, Iowa is shown in Figure 4.1.

4.5 Treated and Control County Data

If a county with a new ethanol biorefinery satisfies the treatment criteria, then it becomes part of the treatment group regardless of the population size of county. A treated county can be eliminated from the treatment group if its production start date and continuous operational status are not confirmed. Then, the list of treated counties is used to find appropriate matches using the Mahalanobis distance metric. Based on the matching process described in Section 3.2.5, six control groups were established. The six treated-control groups with 97 matched pairs are shown in Table 4.2 through Table 4.7. For the purposes of this research, treated counties with populations less than 25,000 are classified as rural. The six rural treated-control groups with 56 matched pairs are shown in Table 4.8 through Table 4.13. The six treated-control groups are classified by region (Midwest or in-state matching) and a secondary sorting process based upon Mahalanobis best match, population best match, and Rural-Urban Continuum Code (RUCC) best match.

4.6 Dependent Variable

The natural log of real per capita earnings is the only dependent variable in this analysis. Per capita earnings data was acquired from the Bureau of Economic Analysis (BEA) for all Midwest counties. Real per capita earnings was calculated by deflating per capita earnings by CPI_{2010}. CPI_{2010} was obtained from Federal Reserve Economic Data (FRED).

4.7 Selection Variables

Selection variables define characteristics for each county in the U.S. Midwest which are deemed important for matching a treated county with an ethanol biorefinery to a set of potential control counties without an ethanol biorefinery. There are 29 selection variables defined for use in the Mahalanobis distance metric calculation and are classified into three categories: growth rates, spatial structure and economic structure as shown in Table 4.14. Table 4.15 shows the 29 selection variables for a hypothetical treated county based on an ethanol biorefinery that starts initial production in 2001 (treatment year). The numerical portions of the variable names are defined relative to the treatment year ("2001" for this example) in order to match the treated county's characteristics with similar characteristics in a potential control county for the pre-treatment period. The growth rates consist of population growth rates and real per capita earnings growth rates for the five pre-treatment years prior to treatment. Spatial structure variables use levels for population, population density, and corn acres in the period just prior to treatment and use corn and soybean production levels for the five pre-treatment years prior to treatment. Finally, the economic structure variables use levels in real per capita earnings, farm real cash receipts (crops), and employment while using share information for farm share of earnings, manufacturing share of earnings and retail share of earnings at the county level.

Descriptions of the dependent variable and selection variable data are provided in Table 4.16. Summary statistics for this data from 1995 to 2016 are shown in Table 4.17. A more detailed summary statistics for the dependent variable used in the difference-in-differences regression (DID) models are shown in Table 4.18. Descriptions for the rural-urban continuum codes (RUCC) are shown in Table 4.19.

4.8 Missing Data

Corn and soybean data acquired from the USDA-NASS Quick Stats website has data missing for certain counties during certain years. Additionally, Bureau of Economic

Analysis (BEA) data has missing values which are generally attributed to confidentially reasons. Since this data is used in mathematical calculations and graphing, it is best to convert the missing data into values based on the history embedded in the data. The following subsections will discuss the techniques used to convert the missing data into values.

4.8.1 Continuity

Continuity adjustment is used when there are clear trends in the harvested acreage data and there are only a few missing years of data for a particular county. In this case, harvested acres are assumed to follow a trend between two known data points. Then, the contemporaneous average yields for the state in which the county resides are used to calculate the production values to fill-in the missing data.

4.8.2 High Uncertainty: Zero-out

When a county displays erratic harvested acres or has no history of production of a particular field crop, then there is a high degree of uncertainty about that county's crop production activities. In this case, the missing data are replaced by zeros.

4.8.3 Phase-in or Phase-out

If a county had no history of production for a crop and later showed strong signals of continuous production of that crop, then a phase-in period is modeled into the harvested acres and then multiplied by the contemporaneous state average yields to calculate the production values to replace the missing values. Similarly, if a county had continuous production of a crop which seemed to sharply decrease, then a phase-out period is modeled into harvested acres and then multiplied by contemporaneous state average yields to calculate the production values used to replace the missing values. Generally, the phase-in and phase-out periods are five years or less.

4.9 Data Movies

Data movies were created to show the spatial-temporal aspects of the data. The content of the movies include three choropleths on crop production, one movie on the expansion of fuel ethanol industry over time (ethanol biorefinery location and start of operations), and two movies on the matching process between treated and control counties. These movies are listed in Table 4.20.

To produce the movies, several R programs were written for mapping the specific type of information (choropleth, biorefinery location, matching) into static data images. These images were imported into a movie maker program and compiled into MP4 movie files.

The three choropleth movies show the evolution of corn production, corn yield and soy production by county over time in the U.S. Midwest. This data is very similar to United States Department of Agricultural (USDA) published choropleths with the only differences being the adjustments made for missing data. In the Midwest Ethanol Biorefinery Location movie, the industry expansion from 54 biorefineries in 2000 to 173 biorefineries in 2016 is animated to show the operational startup of these biorefineries across the U.S. Midwest over time. Ethanol industry expansion is quite dynamic between the years of 2004 to 2009 which is the period when the RFS1 and RFS2 were put into effect, 2006 and 2008, respectively. In the two matching movies, the contemporaneous matches between treated and control counties as a result of the Mahalanobis distance matching process are shown for both matching over the entire Midwest region (RUCC match) and matching within state (Best Match) for the treated counties over time. Treated counties are shown in blue while control counties are displayed in red. A gold line is used to connect the contemporaneously matched treated and control county pairs. The number of treated counties is displayed in the upper right-hand corner for each year. Additionally, the difference between New Plants and Treated counties relates to new ethanol biorefineries that did not qualify as a treatment event for this research. These non-treatment event new plants are displayed on the map as deep pink location points only for the period where they supposedly go operational.

4.10 Data Sources and Links

The data used in this research was acquired from multiple sources. This section will provide a brief overview of the data and how it is used while a complete list of the data sources and links is provided in Table 4.21.

The data used in the Mahalanobis distance metric calculation comes from four sources: Bureau of Economic Analysis (BEA), Federal Reserve Economic Data (FRED), Census, and United States Department of Agriculture – National Agricultural Statistics Service (USDA-NASS). All county level data are setup in a cross-sectional (county) and longitudinal format (time: 1995 to 2015). The following county-level variables were accessed from BEA: personal earnings, per capita earnings, population, farm earnings, manufacturing earnings, and retail earnings. All nominal value data are converted to real data using CPI_{2010} obtained from FRED. In particular, personal earnings and per capita earnings are deflated by CPI_{2010} to produce real personal earnings and real per capita earnings. All growth rate data are calculated using the standard growth rate equation (growth rate = $Value_{t+1}/Value_t - 1$). Farm share is calculated as farm earnings in the numerator with the sum of farm earnings, manufacturing earnings and retail earnings in the denominator. Manufacture share and retail share are calculated in a similar fashion. County population density is calculated from the BEA county-level population data divided by the 2010 Census value for the county-level area. Agricultural data on corn and soybean production at a county-level were retrieved from the USDA-NASS database. All the selection variable data was organized into 13 separate csv files for use in the Mahalanobis distance metric R programs for matching on the Midwest region (R program in Appendix 1) and for matching in-state (R program in Appendix 2).

4.11 Tables in Chapter 4

Table 4.1 U.S. Midwest States: Number of Counties, Corn Arces, & Biorefineries

Major Corn Producing States in U.S. Midwest			
State	**Number of Counties**	**Planted Corn (Acres, 2015)**	**Biorefineries (2015)**
Iowa	99	13,600,000	41
Illinois	102	11,700,000	14
Nebraska	93	9,300,000	25
Minnesota	87	8,500,000	21
Indiana	92	5,800,000	14
South Dakota	66	5,200,000	15
Wisconsin	72	4,100,000	9
Kansas	105	4,050,000	11
Ohio	88	3,500,000	7
Missouri	114	3,300,000	6
North Dakota	53	2,700,000	5
Michigan	83	2,450,000	5
Totals	**1054**	**74,200,000**	**173**

Data Sources: USDA-NASS (counties and planted arces) and RFA (biorefineries)

Table 4.2 Midwest Best Match Treated-Control County Pairs

	Treated		Control			Treated		Control	
ProdYear	**County**	**State**	**County**	**State**	**ProdYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Montgomery*	IA	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Clay*	NE	2007	*Gage*	NE	*Nodaway*	MO
2003	*Brookings*	SD	*Deuel*	SD	2007	*Madison*	NE	*Jefferson*	NE
2003	*Turner*	SD	*Hutchinson*	SD	2007	*Perkins*	NE	*Frontier*	NE
2003	*Winnebago*	WI	*La Crosse*	WI	2007	*Valley*	NE	*Howard*	NE
2004	*Crawford*	IL	*Fulton*	IN	2007	*McLean*	ND	*Box Butte*	NE
2004	*Cerro Gordo*	IA	*Poweshiek*	IA	2007	*Stark*	ND	*Barton*	KS
2004	*Des Moines*	IA	*Lee*	IA	2007	*Spink*	SD	*Clark*	SD
2004	*Osceola*	IA	*Watonwan*	MN	2007	*Juneau*	WI	*Adams*	WI
2004	*Worth*	IA	*Audubon*	IA	2007	*Rock*	WI	*LaPorte*	IN
2004	*Gove*	KS	*Florence*	WI	2008	*Henry*	IL	*Knox*	IL
2004	*Hitchcock*	NE	*Harding*	SD	2008	*Putnam*	IL	*Osage*	MO
2004	*Merrick*	NE	*Howard*	NE	2008	*St. Clair*	IL	*St. Charles*	MO
2004	*Lincoln*	SD	*Chisago*	MN	2008	*Madison*	IN	*Columbiana*	OH
2005	*Crawford*	IA	*Carroll*	IA	2008	*Randolph*	IN	*Fulton*	IN
2005	*Webster*	IA	*Boone*	IA	2008	*Wabash*	IN	*Huntington*	IN
2005	*Wright*	IA	*Humboldt*	IA	2008	*Wells*	IN	*Huntington*	IN
2005	*Anderson*	KS	*Crawford*	KS	2008	*Butler*	IA	*Franklin*	IA
2005	*Blue Earth*	MN	*Brown*	MN	2008	*Delaware*	IA	*Winneshiek*	IA
2005	*Chippewa*	MN	*Brown*	MN	2008	*Dickinson*	IA	*Dickinson*	MI
2005	*Kandiyohi*	MN	*Goodhue*	MN	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Saline*	MO	*Clark*	IL	2008	*O'Brien*	IA	*Emmet*	IA
2005	*Columbia*	WI	*Oconto*	WI	2008	*Plymouth*	IA	*Monona*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Franklin*	IL
2006	*Fayette*	IA	*Bremer*	IA	2008	*Rice*	KS	*Forest*	WI
2006	*Hamilton*	IA	*Warren*	IL	2008	*Otter Tail*	MN	*Pope*	MN
2006	*Story*	IA	*Poweshiek*	IA	2008	*Carroll*	MO	*Montgomery*	IA
2006	*Phillips*	KS	*Rooks*	KS	2008	*Furnas*	NE	*Norton*	KS
2006	*Calhoun*	MI	*Ashtabula*	OH	2008	*Hall*	NE	*Clayton*	IA
2006	*Audrain*	MO	*Chariton*	MO	2008	*Holt*	NE	*Cloud*	KS
2006	*Dawson*	NE	*Custer*	NE	2008	*Morrill*	NE	*Smith*	KS
2006	*Davison*	SD	*Yankton*	SD	2008	*Cass*	ND	*Barnes*	ND
2006	*Dunn*	WI	*Polk*	WI	2008	*Richland*	ND	*Wilkin*	MN
2007	*Ogle*	IL	*Kankakee*	IL	2008	*Darke*	OH	*Mercer*	OH
2007	*Cass*	IN	*Adams*	IN	2008	*Marion*	OH	*Morrow*	OH
2007	*Grant*	IN	*Delaware*	IN	2008	*Putnam*	OH	*Fulton*	OH
2007	*Jasper*	IN	*Pulaski*	IN	2008	*Seneca*	OH	*Crawford*	OH
2007	*Jay*	IN	*Huntington*	IN	2008	*Edmunds*	SD	*Clark*	SD
2007	*Adams*	IA	*Worth*	MO	2008	*Jefferson*	WI	*Knox*	OH
2007	*Floyd*	IA	*Bremer*	IA	2009	*Ford*	IL	*Poweshiek*	IA
2007	*Fremont*	IA	*Page*	IA	2009	*Chickasaw*	IA	*Bremer*	IA
2007	*Seward*	KS	*Ford*	KS	2009	*Greene*	IA	*Boone*	IA
2007	*Lenawee*	MI	*Ashtabula*	OH	2009	*Pottawattamie*	IA	*Cass*	IA
2007	*St. Clair*	MI	*Licking*	OH	2009	*Redwood*	MN	*Murray*	MN
2007	*Jackson*	MN	*Watonwan*	MN	2009	*Waseca*	MN	*McLeod*	MN
2007	*Buchanan*	MO	*Douglas*	KS	2010	*Madison*	IL	*St. Charles*	MO
2007	*Boone*	NE	*Webster*	NE	2011	*Putnam*	IN	*Parke*	IN
2007	*Buffalo*	NE	*Custer*	NE	2015	*Stutsman*	ND	*Barnes*	ND
2007	*Dakota*	NE	*Osceola*	MI					

Table 4.3 Midwest Population Match Treated-Control County Pairs

	Treated		Control			Treated		Control	
ProdYear	**County**	**State**	**County**	**State**	**ProdYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Mitchell*	KS	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Clay*	NE	2007	*Gage*	NE	*Nodaway*	MO
2003	*Brookings*	SD	*Morgan*	IL	2007	*Madison*	NE	*Douglas*	MN
2003	*Turner*	SD	*McCook*	SD	2007	*Perkins*	NE	*Frontier*	NE
2003	*Winnebago*	WI	*Outagamie*	WI	2007	*Valley*	NE	*Harlan*	NE
2004	*Crawford*	IL	*Fulton*	IN	2007	*McLean*	ND	*Mercer*	ND
2004	*Cerro Gordo*	IA	*Marshall*	IA	2007	*Stark*	ND	*Lawrence*	SD
2004	*Des Moines*	IA	*Lee*	IA	2007	*Spink*	SD	*Burt*	NE
2004	*Osceola*	IA	*Adair*	IA	2007	*Juneau*	WI	*Mille Lacs*	MN
2004	*Worth*	IA	*Van Buren*	IA	2007	*Rock*	WI	*Berrien*	MI
2004	*Gove*	KS	*Logan*	KS	2008	*Henry*	IL	*Knox*	IL
2004	*Hitchcock*	NE	*Clark*	KS	2008	*Putnam*	IL	*Sullivan*	MO
2004	*Merrick*	NE	*Clay*	NE	2008	*St. Clair*	IL	*Mahoning*	OH
2004	*Lincoln*	SD	*Insanti*	MN	2008	*Madison*	IN	*Clark*	OH
2005	*Crawford*	IA	*Poweshiek*	IA	2008	*Randolph*	IN	*Decatur*	IN
2005	*Webster*	IA	*Morgan*	IL	2008	*Wabash*	IN	*Adams*	IN
2005	*Wright*	IA	*Winnebago*	IA	2008	*Wells*	IN	*Van Wert*	OH
2005	*Anderson*	KS	*Dawes*	NE	2008	*Butler*	IA	*Newton*	IN
2005	*Blue Earth*	MN	*Rice*	MN	2008	*Delaware*	IA	*Clayton*	IA
2005	*Chippewa*	MN	*Union*	IA	2008	*Dickinson*	IA	*Clay*	IA
2005	*Kandiyohi*	MN	*Goodhue*	MN	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Saline*	MO	*Perry*	IL	2008	*O'Brien*	IA	*Clay*	IA
2005	*Columbia*	WI	*Newton*	MO	2008	*Plymouth*	IA	*Todd*	MN
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Thayer*	NE
2006	*Fayette*	IA	*Buchanan*	IA	2008	*Rice*	KS	*Forest*	WI
2006	*Hamilton*	IA	*Carroll*	IL	2008	*Otter Tail*	MN	*Waupaca*	WI
2006	*Story*	IA	*Platte*	MO	2008	*Carroll*	MO	*Howard*	MO
2006	*Phillips*	KS	*Rooks*	KS	2008	*Furnas*	NE	*Norton*	KS
2006	*Calhoun*	MI	*Richland*	OH	2008	*Hall*	NE	*Winona*	MN
2006	*Audrain*	MO	*Van Wert*	OH	2008	*Holt*	NE	*Emmet*	IA
2006	*Dawson*	NE	*Jackson*	IA	2008	*Morrill*	NE	*Polk*	NE
2006	*Davison*	SD	*Yankton*	SD	2008	*Cass*	ND	*Olmsted*	MN
2006	*Dunn*	WI	*Polk*	WI	2008	*Richland*	ND	*Kanabec*	MN
2007	*Ogle*	IL	*Whiteside*	IL	2008	*Darke*	OH	*Logan*	OH
2007	*Cass*	IN	*Huntington*	IN	2008	*Marion*	OH	*Wayne*	IN
2007	*Grant*	IN	*Shiawassee*	MI	2008	*Putnam*	OH	*Williams*	OH
2007	*Jasper*	IN	*Lafayette*	MO	2008	*Seneca*	OH	*Ashland*	OH
2007	*Jay*	IN	*Fulton*	IN	2008	*Edmunds*	SD	*Clark*	SD
2007	*Adams*	IA	*Ringgold*	IA	2008	*Jefferson*	WI	*Muskingum*	OH
2007	*Floyd*	IA	*Poweshiek*	IA	2009	*Ford*	IL	*Madison*	IA
2007	*Fremont*	IA	*Jefferson*	NE	2009	*Chickasaw*	IA	*Union*	IA
2007	*Seward*	KS	*Osceola*	MI	2009	*Greene*	IA	*Howard*	IA
2007	*Lenawee*	MI	*Ashtabula*	OH	2009	*Pottawattamie*	IA	*LaPorte*	IN
2007	*St. Clair*	MI	*Licking*	OH	2009	*Redwood*	MN	*Clay*	IA
2007	*Jackson*	MN	*Watonwan*	MN	2009	*Waseca*	MN	*Fountain*	IN
2007	*Buchanan*	MO	*Grand Traverse*	MI	2010	*Madison*	IL	*Trumbull*	OH
2007	*Boone*	NE	*Polk*	NE	2011	*Putnam*	IN	*Perry*	OH
2007	*Buffalo*	NE	*Branch*	MI	2015	*Stutsman*	ND	*Nobles*	MN
2007	*Dakota*	NE	*Owen*	IN					

Table 4.4 Midwest RUCC Match Treated-Control County Pairs

	Treated		Control	
ProdYear	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Cloud*	KS
2003	*Kearney*	NE	*Phelps*	NE
2003	*Brookings*	SD	*Wapello*	IA
2003	*Turner*	SD	*McCook*	SD
2003	*Winnebago*	WI	*La Crosse*	WI
2004	*Crawford*	IL	*Fulton*	IN
2004	*Cerro Gordo*	IA	*Wapello*	IA
2004	*Des Moines*	IA	*Lee*	IA
2004	*Osceola*	IA	*Emmet*	IA
2004	*Worth*	IA	*Lac qui Parle*	MN
2004	*Gove*	KS	*Logan*	NE
2004	*Hitchcock*	NE	*Webster*	NE
2004	*Merrick*	NE	*Howard*	NE
2004	*Lincoln*	SD	*Grundy*	IA
2005	*Crawford*	IA	*Carroll*	IA
2005	*Webster*	IA	*McDonough*	IL
2005	*Wright*	IA	*Humboldt*	IA
2005	*Anderson*	KS	*Henry*	MO
2005	*Blue Earth*	MN	*Nicollet*	MN
2005	*Chippewa*	MN	*Barton*	KS
2005	*Kandiyohi*	MN	*Goodhue*	MN
2005	*Saline*	MO	*Clark*	IL
2005	*Columbia*	WI	*Oconto*	WI
2006	*Buena Vista*	IA	*Carroll*	IA
2006	*Fayette*	IA	*Buchanan*	IA
2006	*Hamilton*	IA	*Warren*	IL
2006	*Story*	IA	*Grundy*	IA
2006	*Phillips*	KS	*Neosho*	KS
2006	*Calhoun*	MI	*Richland*	OH
2006	*Audrain*	MO	*Highland*	OH
2006	*Dawson*	NE	*Custer*	NE
2006	*Davison*	SD	*Yankton*	SD
2006	*Dunn*	WI	*Polk*	WI
2007	*Ogle*	IL	*Walworth*	WI
2007	*Cass*	IN	*Henry*	IN
2007	*Grant*	IN	*Henry*	IN
2007	*Jasper*	IN	*Lafayette*	MO
2007	*Jay*	IN	*Huntington*	IN
2007	*Adams*	IA	*Worth*	MO
2007	*Floyd*	IA	*Poweshiek*	IA
2007	*Fremont*	IA	*Morris*	KS
2007	*Seward*	KS	*Ford*	KS
2007	*Lenawee*	MI	*Ionia*	MI
2007	*St. Clair*	MI	*Licking*	OH
2007	*Jackson*	MN	*Nobles*	MN
2007	*Buchanan*	MO	*Douglas*	KS
2007	*Boone*	NE	*Webster*	NE
2007	*Buffalo*	NE	*Wayne*	OH
2007	*Dakota*	NE	*Owen*	IN
2007	*Fillmore*	NE	*Clay*	NE
2007	*Gage*	NE	*Nodaway*	MO
2007	*Madison*	NE	*Wayne*	IN
2007	*Perkins*	NE	*Frontier*	NE
2007	*Valley*	NE	*Kiowa*	KS
2007	*McLean*	ND	*Dade*	MO
2007	*Stark*	ND	*Barton*	KS
2007	*Spink*	SD	*Clark*	SD
2007	*Juneau*	WI	*Wadena*	MN
2007	*Rock*	WI	*LaPorte*	IN
2008	*Henry*	IL	*Woodford*	IL
2008	*Putnam*	IL	*Elk*	KS
2008	*St. Clair*	IL	*St. Charles*	MO
2008	*Madison*	IN	*Porter*	IN
2008	*Randolph*	IN	*Miami*	IN
2008	*Wabash*	IN	*Huntington*	IN
2008	*Wells*	IN	*Whitley*	IN
2008	*Butler*	IA	*Warren*	IN
2008	*Delaware*	IA	*Allamakee*	IA
2008	*Dickinson*	IA	*Clay*	IA
2008	*Mitchell*	IA	*Humboldt*	IA
2008	*O'Brien*	IA	*Emmet*	IA
2008	*Plymouth*	IA	*Woodbury*	IA
2008	*Republic*	KS	*Thayer*	NE
2008	*Rice*	KS	*Linn*	MO
2008	*Otter Tail*	MN	*Barron*	WI
2008	*Carroll*	MO	*Montgomery*	IA
2008	*Furnas*	NE	*Knox*	NE
2008	*Hall*	NE	*Dubuque*	IA
2008	*Holt*	NE	*Cloud*	KS
2008	*Morrill*	NE	*Smith*	KS
2008	*Cass*	ND	*Clay*	MN
2008	*Richland*	ND	*Wilkin*	MN
2008	*Darke*	OH	*Rush*	IN
2008	*Marion*	OH	*Crawford*	OH
2008	*Putnam*	OH	*Paulding*	OH
2008	*Seneca*	OH	*Crawford*	OH
2008	*Edmunds*	SD	*Clark*	SD
2008	*Jefferson*	WI	*Knox*	OH
2009	*Ford*	IL	*Washington*	IA
2009	*Chickasaw*	IA	*Howard*	IA
2009	*Greene*	IA	*Boone*	IA
2009	*Pottawattamie*	IA	*Cass*	NE
2009	*Redwood*	MN	*Clay*	IA
2009	*Waseca*	MN	*McLeod*	MN
2010	*Madison*	IL	*St. Charles*	MO
2011	*Putnam*	IN	*Hocking*	OH
2015	*Stutsman*	ND	*Nobles*	MN

Table 4.5 State Best Match Treated-Control County Pairs

	Treated		Control			Treated		Control	
ProdYear	**County**	**State**	**County**	**State**	**ProdYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Gregory*	SD	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Phelps*	NE	2007	*Gage*	NE	*Jefferson*	NE
2003	*Brookings*	SD	*Yankton*	SD	2007	*Madison*	NE	*Knox*	NE
2003	*Turner*	SD	*Miner*	SD	2007	*Perkins*	NE	*Frontier*	NE
2003	*Winnebago*	WI	*Outagamie*	WI	2007	*Valley*	NE	*Webster*	NE
2004	*Crawford*	IL	*Massac*	IL	2007	*McLean*	ND	*McHenry*	ND
2004	*Cerro Gordo*	IA	*Marshall*	IA	2007	*Stark*	ND	*Pierce*	ND
2004	*Des Moines*	IA	*Lee*	IA	2007	*Spink*	SD	*Todd*	SD
2004	*Osceola*	IA	*Lyon*	IA	2007	*Juneau*	WI	*Marquette*	WI
2004	*Worth*	IA	*Winnebago*	IA	2007	*Rock*	WI	*Lincoln*	WI
2004	*Gove*	KS	*Smith*	KS	2008	*Henry*	IL	*Knox*	IL
2004	*Hitchcock*	NE	*Webster*	NE	2008	*Putnam*	IL	*Union*	IL
2004	*Merrick*	NE	*Polk*	NE	2008	*St. Clair*	IL	*McHenry*	IL
2004	*Lincoln*	SD	*Hamlin*	SD	2008	*Madison*	IN	*Delaware*	IN
2005	*Crawford*	IA	*Marion*	IA	2008	*Randolph*	IN	*DeKalb*	IN
2005	*Webster*	IA	*Boone*	IA	2008	*Wabash*	IN	*Huntington*	IN
2005	*Wright*	IA	*Humboldt*	IA	2008	*Wells*	IN	*Adams*	IN
2005	*Anderson*	KS	*Crawford*	KS	2008	*Butler*	IA	*Franklin*	IA
2005	*Blue Earth*	MN	*Brown*	MN	2008	*Delaware*	IA	*Bremer*	IA
2005	*Chippewa*	MN	*Brown*	MN	2008	*Dickinson*	IA	*Humboldt*	IA
2005	*Kandiyohi*	MN	*Rice*	MN	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Saline*	MO	*Chariton*	MO	2008	*O'Brien*	IA	*Emmet*	IA
2005	*Columbia*	WI	*Green Lake*	WI	2008	*Plymouth*	IA	*Lyon*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Ness*	KS
2006	*Fayette*	IA	*Grundy*	IA	2008	*Rice*	KS	*Rooks*	KS
2006	*Hamilton*	IA	*Grundy*	IA	2008	*Otter Tail*	MN	*Rice*	MN
2006	*Story*	IA	*Grundy*	IA	2008	*Carroll*	MO	*Randolph*	MO
2006	*Phillips*	KS	*Rooks*	KS	2008	*Furnas*	NE	*Red Willow*	NE
2006	*Calhoun*	MI	*Jackson*	MI	2008	*Hall*	NE	*Scotts Bluff*	NE
2006	*Audrain*	MO	*Ray*	MO	2008	*Holt*	NE	*Butler*	NE
2006	*Dawson*	NE	*Colfax*	NE	2008	*Morrill*	NE	*Sheridan*	NE
2006	*Davison*	SD	*Yankton*	SD	2008	*Cass*	ND	*Adams*	ND
2006	*Dunn*	WI	*Taylor*	WI	2008	*Richland*	ND	*McHenry*	ND
2007	*Ogle*	IL	*DeKalb*	IL	2008	*Darke*	OH	*Miami*	OH
2007	*Cass*	IN	*Lawrence*	IN	2008	*Marion*	OH	*Crawford*	OH
2007	*Grant*	IN	*Delaware*	IN	2008	*Putnam*	OH	*Tuscarawas*	OH
2007	*Jasper*	IN	*Pulaski*	IN	2008	*Seneca*	OH	*Ashland*	OH
2007	*Jay*	IN	*Adams*	IN	2008	*Edmunds*	SD	*McPherson*	SD
2007	*Adams*	IA	*Humboldt*	IA	2008	*Jefferson*	WI	*Waupaca*	WI
2007	*Floyd*	IA	*Bremer*	IA	2009	*Ford*	IL	*Edgar*	IL
2007	*Fremont*	IA	*Page*	IA	2009	*Chickasaw*	IA	*Bremer*	IA
2007	*Seward*	KS	*Ford*	KS	2009	*Greene*	IA	*Boone*	IA
2007	*Lenawee*	MI	*Hillsdale*	MI	2009	*Pottawattamie*	IA	*Cass*	IA
2007	*St. Clair*	MI	*Lapeer*	MI	2009	*Redwood*	MN	*Meeker*	MN
2007	*Jackson*	MN	*Murray*	MN	2009	*Waseca*	MN	*McLeod*	MN
2007	*Buchanan*	MO	*St. Francois*	MO	2010	*Madison*	IL	*McHenry*	IL
2007	*Boone*	NE	*Nance*	NE	2011	*Putnam*	IN	*Parke*	IN
2007	*Buffalo*	NE	*Red Willow*	NE	2015	*Stutsman*	ND	*Foster*	ND
2007	*Dakota*	NE	*Thayer*	NE					

Table 4.6 State Population Match Treated-Control County Pairs

	Treated		Control	
ProdYear	County	State	County	State
2002	*Grant*	SD	*Butte*	SD
2003	*Kearney*	NE	*Antelope*	NE
2003	*Brookings*	SD	*Lawrence*	SD
2003	*Turner*	SD	*Custer*	SD
2003	*Winnebago*	WI	*Outagamie*	WI
2004	*Crawford*	IL	*Hancock*	IL
2004	*Cerro Gordo*	IA	*Marshall*	IA
2004	*Des Moines*	IA	*Marshall*	IA
2004	*Osceola*	IA	*Adair*	IA
2004	*Worth*	IA	*Pocahontas*	IA
2004	*Gove*	KS	*Logan*	KS
2004	*Hitchcock*	NE	*Webster*	NE
2004	*Merrick*	NE	*Jefferson*	NE
2004	*Lincoln*	SD	*Meade*	SD
2005	*Crawford*	IA	*Poweshiek*	IA
2005	*Webster*	IA	*Jasper*	IA
2005	*Wright*	IA	*Lyon*	IA
2005	*Anderson*	KS	*Greenwood*	KS
2005	*Blue Earth*	MN	*Crow Wing*	MN
2005	*Chippewa*	MN	*Koochiching*	MN
2005	*Kandiyohi*	MN	*Itasca*	MN
2005	*Saline*	MO	*Henry*	MO
2005	*Columbia*	WI	*Sauk*	WI
2006	*Buena Vista*	IA	*Carroll*	IA
2006	*Fayette*	IA	*Buchanan*	IA
2006	*Hamilton*	IA	*Allamakee*	IA
2006	*Story*	IA	*Dubuque*	IA
2006	*Phillips*	KS	*Rooks*	KS
2006	*Calhoun*	MI	*Jackson*	MI
2006	*Audrain*	MO	*Adair*	MO
2006	*Dawson*	NE	*Scotts Bluff*	NE
2006	*Davison*	SD	*Yankton*	SD
2006	*Dunn*	WI	*Monroe*	WI
2007	*Ogle*	IL	*Marion*	IL
2007	*Cass*	IN	*Steuben*	IN
2007	*Grant*	IN	*Howard*	IN
2007	*Jasper*	IN	*Washington*	IN
2007	*Jay*	IN	*Orange*	IN
2007	*Adams*	IA	*Ringgold*	IA
2007	*Floyd*	IA	*Poweshiek*	IA
2007	*Fremont*	IA	*Adair*	IA
2007	*Seward*	KS	*Franklin*	KS
2007	*Lenawee*	MI	*Midland*	MI
2007	*St. Clair*	MI	*Jackson*	MI
2007	*Jackson*	MN	*Murray*	MN
2007	*Buchanan*	MO	*Jasper*	MO
2007	*Boone*	NE	*Clay*	NE
2007	*Buffalo*	NE	*Scotts Bluff*	NE
2007	*Dakota*	NE	*Red Willow*	NE

	Treated		Control	
ProdYear	County	State	County	State
2007	*Fillmore*	NE	*Clay*	NE
2007	*Gage*	NE	*Saunders*	NE
2007	*Madison*	NE	*Scotts Bluff*	NE
2007	*Perkins*	NE	*Frontier*	NE
2007	*Valley*	NE	*Webster*	NE
2007	*McLean*	ND	*Mercer*	ND
2007	*Stark*	ND	*Williams*	ND
2007	*Spink*	SD	*Fall River*	SD
2007	*Juneau*	WI	*Lincoln*	WI
2007	*Rock*	WI	*Racine*	WI
2008	*Henry*	IL	*Knox*	IL
2008	*Putnam*	IL	*Pulaski*	IL
2008	*St. Clair*	IL	*McHenry*	IL
2008	*Madison*	IN	*Johnson*	IN
2008	*Randolph*	IN	*Decatur*	IN
2008	*Wabash*	IN	*Huntington*	IN
2008	*Wells*	IN	*Decatur*	IN
2008	*Butler*	IA	*Cass*	IA
2008	*Delaware*	IA	*Jackson*	IA
2008	*Dickinson*	IA	*Clay*	IA
2008	*Mitchell*	IA	*Humboldt*	IA
2008	*O'Brien*	IA	*Page*	IA
2008	*Plymouth*	IA	*Bremer*	IA
2008	*Republic*	KS	*Barber*	KS
2008	*Rice*	KS	*Cloud*	KS
2008	*Otter Tail*	MN	*Rice*	MN
2008	*Carroll*	MO	*St. Clair*	MO
2008	*Furnas*	NE	*Sheridan*	NE
2008	*Hall*	NE	*Scotts Bluff*	NE
2008	*Holt*	NE	*Cheyenne*	NE
2008	*Morrill*	NE	*Sheridan*	NE
2008	*Cass*	ND	*Burleigh*	ND
2008	*Richland*	ND	*Williams*	ND
2008	*Darke*	OH	*Ashland*	OH
2008	*Marion*	OH	*Knox*	OH
2008	*Putnam*	OH	*Williams*	OH
2008	*Seneca*	OH	*Ashland*	OH
2008	*Edmunds*	SD	*Gregory*	SD
2008	*Jefferson*	WI	*Wood*	WI
2009	*Ford*	IL	*Marshall*	IL
2009	*Chickasaw*	IA	*Cass*	IA
2009	*Greene*	IA	*Keokuk*	IA
2009	*Pottawattamie*	IA	*Black Hawk*	IA
2009	*Redwood*	MN	*Meeker*	MN
2009	*Waseca*	MN	*Kanabec*	MN
2010	*Madison*	IL	*Winnebago*	IL
2011	*Putnam*	IN	*Dearborn*	IN
2015	*Stutsman*	ND	*Walsh*	ND

Table 4.7 State RUCC Match Treated-Control County Pairs

	Treated		Control			Treated		Control	
ProdYear	County	State	County	State	ProdYear	County	State	County	State
2002	*Grant*	SD	*Yankton*	SD	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Phelps*	NE	2007	*Gage*	NE	*Saline*	NE
2003	*Brookings*	SD	*Butte*	SD	2007	*Madison*	NE	*Scotts Bluff*	NE
2003	*Turner*	SD	*McCook*	SD	2007	*Perkins*	NE	*Frontier*	NE
2003	*Winnebago*	WI	*Outagamie*	WI	2007	*Valley*	NE	*Webster*	NE
2004	*Crawford*	IL	*Massac*	IL	2007	*McLean*	ND	*Kidder*	ND
2004	*Cerro Gordo*	IA	*Wapello*	IA	2007	*Stark*	ND	*Williams*	ND
2004	*Des Moines*	IA	*Lee*	IA	2007	*Spink*	SD	*Todd*	SD
2004	*Osceola*	IA	*Emmet*	IA	2007	*Juneau*	WI	*Sawyer*	WI
2004	*Worth*	IA	*Pocahontas*	IA	2007	*Rock*	WI	*Racine*	WI
2004	*Gove*	KS	*Smith*	KS	2008	*Henry*	IL	*Woodford*	IL
2004	*Hitchcock*	NE	*Webster*	NE	2008	*Putnam*	IL	*Pulaski*	IL
2004	*Merrick*	NE	*Howard*	NE	2008	*St. Clair*	IL	*McHenry*	IL
2004	*Lincoln*	SD	*Custer*	SD	2008	*Madison*	IN	*Washington*	IN
2005	*Crawford*	IA	*Poweshiek*	IA	2008	*Randolph*	IN	*Decatur*	IN
2005	*Webster*	IA	*Wapello*	IA	2008	*Wabash*	IN	*Huntington*	IN
2005	*Wright*	IA	*Humboldt*	IA	2008	*Wells*	IN	*Whitley*	IN
2005	*Anderson*	KS	*Greenwood*	KS	2008	*Butler*	IA	*Audubon*	IA
2005	*Blue Earth*	MN	*Nicollet*	MN	2008	*Delaware*	IA	*Jackson*	IA
2005	*Chippewa*	MN	*Beltrami*	MN	2008	*Dickinson*	IA	*Humboldt*	IA
2005	*Kandiyohi*	MN	*Crow Wing*	MN	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Saline*	MO	*Howard*	MO	2008	*O'Brien*	IA	*Emmet*	IA
2005	*Columbia*	WI	*Oconto*	WI	2008	*Plymouth*	IA	*Bremer*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Ness*	KS
2006	*Fayette*	IA	*Buchanan*	IA	2008	*Rice*	KS	*Ellsworth*	KS
2006	*Hamilton*	IA	*Allamakee*	IA	2008	*Otter Tail*	MN	*Brown*	MN
2006	*Story*	IA	*Grundy*	IA	2008	*Carroll*	MO	*Randolph*	MO
2006	*Phillips*	KS	*Ellsworth*	KS	2008	*Furnas*	NE	*Harlan*	NE
2006	*Calhoun*	MI	*Jackson*	MI	2008	*Hall*	NE	*Howard*	NE
2006	*Audrain*	MO	*Randolph*	MO	2008	*Holt*	NE	*Box Butte*	NE
2006	*Dawson*	NE	*Colfax*	NE	2008	*Morrill*	NE	*Sheridan*	NE
2006	*Davison*	SD	*Yankton*	SD	2008	*Cass*	ND	*Oliver*	ND
2006	*Dunn*	WI	*Taylor*	WI	2008	*Richland*	ND	*Walsh*	ND
2007	*Ogle*	IL	*Marion*	IL	2008	*Darke*	OH	*Guernsey*	OH
2007	*Cass*	IN	*Henry*	IN	2008	*Marion*	OH	*Crawford*	OH
2007	*Grant*	IN	*Jackson*	IN	2008	*Putnam*	OH	*Guernsey*	OH
2007	*Jasper*	IN	*Washington*	IN	2008	*Seneca*	OH	*Ashland*	OH
2007	*Jay*	IN	*Adams*	IN	2008	*Edmunds*	SD	*McPherson*	SD
2007	*Adams*	IA	*Wayne*	IA	2008	*Jefferson*	WI	*Wood*	WI
2007	*Floyd*	IA	*Humboldt*	IA	2009	*Ford*	IL	*Williamson*	IL
2007	*Fremont*	IA	*Adair*	IA	2009	*Chickasaw*	IA	*Cass*	IA
2007	*Seward*	KS	*Ford*	KS	2009	*Greene*	IA	*Boone*	IA
2007	*Lenawee*	MI	*Ionia*	MI	2009	*Pottawattamie*	IA	*Jones*	IA
2007	*St. Clair*	MI	*Lapeer*	MI	2009	*Redwood*	MN	*Beltrami*	MN
2007	*Jackson*	MN	*Beltrami*	MN	2009	*Waseca*	MN	*McLeod*	MN
2007	*Buchanan*	MO	*Jasper*	MO	2010	*Madison*	IL	*McHenry*	IL
2007	*Boone*	NE	*Webster*	NE	2011	*Putnam*	IN	*Dearborn*	IN
2007	*Buffalo*	NE	*Dodge*	NE	2015	*Stutsman*	ND	*Pierce*	ND
2007	*Dakota*	NE	*Dixon*	NE					

Table 4.8 Midwest Best Match Treated-Control County Pairs, LT25k

	Treated		Control			Treated		Control	
ProdYear	**County**	**State**	**County**	**State**	**ProdYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Montgomery*	IA	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Clay*	NE	2007	*Gage*	NE	*Nodaway*	MO
2003	*Turner*	SD	*Hutchinson*	SD	2007	*Perkins*	NE	*Frontier*	NE
2004	*Crawford*	IL	*Fulton*	IN	2007	*Valley*	NE	*Howard*	NE
2004	*Osceola*	IA	*Watonwan*	MN	2007	*McLean*	ND	*Box Butte*	NE
2004	*Worth*	IA	*Audubon*	IA	2007	*Stark*	ND	*Barton*	KS
2004	*Gove*	KS	*Florence*	WI	2007	*Spink*	SD	*Clark*	SD
2004	*Hitchcock*	NE	*Harding*	SD	2008	*Putnam*	IL	*Osage*	MO
2004	*Merrick*	NE	*Howard*	NE	2008	*Butler*	IA	*Franklin*	IA
2005	*Crawford*	IA	*Carroll*	IA	2008	*Delaware*	IA	*Winneshiek*	IA
2005	*Wright*	IA	*Humboldt*	IA	2008	*Dickinson*	IA	*Dickinson*	MI
2005	*Anderson*	KS	*Crawford*	KS	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Chippewa*	MN	*Brown*	MN	2008	*O'Brien*	IA	*Emmet*	IA
2005	*Saline*	MO	*Clark*	IL	2008	*Plymouth*	IA	*Monona*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Franklin*	IL
2006	*Fayette*	IA	*Bremer*	IA	2008	*Rice*	KS	*Forest*	WI
2006	*Hamilton*	IA	*Warren*	IL	2008	*Carroll*	MO	*Montgomery*	IA
2006	*Phillips*	KS	*Rooks*	KS	2008	*Furnas*	NE	*Norton*	KS
2006	*Dawson*	NE	*Custer*	NE	2008	*Holt*	NE	*Cloud*	KS
2006	*Davison*	SD	*Yankton*	SD	2008	*Morrill*	NE	*Smith*	KS
2007	*Jay*	IN	*Huntington*	IN	2008	*Richland*	ND	*Wilkin*	MN
2007	*Adams*	IA	*Worth*	MO	2008	*Edmunds*	SD	*Clark*	SD
2007	*Floyd*	IA	*Bremer*	IA	2009	*Ford*	IL	*Poweshiek*	IA
2007	*Fremont*	IA	*Page*	IA	2009	*Chickasaw*	IA	*Bremer*	IA
2007	*Seward*	KS	*Ford*	KS	2009	*Greene*	IA	*Boone*	IA
2007	*Jackson*	MN	*Watonwan*	MN	2009	*Redwood*	MN	*Murray*	MN
2007	*Boone*	NE	*Webster*	NE	2009	*Waseca*	MN	*McLeod*	MN
2007	*Dakota*	NE	*Osceola*	MI	2015	*Stutsman*	ND	*Barnes*	ND

Table 4.9 Midwest Population Match Treated-Control County Pairs, LT25k

	Treated		Control			Treated		Control	
ProdYear	County	State	County	State	ProdYear	County	State	County	State
2002	*Grant*	SD	*Mitchell*	KS	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Clay*	NE	2007	*Gage*	NE	*Nodaway*	MO
2003	*Turner*	SD	*McCook*	SD	2007	*Perkins*	NE	*Frontier*	NE
2004	*Crawford*	IL	*Fulton*	IN	2007	*Valley*	NE	*Harlan*	NE
2004	*Osceola*	IA	*Adair*	IA	2007	*McLean*	ND	*Mercer*	ND
2004	*Worth*	IA	*Van Buren*	IA	2007	*Stark*	ND	*Lawrence*	SD
2004	*Gove*	KS	*Logan*	KS	2007	*Spink*	SD	*Burt*	NE
2004	*Hitchcock*	NE	*Clark*	KS	2008	*Putnam*	IL	*Sullivan*	MO
2004	*Merrick*	NE	*Clay*	NE	2008	*Butler*	IA	*Newton*	IN
2005	*Crawford*	IA	*Poweshiek*	IA	2008	*Delaware*	IA	*Clayton*	IA
2005	*Wright*	IA	*Winnebago*	IA	2008	*Dickinson*	IA	*Clay*	IA
2005	*Anderson*	KS	*Dawes*	NE	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Chippewa*	MN	*Union*	IA	2008	*O'Brien*	IA	*Clay*	IA
2005	*Saline*	MO	*Perry*	IL	2008	*Plymouth*	IA	*Todd*	MN
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Thayer*	NE
2006	*Fayette*	IA	*Buchanan*	IA	2008	*Rice*	KS	*Forest*	WI
2006	*Hamilton*	IA	*Carroll*	IL	2008	*Carroll*	MO	*Howard*	MO
2006	*Phillips*	KS	*Rooks*	KS	2008	*Furnas*	NE	*Norton*	KS
2006	*Dawson*	NE	*Jackson*	IA	2008	*Holt*	NE	*Emmet*	IA
2006	*Davison*	SD	*Yankton*	SD	2008	*Morrill*	NE	*Polk*	NE
2007	*Jay*	IN	*Fulton*	IN	2008	*Richland*	ND	*Kanabec*	MN
2007	*Adams*	IA	*Ringgold*	IA	2008	*Edmunds*	SD	*Clark*	SD
2007	*Floyd*	IA	*Poweshiek*	IA	2009	*Ford*	IL	*Madison*	IA
2007	*Fremont*	IA	*Jefferson*	NE	2009	*Chickasaw*	IA	*Union*	IA
2007	*Seward*	KS	*Osceola*	MI	2009	*Greene*	IA	*Howard*	IA
2007	*Jackson*	MN	*Watonwan*	MN	2009	*Redwood*	MN	*Clay*	IA
2007	*Boone*	NE	*Polk*	NE	2009	*Waseca*	MN	*Fountain*	IN
2007	*Dakota*	NE	*Owen*	IN	2015	*Stutsman*	ND	*Nobles*	MN

Table 4.10 Midwest RUCC Match Treated-Control County Pairs, LT25k

	Treated		Control			Treated		Control	
ProdYear	**County**	**State**	**County**	**State**	**ProdYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Cloud*	KS	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Phelps*	NE	2007	*Gage*	NE	*Nodaway*	MO
2003	*Turner*	SD	*McCook*	SD	2007	*Perkins*	NE	*Frontier*	NE
2004	*Crawford*	IL	*Fulton*	IN	2007	*Valley*	NE	*Kiowa*	KS
2004	*Osceola*	IA	*Emmet*	IA	2007	*McLean*	ND	*Dade*	MO
2004	*Worth*	IA	*Lac qui Parle*	MN	2007	*Stark*	ND	*Barton*	KS
2004	*Gove*	KS	*Logan*	NE	2007	*Spink*	SD	*Clark*	SD
2004	*Hitchcock*	NE	*Webster*	NE	2008	*Putnam*	IL	*Elk*	KS
2004	*Merrick*	NE	*Howard*	NE	2008	*Butler*	IA	*Warren*	IN
2005	*Crawford*	IA	*Carroll*	IA	2008	*Delaware*	IA	*Allamakee*	IA
2005	*Wright*	IA	*Humboldt*	IA	2008	*Dickinson*	IA	*Clay*	IA
2005	*Anderson*	KS	*Henry*	MO	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Chippewa*	MN	*Barton*	KS	2008	*O'Brien*	IA	*Emmet*	IA
2005	*Saline*	MO	*Clark*	IL	2008	*Plymouth*	IA	*Woodbury*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Thayer*	NE
2006	*Fayette*	IA	*Buchanan*	IA	2008	*Rice*	KS	*Linn*	MO
2006	*Hamilton*	IA	*Warren*	IL	2008	*Carroll*	MO	*Montgomery*	IA
2006	*Phillips*	KS	*Neosho*	KS	2008	*Furnas*	NE	*Knox*	NE
2006	*Dawson*	NE	*Custer*	NE	2008	*Holt*	NE	*Cloud*	KS
2006	*Davison*	SD	*Yankton*	SD	2008	*Morrill*	NE	*Smith*	KS
2007	*Jay*	IN	*Huntington*	IN	2008	*Richland*	ND	*Wilkin*	MN
2007	*Adams*	IA	*Worth*	MO	2008	*Edmunds*	SD	*Clark*	SD
2007	*Floyd*	IA	*Poweshiek*	IA	2009	*Ford*	IL	*Washington*	IA
2007	*Fremont*	IA	*Morris*	KS	2009	*Chickasaw*	IA	*Howard*	IA
2007	*Seward*	KS	*Ford*	KS	2009	*Greene*	IA	*Boone*	IA
2007	*Jackson*	MN	*Nobles*	MN	2009	*Redwood*	MN	*Clay*	IA
2007	*Boone*	NE	*Webster*	NE	2009	*Waseca*	MN	*McLeod*	MN
2007	*Dakota*	NE	*Owen*	IN	2015	*Stutsman*	ND	*Nobles*	MN

Table 4.11 State Best Match Treated-Control County Pairs, LT25k

	Treated		Control			Treated		Control	
TreatYear	**County**	**State**	**County**	**State**	**TreatYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Gregory*	SD	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Phelps*	NE	2007	*Gage*	NE	*Jefferson*	NE
2003	*Turner*	SD	*Miner*	SD	2007	*Perkins*	NE	*Frontier*	NE
2004	*Crawford*	IL	*Massac*	IL	2007	*Valley*	NE	*Webster*	NE
2004	*Osceola*	IA	*Lyon*	IA	2007	*McLean*	ND	*McHenry*	ND
2004	*Worth*	IA	*Winnebago*	IA	2007	*Stark*	ND	*Pierce*	ND
2004	*Gove*	KS	*Smith*	KS	2007	*Spink*	SD	*Todd*	SD
2004	*Hitchcock*	NE	*Webster*	NE	2008	*Putnam*	IL	*Union*	IL
2004	*Merrick*	NE	*Polk*	NE	2008	*Butler*	IA	*Franklin*	IA
2005	*Crawford*	IA	*Marion*	IA	2008	*Delaware*	IA	*Bremer*	IA
2005	*Wright*	IA	*Humboldt*	IA	2008	*Dickinson*	IA	*Humboldt*	IA
2005	*Anderson*	KS	*Crawford*	KS	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Chippewa*	MN	*Brown*	MN	2008	*O'Brien*	IA	*Emmet*	IA
2005	*Saline*	MO	*Chariton*	MO	2008	*Plymouth*	IA	*Lyon*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Ness*	KS
2006	*Fayette*	IA	*Grundy*	IA	2008	*Rice*	KS	*Rooks*	KS
2006	*Hamilton*	IA	*Grundy*	IA	2008	*Carroll*	MO	*Randolph*	MO
2006	*Phillips*	KS	*Rooks*	KS	2008	*Furnas*	NE	*Red Willow*	NE
2006	*Dawson*	NE	*Colfax*	NE	2008	*Holt*	NE	*Butler*	NE
2006	*Davison*	SD	*Yankton*	SD	2008	*Morrill*	NE	*Sheridan*	NE
2007	*Jay*	IN	*Adams*	IN	2008	*Richland*	ND	*McHenry*	ND
2007	*Adams*	IA	*Humboldt*	IA	2008	*Edmunds*	SD	*McPherson*	SD
2007	*Floyd*	IA	*Bremer*	IA	2009	*Ford*	IL	*Edgar*	IL
2007	*Fremont*	IA	*Page*	IA	2009	*Chickasaw*	IA	*Bremer*	IA
2007	*Seward*	KS	*Ford*	KS	2009	*Greene*	IA	*Boone*	IA
2007	*Jackson*	MN	*Murray*	MN	2009	*Redwood*	MN	*Meeker*	MN
2007	*Boone*	NE	*Nance*	NE	2009	*Waseca*	MN	*McLeod*	MN
2007	*Dakota*	NE	*Thayer*	NE	2015	*Stutsman*	ND	*Foster*	ND

Table 4.12 State Population Match Treated-Control County Pairs, LT25k

	Treated		Control			Treated		Control	
ProdYear	**County**	**State**	**County**	**State**	**ProdYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Butte*	SD	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Antelope*	NE	2007	*Gage*	NE	*Saunders*	NE
2003	*Turner*	SD	*Custer*	SD	2007	*Perkins*	NE	*Frontier*	NE
2004	*Crawford*	IL	*Hancock*	IL	2007	*Valley*	NE	*Webster*	NE
2004	*Osceola*	IA	*Adair*	IA	2007	*McLean*	ND	*Mercer*	ND
2004	*Worth*	IA	*Pocahontas*	IA	2007	*Stark*	ND	*Williams*	ND
2004	*Gove*	KS	*Logan*	KS	2007	*Spink*	SD	*Fall River*	SD
2004	*Hitchcock*	NE	*Webster*	NE	2008	*Putnam*	IL	*Pulaski*	IL
2004	*Merrick*	NE	*Jefferson*	NE	2008	*Butler*	IA	*Cass*	IA
2005	*Crawford*	IA	*Poweshiek*	IA	2008	*Delaware*	IA	*Jackson*	IA
2005	*Wright*	IA	*Lyon*	IA	2008	*Dickinson*	IA	*Clay*	IA
2005	*Anderson*	KS	*Greenwood*	KS	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Chippewa*	MN	*Koochiching*	MN	2008	*O'Brien*	IA	*Page*	IA
2005	*Saline*	MO	*Henry*	MO	2008	*Plymouth*	IA	*Bremer*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Barber*	KS
2006	*Fayette*	IA	*Buchanan*	IA	2008	*Rice*	KS	*Cloud*	KS
2006	*Hamilton*	IA	*Allamakee*	IA	2008	*Carroll*	MO	*St. Clair*	MO
2006	*Phillips*	KS	*Rooks*	KS	2008	*Furnas*	NE	*Sheridan*	NE
2006	*Dawson*	NE	*Scotts Bluff*	NE	2008	*Holt*	NE	*Cheyenne*	NE
2006	*Davison*	SD	*Yankton*	SD	2008	*Morrill*	NE	*Sheridan*	NE
2007	*Jay*	IN	*Orange*	IN	2008	*Richland*	ND	*Williams*	ND
2007	*Adams*	IA	*Ringgold*	IA	2008	*Edmunds*	SD	*Gregory*	SD
2007	*Floyd*	IA	*Poweshiek*	IA	2009	*Ford*	IL	*Marshall*	IL
2007	*Fremont*	IA	*Adair*	IA	2009	*Chickasaw*	IA	*Cass*	IA
2007	*Seward*	KS	*Franklin*	KS	2009	*Greene*	IA	*Keokuk*	IA
2007	*Jackson*	MN	*Murray*	MN	2009	*Redwood*	MN	*Meeker*	MN
2007	*Boone*	NE	*Clay*	NE	2009	*Waseca*	MN	*Kanabec*	MN
2007	*Dakota*	NE	*Red Willow*	NE	2015	*Stutsman*	ND	*Walsh*	ND

Table 4.13 State RUCC Match Treated-Control County Pairs, LT25k

	Treated		Control			Treated		Control	
ProdYear	**County**	**State**	**County**	**State**	**ProdYear**	**County**	**State**	**County**	**State**
2002	*Grant*	SD	*Yankton*	SD	2007	*Fillmore*	NE	*Clay*	NE
2003	*Kearney*	NE	*Phelps*	NE	2007	*Gage*	NE	*Saline*	NE
2003	*Turner*	SD	*McCook*	SD	2007	*Perkins*	NE	*Frontier*	NE
2004	*Crawford*	IL	*Massac*	IL	2007	*Valley*	NE	*Webster*	NE
2004	*Osceola*	IA	*Emmet*	IA	2007	*McLean*	ND	*Kidder*	ND
2004	*Worth*	IA	*Pocahontas*	IA	2007	*Stark*	ND	*Williams*	ND
2004	*Gove*	KS	*Smith*	KS	2007	*Spink*	SD	*Todd*	SD
2004	*Hitchcock*	NE	*Webster*	NE	2008	*Putnam*	IL	*Pulaski*	IL
2004	*Merrick*	NE	*Howard*	NE	2008	*Butler*	IA	*Audubon*	IA
2005	*Crawford*	IA	*Poweshiek*	IA	2008	*Delaware*	IA	*Jackson*	IA
2005	*Wright*	IA	*Humboldt*	IA	2008	*Dickinson*	IA	*Humboldt*	IA
2005	*Anderson*	KS	*Greenwood*	KS	2008	*Mitchell*	IA	*Humboldt*	IA
2005	*Chippewa*	MN	*Beltrami*	MN	2008	*O'Brien*	IA	*Emmet*	IA
2005	*Saline*	MO	*Howard*	MO	2008	*Plymouth*	IA	*Bremer*	IA
2006	*Buena Vista*	IA	*Carroll*	IA	2008	*Republic*	KS	*Ness*	KS
2006	*Fayette*	IA	*Buchanan*	IA	2008	*Rice*	KS	*Ellsworth*	KS
2006	*Hamilton*	IA	*Allamakee*	IA	2008	*Carroll*	MO	*Randolph*	MO
2006	*Phillips*	KS	*Ellsworth*	KS	2008	*Furnas*	NE	*Harlan*	NE
2006	*Dawson*	NE	*Colfax*	NE	2008	*Holt*	NE	*Box Butte*	NE
2006	*Davison*	SD	*Yankton*	SD	2008	*Morrill*	NE	*Sheridan*	NE
2007	*Jay*	IN	*Adams*	IN	2008	*Richland*	ND	*Walsh*	ND
2007	*Adams*	IA	*Wayne*	IA	2008	*Edmunds*	SD	*McPherson*	SD
2007	*Floyd*	IA	*Humboldt*	IA	2009	*Ford*	IL	*Williamson*	IL
2007	*Fremont*	IA	*Adair*	IA	2009	*Chickasaw*	IA	*Cass*	IA
2007	*Seward*	KS	*Ford*	KS	2009	*Greene*	IA	*Boone*	IA
2007	*Jackson*	MN	*Beltrami*	MN	2009	*Redwood*	MN	*Beltrami*	MN
2007	*Boone*	NE	*Webster*	NE	2009	*Waseca*	MN	*McLeod*	MN
2007	*Dakota*	NE	*Dixon*	NE	2015	*Stutsman*	ND	*Pierce*	ND

Table 4.14 Selection Variable Categories

Growth Rate	Earnings Growth Rate and Population Growth Rate
Spatial Structure	Population, Population Density, Corn Production, Corn Acres and Soybean Production
Economic Structure	Real Per Capita Earnings, Farm Real Cash Receipts, Employment, Farm Earnings Share, Manufacturing Earnings Share, and Retail Earnings Share

Table 4.15 Selection Variables List (Ethanol Biorefinery, 2001)

Growth Variables	
VarName	*VarDescription*
POPG9695	Population growth between 1995 to 1996
POPG9796	Population growth between 1996 to 1997
POPG9897	Population growth between 1997 to 1998
POPG9998	Population growth between 1998 to 1999
POPG0099	Population growth between 1999 to 2000
RPEG9695	Real Per Capita Earnings growth between 1995 to 1996
RPEG9796	Real Per Capita Earnings growth between 1996 to 1997
RPEG9897	Real Per Capita Earnings growth between 1997 to 1998
RPEG9998	Real Per Capita Earnings growth between 1998 to 1999
RPEG0099	Real Per Capita Earnings growth between 1999 to 2000
Spatial Structure Variables	
VarName	*VarDescription*
POP2000	Population in 2000
POPDEN00	Population Density in persons per square mile in 2000
CORN96	Corn production in 1996 (production=yield*harvested_acres)
CORN97	Corn production in 1997
CORN98	Corn production in 1998
CORN99	Corn production in 1999
CORN00	Corn production in 2000
CORNAC00	Corn acres in 2000
SOY96	Soybean production in 1996
SOY97	Soybean production in 1997
SOY98	Soybean production in 1998
SOY99	Soybean production in 1999
SOY00	Soybean production in 2000
Economic Structure Variables	
VarName	*VarDescription*
RPCE2000	Real Per Capita Earnings in 2000
FMRCR00	Farm real cash receipts in 2000
EMPLY00	Employment in 2000
FRMSH00	Farm earnings as a percentage of total earnings in 2000
MFGSH00	Manufacturing earnings as a percentage of total earnings in 2000
RETSH00	Retail trade as a percentage of total earnings in 2000

Table 4.16 Data Description

Variable Name	Abbv. VarName	Definitions
Growth Variables		
Real Per Capita Earnings Growth	RPEG	Real per capita earnings growth at the county level.
Population Growth	POPG	Population growth rate for a county.
Spatial Structure Variables		
Corn Production	CORN	Total production of corn in bushels at a county level.
Corn Acres	CORNAC	Corn acres harvested.
Soybean Production	SOY	Total production of soybeans in bushels at a county level.
Population	POP	Total number of residents in a county.
Population Density	POPDEN	Total number of residents in a county divided by the land area of the county.
Economic Structure Variables		
Real Per Capita Earnings	RPCE	Real per capita earnings uses BEA's per capita net earnings (net earnings by place of residence/population) and deflates by CPI2010 to convert into real per capita earnings.
Farm Real Cash Receipts-Crops	FMRCR	Farm real cash receipts uses BEA’s Cash receipts from crops and deflates by CPI2010 to convert into real cash receipts from crops.
Employment	EMPLY	The BEA employment series for local areas comprises estimates of the number of jobs, full-time plus part-time, by place of work. Full-time and part-time jobs are counted at equal weight. BEA's estimates of local area employment consist of the number of wage and salary jobs, sole proprietorships, and general partners.
Farm Share of Earnings	FRMSH	Farm earnings divided by farm earnings + non-farm earnings. If farm earnings are negative, it is eliminated from the denominator since it decreases the denominator which can produce large negative values for farm share of earnings.
Manufacturing Share of Earnings	MFGSH	Manufacturing earnings divided by farm earnings + non-farm earnings. If farm earnings are negative, it is eliminated from the denominator.
Retail Share of Earnings	RETSH	Retail earnings divided by farm earnings + non-farm earnings. If farm earnings are negative, it is eliminated from the denominator.

Table 4.17 Summary Statistics (Selection Variables & Dependent Variable)

			Summary Statistics (1995 to 2016)						
	Abbv. VarName	Unit	Minimum	1st Quartile	Mean	Median	3rd Quartile	Maximum	Source
Growth Variables									
Real Per Capita Earnings Growth	RPEG	%	-27.2596	-0.0232	0.0212	0.0132	0.0502	11.1207	BEA/FRED
Population Growth	POPG	%	-0.1477	-0.0059	0.0005	0.0000	0.0064	0.1935	BEA
Spatial Structure Variables									
Corn Production	CORN	Bu	0	1,283,025	9,298,395	5,823,500	13,899,000	77,224,000	USDA
Corn Acres	CORNAC	Acres	0	11,800	62,264	46,400	93,900	394,000	USDA
Soybean Production	SOY	Bu	0	287,000	2,432,899	1,768,500	3,924,750	21,586,000	USDA
Population	POP	#	421	8,441	62,281	19,946	43,568	5,373,418	BEA
Population Density	POPDEN	#/sq. mi.	0.47	13.05	122.81	33.04	80.59	5,957.39	BEA/Census
Economic Structure Variables									
Real Per Capita Earnings	RPCE	$	-4,887	17,220	21,216	20,353	24,092	106,726	BEA/FRED
Farm Real Cash Receipts-Crops	FMRCR	$	0	21,454	65,495	49,796	91,501	690,121	BEA/FRED
Employment	EMPLY	#	226	4,256	37,220	9,995	23,646	3,513,899	BEA
Farm Share of Earnings	FRMSH	%	-1.8293	0.0081	0.1002	0.0414	0.1487	0.8764	BEA
Manufacturing Share of Earnings	MFGSH	%	-0.0024	0.0633	0.1660	0.1473	0.2425	0.6854	BEA
Retail Share of Earnings	RETSH	%	0	0.0532	0.0750	0.0711	0.0912	0.7099	BEA

Note: Detailed source information is listed in Table 4.21.

Table 4.18 Summary Statistics (DID Model Dependent Variable)

Midwest Population Match Treated-Control County Group			Summary Statistics for Data used in DID Regression Model						
Varible	County	Time	Minimum	1st Quartile	Mean	Median	3rd Quartile	Maximum	Std. Dev.
lrpce1	Treated	Before	9.6437	9.8634	9.9590	9.9621	10.0710	10.2069	0.1582
lrpce1	Treated	After	9.7571	9.9300	10.0471	10.0311	10.1721	10.4741	0.1651
lrpce1	Control	Before	9.5484	9.8255	9.9127	9.8970	10.0096	10.2838	0.1612
lrpce1	Control	After	9.5122	9.8408	9.9417	9.9520	10.0695	10.2513	0.1712
lrpce2	Treated	Before	9.6437	9.8624	9.9545	9.9619	10.0701	10.2069	0.1560
lrpce2	Treated	After	9.5405	9.9505	10.0917	10.0991	10.2233	10.5837	0.1983
lrpce2	Control	Before	9.5484	9.8250	9.9086	9.8955	10.0026	10.2838	0.1598
lrpce2	Control	After	9.4797	9.8175	9.9457	9.9578	10.0537	10.3557	0.1792
lrpce3	Treated	Before	9.6437	9.8624	9.9545	9.9619	10.0701	10.2069	0.1560
lrpce3	Treated	After	9.7088	10.0298	10.1899	10.1886	10.3255	10.7087	0.2373
lrpce3	Control	Before	9.5484	9.8250	9.9086	9.8955	10.0026	10.2838	0.1598
lrpce3	Control	After	9.5376	9.8685	9.9969	9.9854	10.1110	10.5613	0.2112
lrpce4	Treated	Before	9.6437	9.8624	9.9545	9.9619	10.0701	10.2069	0.1560
lrpce4	Treated	After	9.7957	10.0706	10.2318	10.1747	10.3596	10.8920	0.2585
lrpce4	Control	Before	9.5484	9.8250	9.9086	9.8955	10.0026	10.2838	0.1598
lrpce4	Control	After	9.5752	9.9084	10.0408	10.0302	10.1982	10.5308	0.2190
lrpce5	Treated	Before	9.6437	9.8624	9.9545	9.9619	10.0701	10.2069	0.1560
lrpce5	Treated	After	9.7895	10.0371	10.2322	10.2353	10.3787	11.0382	0.2480
lrpce5	Control	Before	9.5484	9.8250	9.9086	9.8955	10.0026	10.2838	0.1598
lrpce5	Control	After	9.5608	9.9131	10.0536	10.0720	10.2031	10.4783	0.2134

Note 1: lrpcex = Natural log of real per capita earnings and where x represents number of years after treatment

Note 2: For lrpce1, there are N=56 observations. For lrpcex where x= 2 to 5, there are N=55 observations.

Note 3: Data is pooled and aligned around the treatment year. Matched county data with less than 25,000 in population and time period from 2001 to 2016.

Note 4: Source: Burearu of Economic Analysis (county level per capita earnings) and FRED (CPI-2010)

Table 4.19 Rural-Urban Continuum Codes (USDA-ERS)

Code	Description
Metropolitan Counties	
1	Counties in metro areas of 1 million population or more
2	Counties in metro areas of 250,000 to 1 million population
3	Counties in metro areas of fewer than 250,000 population
Nonmetropolitan Counties	
4	Urban population of 20,000 or more, adjacent to a metro area
5	Urban population of 20,000 or more, not adjacent to a metro area
6	Urban population of 2,500 to 19,999, adjacent to a metro area
7	Urban population of 2,500 to 19,999, not adjacent to a metro area
8	Completely rural or less than 2,500 urban population, adjacent to a metro area
9	Completely rural or less than 2,500 urban population, not adjacent to a metro area

Table 4.20 List of Data Movies

File Name	**Contents**
MidwestCornChoropleth.mp4	Corn choropleth (1995-2017)
MidwestCornYield.mp4	Corn yield choropleth (1995-2017)
MidwestSoybeanChoropleth.mp4	Soybean choropleth (1995-2017)
MidwestEthanolBiorefineryLocations.mp4	Biorefinery Locations (2000-2016)
MidwestTreatedControlMatch_RUCC.mp4	Midwest Matched Counties (2001-2015)
StateTreatedControlMatch_BestMatch.mp4	State Matched Counties (2001-2015)

Table 4.21 Data Sources and Internet Links

United States Department of Agriculture (USDA)

National Agricultural Statistics Service (NASS)

Economic Research Service (ERS)

- General Agricultural Information:
  https://www.usda.gov
- Midwest County Corn & Soybean Acreage, Yield and Production:
  https://www.nass.usda.gov/Quick_Stats/Lite/index.php
- 2013 Rural-Urban Continuum Codes:
  https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

Renewable Fuels Association (RFA):

- General Ethanol Industry Information:
  https://ethanolrfa.org/
- Biorefinery Locations:
  https://ethanolrfa.org/resources/biorefinery-locations/
- Ethanol Industry Outlook Publication:
  https://ethanolrfa.org/resources/publications/outlook/
- Archive Information on RFA:
  https://web.archive.org/web/*/ethanolrfa.org

United States Census Bureau

- U.S. Maps
  https://www.census.gov/quickfacts/table/PST045215/00
- U.S. County Area (File: GCT-PH1-Geography-United States)
  https://www.census.gov

United States Bureau of Economic Analysis (BEA)

- General Information:
  http://www.bea.gov
- County Level Data (per capita earnings, employment, population, manufacturing earnings, farm earnings, retail earnings, etc.)
  https://www.bea.gov/itable/index_regional.cfm

FRED/OECD:

- Consumer Price Index (CPI or CPI2010):
  https://fred.stlouisfed.org/series/CPALTT01USA661S

Google Earth:

- Mapping Biorefinery Locations
- Extract Latitude & Longitude Data (based on address information)

4.12 Figures in Chapter 4

Figure 4.1 Flint Hills Resources, Iowa Falls, Iowa (105 mgy)

CHAPTER 5. EMPIRICAL RESULTS AND DISCUSSION

The research hypothesis is tested using six difference-in-differences (DID) models with various combinations of control variables. An analysis of the DID models' residuals is presented to demonstrate that the residuals have mostly normal distribution characteristics. A robustness check is performed against all control groups to more fully understand the variation in results and how it might affect acceptance or rejection of the null hypothesis. Employment and employment rate impacts are examined using means comparisons. An employment multiplier is derived based on average ethanol biorefinery employment and the expansion of new employment in treated counties over the five year period after treatment. Exploratory analysis is presented using the full treated county data set to examine whether significant results can be obtained by including large population treated counties. Additional exploratory analysis is presented on initial plant capacity versus treated county population.

5.1 Research Hypothesis Results and Discussion

5.1.1 Parallel Trend Assumption and Treatment Response

First, the parallel trend must be established in the pre-treatment period to satisfy the identification criteria of Section 3.2.1.1. Figure 5.1 through Figure 5.6 show the parallel trend for real per capita earnings in the pre-treatment period prior to treatment for all treated counties with populations less than 25,000. In each graph, the treated group is the same data, but the control groups are based on the matching by region (Midwest [MW] or State) and the secondary matching of Best Match, Population Match, and RUCC. Data points for the treated and control group counties are plotted adjacent to the vertical time lines and slightly offset so that the two groups of points do not overlap. In the period prior to treatment (t=0), the parallel trend does have a slight bump in the treated county data at period t-3, but otherwise seems to recover prior to treatment. As an additional note, the graphs are using real per capita earnings as the plotted variable, whereas the dependent variable in the regression model is the natural log of real per capita earnings. Since all the data are on the same order of magnitude, a log transformation of the data will still produce similar scaled treated and control group characteristic curves, but the larger values will be slightly more compressed than the smaller values which tends to produce earnings data patterns that are more symmetric about the means. Overall, there does not seem to be any reason to reject the parallel trend in the data prior to treatment.

After treatment (t=0), the graphs in Figure 5.1 through Figure 5.6 show a noticeable response in the treated group. Though the treatment response for the treated group is the same in all of the graphs, the graphs in Figure 5.4 through Figure 5.6 show that the State control groups are more influenced by the treatment than the Midwest control groups shown in Figure 5.1 through Figure 5.3. It was anticipated that State control groups might be more susceptible to spillover effects; therefore, matching over the entire Midwest region could produce more robust control groups (i.e. control groups less affected by the treatment). Furthermore, the Midwest control groups seem to be more stable in the post-treatment period. Thus, in the opinion of this researcher, the Midwest control groups are the best selection for testing the null hypothesis of this research.

5.1.2 Empirical Results

Two approaches will be used to test the hypothesis of this research. First, the hypothesis will be tested over different regression models based on controlling for state effects, time effects and spillover effects. Also, these models will be tested over time from one year after treatment to five years after treatment to capture the treatment response relative to the control group. In this first approach, only one control group will be used to evaluate the hypothesis which is based on the Midwest Population Match control group as shown in Figure 5.2. Second, an appropriate model will be selected from the first approach to evaluate the hypothesis against all control groups, shown in Figure 5.1 through Figure 5.6, over the five year period of time after treatment. The second approach provides a robustness check for the average treatment effect on the treated (ATOT) parameter over the period of time after treatment.

There were 119 new ethanol biorefineries that started operations in the U.S. Midwest from 2001 to 2015, but only 97 qualified as treatment events for the counties where they reside. Due to the population restrictions imposed by the research hypothesis where rural is defined as counties with populations less than 25,000, the set of eligible treated counties reduces to 56 counties. Based on the treatment response discussion in Section 5.1.1, the Midwest control groups had the most stable control groups which seem minimally affected by spillover effects as opposed to the State control groups. Since all three Midwest control groups seem to have fairly similar curves after treatment on the treated counties, the Midwest population match control group will be used to evaluate the null hypothesis.

The six regression models used in the hypothesis test are based on Equation 3.1 with various sets of control variables (state fixed effects, time fixed effects and spillover effects) and using the Midwest population match control group. Results for the average treatment effect on the treated (ATOT) one year after treatment are shown in Table 5.1. The dependent variable is the natural log of real per capita earnings which is abbreviated as lrpce1 where the numerical index represents the number of years after treatment. In the results, ATOT represents the growth rate in real per capita earnings from the base year (year before treatment) to the first year after treatment relative to a control group. In five of the six models, ATOT had a one-sided significance of at least 10% except for Model 5 which used all of the control variables and was not significant. As expected, the models with the spillover control variables had higher growth rates as explained in Section 3.2.8 except for Model 5 where the time fixed effects affected the ATOT parameter. The growth rates vary over the models from 5.53% to 7.63% (uncorrected) or 5.69% to 7.93% (corrected[6]) over the two year period[7]. An interpretation of this growth rate is that Midwest counties with newly operational ethanol biorefineries from 2001 to 2015 and with populations less than 25,000 grew, on average, at an annual growth rate between 2.84% to 3.96% (corrected) relative to matched control counties with similar population levels for the two year period which spans from the year-end of the base year, immediately before the treatment year, to one year after the treatment year.

Regression results for two years after treatment are shown in Table 5.2. ATOT results are significant for all models at a 5% level except for Model 5 which is only

[6] Growth rates are corrected using 100*(exp(uncorrected rate) – 1) as described in Wooldridge (2016).
[7] The two year period covers the treatment year and the first year after treatment. In other words, the treatment effect, in this case, is measured relative to the year-end observation in the base year (year before treatment) to the year-end observation of the first year after the treatment year which is two years.

significant at a 10% level. The average treatment effect on the treated varies from 10% to 12% (uncorrected) growth rate over the three year period (from year-end base year to two years after the treatment year). Three years after treatment, ATOT is significant at a 1% level for all models shown in Table 5.3 and varies in value from 14.7% to 19.6% (uncorrected) growth rate over the four year period (from year-end base year to three years after the treatment year). Similarly, four and five years after treatment results are shown in Table 5.4 and Table 5.5, respectively, in which all models show ATOT is estimated at a 1% significance level. Growth rates in the fourth year after treatment varied from 14.5% to 18.3%. In the fifth year after treatment, ATOT varied between 13.3% and 18.9% across the models. These results can be viewed over time after treatment by graphing the maximum and minimum ATOT growth rates (uncorrected) as shown in Figure 5.7. Up until three years after treatment, there is upward movement in the growth rate, but declines after the third year. This data was transformed by simple averaging over the appropriate period to produce the maximum and minimum average annual growth rates (uncorrected) as shown in Figure 5.8. Using the annual growth rates over the respective period, it is much clearer that there is growth in the first three-years after treatment and it appreciably declines in the fourth and fifth years after treatment.

Though the preceding DID model results are based on one control group, the significance of the ATOT parameter across several models and over time suggest that the null hypothesis should be rejected in favor of the alternative hypothesis. Thus, treated counties with newly operational ethanol biorefineries and populations less than 25,000 did experience economic benefits relative to similar counties without ethanol biorefineries. This hypothesis will continue to be tested across all control groups in Section 5.1.4.

5.1.3 Residual Analysis on the Empirical Models

Normal qq-plots are used to assess the normality of the residuals for the models discussed in Section 5.1.2. The residuals of the models presented in Table 5.1 through Table 5.5 are shown in Figure 5.9 through Figure 5.13, respectively. The normal qq-plots of the residuals for Model 1 through Model 6 in each table are shown in Panel A through Panel F, respectively, as displayed in Figure 5.9 through Figure 5.13. Overall, most of the residuals follow a normal distribution (as demonstrated by the clustering of residuals along the red qq-line) though some residuals in the tails do show significant deviation from the normal distribution. Despite the deviation in the tails, the residuals are sufficiently normally distributed such that the standard errors of the DID econometric models are essentially normally distributed. Thus, the error structure of the econometric DID models allows the hypothesis test to be evaluated based on the approach discussed in CHAPTER 3.

5.1.4 Robustness Check (ATOT Consistency over Control Groups)

Since using a single control group could produce misleading results, the hypothesis test should be analyzed over all control groups to test the robustness of the acceptance of the alternative hypothesis. Based on the multi-model analysis in the Section 5.1.2, the most naïve[8] DID model (Model 1), without control variables, was selected for evaluation across control groups. The naïve DID model was selected for

[8] A naïve DID model refers to a basic difference-in-difference model without additional control variables.

several reasons. First, its ATOT coefficients were consistently the smallest coefficients of all the models and this robustness check is essentially a test of the lower bounds of growth associated with testing the hypothesis. Second, the Bayesian Information Criteria (BIC) consistently ranked the naïve DID model as the best model out of the six models examined and the F-statistics were significant for all years after treatment. Finally, most of the other models produced lower p-values for the ATOT coefficient and this robustness test is essentially testing the significance of the lower bound and how that changes over time. Thus, the naïve DID model was the best selection for this analysis.

In this robustness check, the natural log of real per capita earnings is the dependent variable and is abbreviated as lrpcex where the x index represents the year after treatment. Again, the population for the treated counties is limited to less than 25,000 in this analysis. The regression results are shown in Table 5.6 through Table 5.10 where each table presents the results across all control groups for a particular year after treatment. In Table 5.6 (one year after treatment), two Midwest control groups had 10% significance levels on the ATOT coefficients, but the other four models were not significant. The lower coefficients for the State control groups (2.36% to 3.88%) was expected based on the control group graphs in Figure 5.4 through Figure 5.6 relative to the Midwest control group graphs in Figure 5.1 through Figure 5.3. Thus, these results bring into question whether the null hypothesis can be rejected for the period one year after treatment. Table 5.7 shows the results for two years after treatment. Midwest control groups are significant at the 5% and 1% level with ATOT values in the range from 9.08% to 10.8%. State control groups were significant at 10% and 5% with values in the range of 6.3% to 8.64%. Results three years after treatment are shown in Table 5.8

with two results having 5% significance and four results at 1% significance. In Table 5.9, the results are still strong for the Midwest control groups at the 1% significance level, but the State control groups significance levels deteriorate to 10% and 5% for four years after treatment. Table 5.10 shows the results for five years after treatment where there are two results with 1% significance, two results with 5% significance, one result with 10% significance and one result that is not significant.

There are several key aspects of the robustness test that are worth noting. Over all years after treatment, the ATOT coefficients for the State control groups have smaller values in any particular year after treatment relative to the ATOT coefficients for the Midwest control groups. There are potentially two main causes for this difference. First, the Midwest control groups were selected from a larger pool of potential control counties and thus, better matches were obtained for the Midwest control groups. Second, the State control groups (in-state matches) potentially suffer from greater spillover effects due to a more limited pool of potential control counties. It is the spillover effects that seem to be reflected in the upward movement of the State control group graphs in Figure 5.4 through Figure 5.6.

The results of the robustness test can be summarized as follows. One year after treatment, the regression results for ATOT only show weak significance for two control groups; therefore, the null hypothesis cannot be rejected for this case. There is moderate support through marginal and strong significance levels for ATOT in rejecting the null hypothesis for two, four and five years after treatment. Finally, the strongest support for rejecting the null hypothesis occurs three years after treatment when ATOT has significance levels of 1% and 5% across all control groups. Thus, in four out of five periods after treatment, the null hypothesis can be rejected in favor of the alternative hypothesis which states that ethanol biorefineries did have positive economic impacts on the counties where they were located relative to control counties without ethanol biorefineries.

One important lesson for testing several periods after treatment is that significant results for the variable of interest did not occur until two years after treatment and became stronger in the third year after treatment. So, studies that only test one period after treatment may actually miss a statistically significant response that is not observed until several years after the treatment. Another point is that the control groups may eventually start responding to the treatment after some period of time; especially, if the treatment has spillover effects that can have economic ripple effects through contiguous counties.

5.1.5 Treated County Year-over-Year Growth

The results from Section 5.1.2 can be presented in year over year changes in growth as shown in Table 5.11. The top table shows the values from Section 5.1.2 which are the growth rates over the period evaluated in the DID model and the bottom table shows the corrected year over year growth rates. The treatment period (t) and first year after treatment (t+1) growth rates are estimates based on the DID model results for one year after treatment. This is calculated by correcting the 7.60% (max) results and then dividing by two[9]. Other results are calculated by taking the difference between the DID model results for the consecutive periods and then correcting that value to obtain the year

[9] This includes the treatment year and the first year after treatment which is two years relative to the base year (year just prior to the treatment year).

over year growth rate. Also, the (max, min) values in the year over year growth table match up with (max, min) values in the uncorrected growth table even though the values may not represent a max or a min in the year over year growth table.

In Table 5.11, it is clear that growth occurred for the first three years after treatment for the treated counties, but growth diminished significantly and became negative four years after treatment. This is exactly what is shown in Figure 5.1 through Figure 5.3 where the treated group response seemed to hit a peak four years after treatment while the control groups continued a slow rise in their response over time.

5.1.6 Employment Impacts

For many years, counties throughout the U.S. Midwest have seen population declines with associated employment level declines, but somewhat stable employment rates. Thus, it is hypothesized that a new ethanol biorefinery in a county would stabilize or slightly increase employment, at least in the short-run. For this analysis, only treated counties with less than 25,000 in population are examined for any potential employment effects. The Midwest and State population match control groups will be used for comparison purposes, since comparing on similar population levels is the only approach that makes sense in this analysis.

In Figure 5.14, the treated counties with populations less than 25,000 are compared to the Midwest population match control group. There is a good parallel trend in the pre-treatment period. In the post-treatment period, the treated counties seem to be mostly stable with some employment gains overall; whereas, the control group had employment loses after the treatment event (t=0). The mean values used in the graph are shown in Table 5.12 under the Employment table. This clearly shows that the treated counties gained employment on average from 7476 in the base year (t-1) to 7687 five years after treatment. Control county employment declined on average from 7138 in the base year to 6969 five years after treatment. The difference between the treated mean and the parallel trend assumption is an employment increase of 380 more positions for the treated counties, on average, in the five year period after treatment. It should be noted that the parallel trend assumption represents the treated group's response in the absence of treatment. In the Employment Rate[10] table, the treated counties employment rate increased from 59.9% to 62.1% from the base year to five years after treatment, respectively. Thus, there are employment gains, on average, in the treated counties after treatment and the counterfactual control group response showed employment declines which seems in-line with regional trends.

In Figure 5.15, the treated counties with populations less than 25,000 are compared to the State population match control group. On average, the treated group means track the control group means up to one year after treatment; then, the control group has relatively strong employment gains over the next four periods. Treated and control group means are shown in Table 5.13 in the Employment table. The treated counties had average employment gains of 211 from base year to five years after treatment while the control counties had average employment gains of 960 over the same period. In the Employment Rate table, treated and control counties increased their employment rate from the base year to five years after treatment by about the same amount. Additionally, the Population table shows that treated counties declined in population by about 235 on average while control counties gained in population by 128

[10] County employment rate is calculated as total county employment divided by total county population.

on average. One possibility for the rise in employment in the State control group is that there could have been a gradual increase in part-time jobs which have the same weight in employment numbers with the Bureau of Economic Analysis (BEA: refer to Table 4.16 for definition of Employment). Further, the jobs could have been related to seasonal workers who benefitted from opportunities associated with spillover effects from treated counties, but this is just speculation. As discussed previously, State control groups were suspected of benefitting from spillover effects, since the State control groups' real per capita earnings had a positive response several years after the treatment was applied on the treated counties.

This investigation into employment effects had somewhat mixed results. As expected, treated counties with populations less than 25,000 showed, on average, an increase in employment and an increase in the employment rate after treatment. In the Midwest control group case, employment in the control counties declined on average which was hypothesized might occur. For the State control group case, employment levels began to rise two years after treatment and continued until five years after treatment. What happened in the State control group case is not known with certainty though some thoughts were presented with the findings. Based on the increased employment finding associated with the State control group, this gives more credence to using the Midwest control groups as the preferred counterfactual for validation of the main hypothesis of this dissertation.

5.1.7 Employment Multiplier

An employment multiplier is calculated based on the treated county employment gains (direct, indirect and induced) shown in Table 5.12 relative to the average employment of an ethanol biorefinery. Since ethanol biorefineries generally employ between 40 to 65 people based on plant size, an average estimate of 50 employees is used to represent the typical ethanol biorefinery. Table 5.14 shows the employment multiplier calculated for the treatment year (time = 0) through five years after treatment. The employment multiplier varies from 1.46 in the treatment year to 7.6 five years after treatment. These results are similar to the results in Low and Isserman (2009) where their modeling showed employment multipliers ranging from 2.8 to 6.4[11] depending on the location and size of the ethanol plant. The employment multiplier results show that there are indirect and induced employment gains for any county with a new ethanol biorefinery beyond just the direct employment gains from the jobs at the plant.

5.2 Exploratory Analysis and Discussion

5.2.1 Revisit Research Hypothesis with All Eligible Treated Counties

In an effort to explore more insights into the data, a naïve DID regression was run for the full set of treated counties (97) against all control groups with no treated county population limits. Parallel trend and treatment response graphs are shown in Figure 5.16 through Figure 5.21 for the six control groups. The parallel trend prior to treatment seems reasonable in all the graphs. Treatment response in these graphs is not as strong as for the rural analysis with treated county populations less than 25,000. Based on the less prominent treatment response, the DID regression model will be run for the period three

[11] These values are calculated based on modeled job estimates from Table 7 of Low and Isserman (2009).

years after treatment which was the optimal period for a highly significant ATOT coefficient in the rural treated county case.

Regression results for all the treated counties across all control groups and three years after treatment are shown in Table 5.15. There are five ATOT coefficients with significance levels of 5% and one coefficient with a significance level of 10%. ATOT coefficients range from 5.99% to 8.98% which is quite lower than the values of 10.2% to 14.7% found in the rural treated county analysis. These lower values are expected since for any given economic impact associated with an ethanol biorefinery that economic impact will diminish on a per capita basis in large population counties which means its economic impact is more difficult to detect in large population counties. Also, the ethanol biorefinery economic impact in large population communities can be masked by other more dominant economic activities.

5.2.2 Treated County Population versus Biorefinery Initial Capacity

Another exploratory question is whether an ethanol biorefinery's initial capacity had a significant economic impact that varied based on the population level of the treated county. It can be hypothesized that high capacity ethanol biorefineries might have a significant economic impact in high population counties, but should definitely have a significant economic impact in low population counties. Also, low capacity ethanol biorefineries are unlikely to have a significant economic impact in large population counties, but could have a significant economic impact in low population counties.

For this analysis, the population break point was set at 25,000 which separated the low population treated counties from the high population treated counties. For the ethanol biorefinery initial capacity, a break point was set at 55 mgy to separate the low capacity biorefineries from the high capacity biorefineries. The break points were set to balance, as best as possible, the number of treated counties divided into the four sectors to be tested. The treated counties in each sector are: 22 (low population, high capacity), 34 (low population, low capacity), 20 (high population, low capacity), and 21 (high population, high capacity). In the analysis, a DID regression model with state fixed effects control variables was selected to test this exploratory question. The dependent variable is the natural log of real per capita earnings (lrpcex) where x is the index that represents the year after treatment and the counterfactual control group is the Midwest population match counties. Also, the average treatment effect on the treated (ATOT) is represented by the delta symbol (δ) and is presented with the standard error (se) when a particular sector has a significant result.

Results of the capacity versus population analysis are shown in the two-by-two diagrams of Figure 5.22 for each year after treatment. For high population treated counties, there are no significant results for any capacity level at any period of time after treatment. Also, there are no significant results for the first year after treatment, but one-sided test p-values are provided which give some indication of the regression model results. Low capacity ethanol biorefineries had significant results from two years after treatment through five years after treatment with significance levels better than 5%. High capacity ethanol biorefineries only have significant results in the fourth and fifth years after treatment and their growth rates are lower than for the low capacity plants.

It is not clear why high capacity ethanol biorefineries economic impact was much smaller than low capacity ethanol biorefineries in low population counties. One speculative thought is that low capacity plants are sufficient in size to demand a majority of the county's corn production output and that high capacity plants contribute more to spillover effects. Another speculative thought is that low capacity plants are more likely to be locally owned which can benefit the local community whereas high capacity plants have corporate structures where the plant's profits are exported out of the community and have less local impact. An investigation into the ownership of the low capacity ethanol biorefineries in low population areas found that 13 of the 34 were locally owned. Thus, local ownership could have had an impact on the results. A deeper dive into these questions could form a basis for future research.

5.3 Threats to Internal Validity

5.3.1 SUTVA I – No Interference

The Stable Unit Treatment Value Assumption (SUTVA) is defined by Imbens and Rubin (2015) as, "the treatment applied to one unit does not affect the outcome for the other units." Essentially, this is a concern about spillover effects which can bias the estimate of the treatment effect. It is clear that spillover effects do occur in this analysis since feedstock production is not exclusively contained within a treated county and ethanol biorefineries are not necessarily centered within a treated county's border which can lead to asymmetrical spillovers into contiguous counties. Though spillover indicator variables are applied in some of the DID models, these are relatively crude instruments for controlling for spillover effects and may not fully correct for the bias in the treatment effect estimate.

Let's examine what happens when there are spillover effects occurring associated with an ethanol biorefinery's operation in a treated county. The primary source of spillovers is through corn production in neighboring counties which is supplied as feedstock to the treated county's ethanol biorefinery. Thus, there are three dynamic aspects to consider. First, not all corn production is captured by the treated county. In one sense, this leads to an underestimation of the treatment effect since not all production is captured by the treated county. In another sense, the treated county is likely producing the majority of corn production and this is the treatment effect being captured but can vary across treated counties. Second, what if one of the contiguous counties becomes a control county match? In this case, the control county's economic performance is higher than it would have been without the treatment in a contiguous treated county. Thus, this will tend to put a downward bias on the treatment effect since the counterfactual control county has a higher economic level of performance than it would have had without the spillover effect. Third, what if one of the contiguous counties becomes another treated county? In the short run, this will lead to greater demand for limited corn supply; thus, elevating local prices. In the long run, producers will respond by more intensively increasing their corn production. This would also spatially shift the spillover patterns around the ethanol biorefinery. In this case, the impact on the ATOT is likely to be positive by capturing more local economic activity and may produce a less biased underestimation of ATOT for the treated county.

On average, positive spillovers from a treated county will lead to an underestimation of the ATOT (lower economic performance in the treated county than would have occurred without spillover). Additionally, positive spillovers into a control county will contribute to an underestimation of the ATOT (higher economic performance in the control county than would have occurred without spillover). Thus, any analysis without control variables to capture spillover effects will most likely underestimate ATOT on the treated unit (county).

5.3.2 SUTVA II – No Hidden Variations of Treatments

There are variations in treatments that are not fully considered in this analysis. For instance, the capacity of the ethanol biorefinery directly relates to feedstock demand from the local community and surrounding communities. Large capacity biorefineries can create large spillover effects outside of their county of residence due to the large quantity of feedstock required for production. Small capacity refiners could pull a major percentage of their feedstock demand from the local community or county; thus, creating less of a spillover effect. Though these variations are not necessarily hidden, the information would be extremely difficult to obtain which essentially makes it virtually hidden. Another potentially hidden variation is the efficiency of each individual ethanol biorefinery. It is expected that each plant's cost structure and efficiency will be fairly similar across the industry due to the competitive nature of the ethanol industry.

5.3.3 Selection on Observables into Treatment

Selection into treatment could be influenced by observable characteristics of the treated counties. The main result from this situation is that average treatment on the treated (ATOT) and average treatment on the control (ATOC) are most likely not equal. The ATOC is a potential outcome on the control group if the control group was given a treatment. ATOC cannot be assessed in this analysis, since it is a concept more understood in randomized control trials (RCT) where subjects can be randomly selected to be either in the treated or control group (equivalent control groups). In a RCT, the parameter of interest would be the average treatment effect (ATE), since ATOT and ATOC are expected to be equal. Thus, ATOT is the parameter of interest in this dissertation and is most likely greater than ATE. This threat should not bias the ATOT parameter, but limits or completely eliminates what can be said about ATE unless observable characteristics are modeled in the econometric analysis.

5.3.4 Population Statistics versus Sample Statistics (Assumption)

In population statistics, means are known with certainty. Therefore, a difference between two means is also known with certainty. In sample statistics, samples are randomly drawn from a population distribution and are used to calculate an estimate of the population mean with an associated standard error for the estimated mean. The sample distribution for the mean is essentially normal based on the Central Limit Theorem. So, for this dissertation, there are three major concerns about using population data as if it were sample data. First, it is obvious that the data is representative of the population since it is mostly population data. Therefore, a sample statistics approach to the data is valid since the data is a representative sample. Second, the data used as a dependent variable should have the form of a normal distribution. This is accomplished by taking the natural log of the real per capita earnings variable, since earnings generally have a log-normal distribution. Third, the residuals of the regression models should have a normal distribution. The QQ-plots in Figure 5.9 through Figure 5.13 show that the residuals have mostly normal distribution characteristics with some deviations in the tails. Thus, using a sample statistics approach is valid for this analysis.

5.4 Tables in Chapter 5

Table 5.1 Ln(RPCE) One Year after Treatment, Treated Population < 25,000

	(1)	(2)	(3)	(4)	(5)	(6)
	lrpce1	lrpce1	lrpce1	lrpce1	lrpce1	lrpce1
Treated	0.0464	0.0289	0.0209	0.0145	0.0157	0.0464
	(0.0302)	(0.0384)	(0.0293)	(0.0367)	(0.0369)	(0.0285)
After	0.029	-0.00591	0.029	0.00293	-0.0257	-0.00347
	(0.0314)	(0.0378)	(0.0275)	(0.0334)	(0.0349)	(0.0331)
ATOT (δ)	0.059°	0.0763°	0.059°	0.0689°	0.0553	0.059°
	(0.0438)	(0.0580)	(0.0403)	(0.0523)	(0.0529)	(0.0419)
Spillover Effects	No	Yes	No	Yes	Yes	No
State Fixed Effect	No	No	Yes	Yes	Yes	No
Time Fixed Effects	No	No	No	No	Yes	Yes
Constant	9.913***	9.913***	9.949***	9.945***	9.845***	9.837***
	(0.0215)	(0.0217)	(0.0250)	(0.0258)	(0.0637)	(0.0149)
R-squared	0.0870	0.1240	0.2620	0.2830	0.3770	0.2070
adj. R-squared	0.074	0.095	0.216	0.224	0.288	0.154
BIC	-156.6	-144.2	-150.2	-135	-107.1	-128.8
F	6.953	4.307	27.78	43.71	55.89	82.49
N	224	224	224	224	224	224

Standard errors in parentheses

° p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.

Note 2: Control group = Midwest Population Match

Note 3: lrpce1 = natural log of real per capita earnings; 1 year after treatment

Table 5.2 Ln(RPCE) Two Years after Treatment, Treated Population < 25,000

	(1)	(2)	(3)	(4)	(5)	(6)
	lrpce2	lrpce2	lrpce2	lrpce2	lrpce2	lrpce2
Treated	0.0459	0.0231	0.0213	0.0135	0.0161	0.0459
	(0.0301)	(0.0381)	(0.0293)	(0.0375)	(0.0380)	(0.0290)
After	0.0371	-0.00161	0.0371	0.00786	-0.0198	0.000931
	(0.0324)	(0.0388)	(0.0277)	(0.0339)	(0.0408)	(0.0400)
ATOT (δ)	0.100*	0.120*	0.100*	0.111*	0.112°	0.100*
	(0.0470)	(0.0687)	(0.0429)	(0.0643)	(0.0679)	(0.0461)
Spillover Effects	No	Yes	No	Yes	Yes	No
State Fixed Effect	No	No	Yes	Yes	Yes	No
Time Fixed Effects	No	No	No	No	Yes	Yes
Constant	9.909***	9.909***	9.974***	9.970***	9.885***	9.837***
	(0.0215)	(0.0218)	(0.0254)	(0.0262)	(0.0561)	(0.0151)
R-squared	0.1390	0.1760	0.3170	0.3340	0.3830	0.2080
adj. R-squared	0.127	0.148	0.274	0.278	0.296	0.158
BIC	-127.2	-115.1	-123.9	-108.2	-70.85	-91.42
F	10.06	9.17	62.25	136.6	33.41	18.49
N	220	220	220	220	220	220

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.

Note 2: Control group = Midwest Population Match

Note 3: lrpce2 = natural log of real per capita earnings; 2 years after treatment

Table 5.3 Ln(RPCE) Three Years after Treatment, Treated Population < 25,000

	(1)	(2)	(3)	(4)	(5)	(6)
	lrpce3	lrpce3	lrpce3	lrpce3	lrpce3	lrpce3
Treated	0.0459	0.0231	0.0189	0.00954	0.0112	0.0459
	(0.0370)	(0.0381)	(0.0300)	(0.0378)	(0.0390)	(0.0297)
After	0.0883*	0.0399	0.0883**	0.0497	0.0182	0.0651
	(0.0370)	(0.0442)	(0.0308)	(0.0383)	(0.0535)	(0.0575)
ATOT (δ)	0.147**	0.196**	0.147**	0.188**	0.192**	0.147**
	(0.0524)	(0.0777)	(0.0488)	(0.0741)	(0.0710)	(0.0512)
Spillover Effects	No	Yes	No	Yes	Yes	No
State Fixed Effect	No	No	Yes	Yes	Yes	No
Time Fixed Effects	No	No	No	No	Yes	Yes
Constant	9.909***	9.909***	9.956***	9.953***	9.819***	9.837***
	(0.0262)	(0.0218)	(0.0263)	(0.0269)	(0.0860)	(0.0155)
R-squared	0.2360	0.2660	0.3680	0.3850	0.4790	0.3070
adj. R-squared	0.226	0.242	0.328	0.333	0.403	0.259
BIC	-79.28	-66.33	-66.81	-51.41	-28.68	-41.17
F	22.29	12.97	21.73	35.89	14.88	20.48
N	220	220	220	220	220	220

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.

Note 2: Control group = Midwest Population Match

Note 3: lrpce3 = natural log of real per capita earnings; 3 years after treatment

Table 5.4 Ln(RPCE) Four Years after Treatment, Treated Population < 25,000

	(1)	(2)	(3)	(4)	(5)	(6)
	lrpce4	lrpce4	lrpce4	lrpce4	lrpce4	lrpce4
Treated	0.0459	0.0231	0.0154	0.00162	0.00587	0.0459
	(0.0387)	(0.0381)	(0.0309)	(0.0384)	(0.0385)	(0.0295)
After	0.132***	0.0683	0.132***	0.0805*	0.0935	0.163*
	(0.0387)	(0.0428)	(0.0314)	(0.0370)	(0.0758)	(0.0826)
ATOT (δ)	0.145**	0.183**	0.145**	0.172**	0.151**	0.145**
	(0.0547)	(0.0723)	(0.0508)	(0.0674)	(0.0643)	(0.0520)
Spillover Effects	No	Yes	No	Yes	Yes	No
State Fixed Effect	No	No	Yes	Yes	Yes	No
Time Fixed Effects	No	No	No	No	Yes	Yes
Constant	9.909***	9.909***	9.939***	9.932***	9.829***	9.837***
	(0.0274)	(0.0218)	(0.0263)	(0.0272)	(0.0764)	(0.0154)
R-squared	0.2750	0.3270	0.4040	0.4400	0.5380	0.3810
adj. R-squared	0.265	0.304	0.367	0.393	0.468	0.336
BIC	-59.99	-54.75	-49.32	-41.54	-19.04	-30.31
F	27.26	12.81	18.97	31.52	76.81	22.23
N	220	220	220	220	220	220

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.
Note 2: Control group = Midwest Population Match
Note 3: lrpce4 = natural log of real per capita earnings; 4 years after treatment

Table 5.5 Ln(RPCE) Five Years after Treatment, Treated Population < 25,000

	(1)	(2)	(3)	(4)	(5)	(6)
	lrpce5	lrpce5	lrpce5	lrpce5	lrpce5	lrpce5
Treated	0.0459	0.0231	0.0127	-0.00616	0.00264	0.0459
	(0.0378)	(0.0381)	(0.0307)	(0.0384)	(0.0386)	(0.0287)
After	0.145***	0.082	0.145***	0.0960*	0.183	0.262*
	(0.0378)	(0.0440)	(0.0308)	(0.0378)	(0.1060)	(0.1160)
ATOT (δ)	0.133**	0.189**	0.133**	0.176**	0.151**	0.133**
	(0.0534)	(0.0760)	(0.0493)	(0.0679)	(0.0633)	(0.0497)
Spillover Effects	No	Yes	No	Yes	Yes	No
State Fixed Effect	No	No	Yes	Yes	Yes	No
Time Fixed Effects	No	No	No	No	Yes	Yes
Constant	9.909***	9.909***	9.939***	9.932***	9.829***	9.837***
	(0.0267)	(0.0218)	(0.0260)	(0.0264)	(0.0514)	(0.0150)
R-squared	0.2860	0.3390	0.4210	0.4610	0.5690	0.4200
adj. R-squared	0.276	0.317	0.384	0.416	0.501	0.374
BIC	-70.55	-66.06	-62.74	-57.04	-36.03	-46.24
F	28.81	16	15.54	21.64	12.99	25.43
N	220	220	220	220	220	220

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.
Note 2: Control group = Midwest Population Match
Note 3: lrpce5 = natural log of real per capita earnings; 5 years after treatment

Table 5.6 Ln(RPCE1) Robustness: Treated with Populations < 25,000

Control Group Matching	Midwest			State		
	Best Match	RUCC	Pop Match	Best Match	Pop Match	RUCC
	lrpce1	lrpce1	lrpce1	lrpce1	lrpce1	lrpce1
Treated	0.0377	0.0299	0.0464	0.0283	0.049	0.0606*
	(0.0295)	(0.0290)	(0.0302)	(0.0320)	(0.0306)	(0.0296)
After	0.0436	0.0273	0.029	0.0645	0.0564	0.0493
	(0.0297)	(0.0277)	(0.0314)	(0.0346)	(0.0339)	(0.0316)
ATOT (δ)	0.0445	0.0607°	0.059°	0.0236	0.0317	0.0388
	(0.0426)	(0.0412)	(0.0438)	(0.0462)	(0.0457)	(0.0439)
_cons	9.921***	9.929***	9.913***	9.931***	9.910***	9.898***
	(0.0206)	(0.0199)	(0.0215)	(0.0240)	(0.0221)	(0.0207)
adj. R-sq	0.065	0.065	0.074	0.048	0.065	0.085
BIC	-169	-184	-156.6	-133.3	-138.2	-155.7
F	6.037	5.726	6.953	4.906	6.622	8.221
N	224	224	224	224	224	224

Standard errors in parentheses

° p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.

Note 2: lrpce1 = natural log of real per capita earnings; 1 year after treatment

Table 5.7 Ln(RPCE2) Robustness: Treated with Populations < 25,000

Control Group Matching	Midwest			State		
	Best Match	RUCC	Pop Match	Best Match	Pop Match	RUCC
	lrpce2	lrpce2	lrpce2	lrpce2	lrpce2	lrpce2
Treated	0.0378	0.0292	0.0459	0.0329	0.0474	0.0573
	(0.0293)	(0.0290)	(0.0301)	(0.0309)	(0.0306)	(0.0298)
After	0.0464	0.0287	0.0371	0.0741*	0.0611	0.0507
	(0.0295)	(0.0286)	(0.0324)	(0.0320)	(0.0366)	(0.0316)
ATOT (δ)	0.0908*	0.108**	0.100*	0.063°	0.076°	0.0864*
	(0.0451)	(0.0445)	(0.0470)	(0.0467)	(0.0500)	(0.0464)
_cons	9.917***	9.925***	9.909***	9.922***	9.907***	9.897***
	(0.0204)	(0.0199)	(0.0215)	(0.0226)	(0.0223)	(0.0211)
adj. R-sq	0.125	0.124	0.127	0.109	0.109	0.14
BIC	-145.4	-151.4	-127.2	-129.4	-99.77	-132.1
F	9.376	8.944	10.06	8.679	9.735	11.15
N	220	220	220	220	220	220

Standard errors in parentheses

° p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.

Note 2: lrpce2 = natural log of real per capita earnings; 2 years after treatment

Table 5.8 Ln(RPCE3) Robustness: Treated with Populations < 25,000

Control Group Matching	Midwest			State		
	Best Match	RUCC	Pop Match	Best Match	Pop Match	RUCC
	lrpce3	lrpce3	lrpce3	lrpce3	lrpce3	lrpce3
Treated	0.0378	0.0292	0.0459	0.0329	0.0474	0.0573
	(0.0293)	(0.0290)	(0.0301)	(0.0309)	(0.0306)	(0.0298)
After	0.0938**	0.0925**	0.0883*	0.133***	0.122**	0.106**
	(0.0328)	(0.0331)	(0.0357)	(0.0346)	(0.0414)	(0.0360)
ATOT (δ)	0.142**	0.143**	0.147**	0.102*	0.114*	0.129**
	(0.0504)	(0.0506)	(0.0524)	(0.0516)	(0.0564)	(0.0525)
_cons	9.917***	9.925***	9.909***	9.922***	9.907***	9.897***
	(0.0204)	(0.0199)	(0.0215)	(0.0226)	(0.0223)	(0.0211)
adj. R-sq	0.231	0.223	0.226	0.222	0.2	0.234
BIC	-95.73	-94.41	-79.28	-85.66	-46.81	-77.78
F	18.4	17.82	18.59	18.77	18.95	20.18
N	220	220	220	220	220	220

Standard errors in parentheses
° p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.
Note 2: lrpce3 = natural log of real per capita earnings; 3 years after treatment

Table 5.9 Ln(RPCE4) Robustness: Treated with Populations < 25,000

Control Group Matching	Midwest			State		
	Best Match	RUCC	Pop Match	Best Match	Pop Match	RUCC
	lrpce4	lrpce4	lrpce4	lrpce4	lrpce4	lrpce4
Treated	0.0378	0.0292	0.0459	0.0329	0.0474	0.0573
	(0.0293)	(0.0290)	(0.0301)	(0.0309)	(0.0306)	(0.0298)
After	0.134***	0.125***	0.132***	0.189***	0.187***	0.168***
	(0.0355)	(0.0351)	(0.0366)	(0.0370)	(0.0480)	(0.0396)
ATOT (δ)	0.143**	0.152**	0.145**	0.0887°	0.0901°	0.109*
	(0.0540)	(0.0538)	(0.0547)	(0.0550)	(0.0630)	(0.0568)
_cons	9.917***	9.925***	9.909***	9.922***	9.907***	9.897***
	(0.0204)	(0.0199)	(0.0215)	(0.0226)	(0.0223)	(0.0211)
adj. R-sq	0.264	0.258	0.265	0.265	0.221	0.262
BIC	-65.54	-67.7	-59.99	-57.5	1.844	-43.63
F	22.72	21.91	22.73	24.8	23.59	25.1
N	220	220	220	220	220	220

Standard errors in parentheses

° p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.

Note 2: lrpce4 = natural log of real per capita earnings; 4 years after treatment

Table 5.10 Ln(RPCE5) Robustness: Treated with Populations < 25,000

Control Group Matching	Midwest			State		
	Best Match	RUCC	Pop Match	Best Match	Pop Match	RUCC
	lrpce5	lrpce5	lrpce5	lrpce5	lrpce5	lrpce5
Treated	0.0378	0.0292	0.0459	0.0329	0.0474	0.0573
	(0.0293)	(0.0290)	(0.0301)	(0.0309)	(0.0306)	(0.0298)
After	0.156***	0.148***	0.145***	0.193***	0.202***	0.171***
	(0.0343)	(0.0341)	(0.0359)	(0.0379)	(0.0488)	(0.0434)
ATOT (δ)	0.121*	0.130**	0.133**	0.085°	0.0759	0.107*
	(0.0523)	(0.0522)	(0.0534)	(0.0547)	(0.0628)	(0.0587)
_cons	9.917***	9.925***	9.909***	9.922***	9.907***	9.897***
	(0.0204)	(0.0199)	(0.0215)	(0.0226)	(0.0223)	(0.0211)
adj. R-sq	0.28	0.274	0.276	0.27	0.227	0.25
BIC	-79.53	-80.85	-70.55	-59.91	0.339	-29.07
F	25.49	24.61	24.74	26.05	25.48	26.28
N	220	220	220	220	220	220

Standard errors in parentheses
° p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Note 1: One-sided p-values for ATOT. All other variables are two-sided p-values.
Note 2: lrpce5 = natural log of real per capita earnings; 5 years after treatment

Table 5.11 Corrected Growth Rates Year over Year

Uncorrected Growth Rates over Period (%)

Growth	t	t+1	t+2	t+3	t+4	t+5	Time
Max		7.60	12.00	19.60	18.30	18.90	
Min		5.50	10.00	14.70	14.50	13.30	

Corrected Growth Rates for Year over Year Growth (%)

Growth	t	t+1	t+2	t+3	t+4	t+5	Time
Max	3.95	3.95	4.50	7.90	-1.29	0.60	
Min	2.83	2.83	4.60	4.81	-0.20	-1.19	

Table 5.12 Employment Treated-Control Groups Means (Midwest)

	Employment						
	Year after Treatment						
	t-1	0	1	2	3	4	5
Treated Mean	7476	7538	7529	7509	7534	7586	7687
Control Mean	7138	7127	7078	7069	7006	6964	6969
Parallel Trend Assumption	7476	7465	7416	7408	7344	7302	7307
Treated Mean - Parallel Trend	0	73	113	101	190	284	380

Note: Midwest Population Match Control Group. Means are simple averages.

	Employment Rate						
	Year after Treatment						
	t-1	0	1	2	3	4	5
Treated Mean	0.5990	0.6072	0.6099	0.6088	0.6118	0.6164	0.6214
Control Mean	0.5844	0.5847	0.5838	0.5857	0.5847	0.5853	0.5871
Parallel Trend Assumption	0.5990	0.5993	0.5985	0.6003	0.5994	0.6000	0.6017
Treated Mean - Parallel Trend	0.0000	0.0079	0.0115	0.0085	0.0124	0.0165	0.0196

Note 1: Midwest Population Match Control Group. Means are simple averages.

Note 2: County Employment Rate = County Employment/County Population

Table 5.13 Employment Treated-Control Groups Means (State)

	Employment						
	Year after Treatment						
	t-1	0	1	2	3	4	5
Treated Mean	7476	7538	7529	7509	7534	7586	7687
Control Mean	7434	7511	7519	7569	7739	8088	8394
Parallel Trend Assumption	7476	7553	7561	7611	7781	8130	8436
Treated Mean - Parallel Trend	0	-15	-32	-102	-248	-544	-749

Note: State Population Match Control Group. Means are simple averages.

	Employment Rate						
	Year after Treatment						
	t-1	0	1	2	3	4	5
Treated Mean	0.5990	0.6072	0.6099	0.6088	0.6118	0.6164	0.6214
Control Mean	0.5845	0.5892	0.5897	0.5912	0.5941	0.5990	0.6034
Parallel Trend Assumption	0.5990	0.6038	0.6043	0.6057	0.6087	0.6135	0.6179
Treated Mean - Parallel Trend	0.0000	0.0034	0.0057	0.0030	0.0031	0.0029	0.0034

Note 1: State Population Match Control Group. Means are simple averages.

Note 2: County Employment Rate = County Employment/County Population

	Population						
	Year after Treatment						
	t-1	0	1	2	3	4	5
Treated Mean	12509	12483	12435	12272	12258	12250	12274
Control Mean	12656	12628	12625	12672	12688	12733	12784
Parallel Trend Assumption	12509	12481	12478	12525	12540	12586	12637
Treated Mean - Parallel Trend	0	2	-43	-253	-283	-336	-363

Note: State Population Match Control Group. Means are simple averages.

Table 5.14 Employment Multiplier (Midwest)

	Year after Treatment						
	t-1	0	1	2	3	4	5
Employment Multiplier		1.46	2.25	2.03	3.79	5.68	7.60

Note 1: Midwest Population Match Control Group.

Note 2: Assumes the ethanol biorefineries under study have an average of 50 employees.

Table 5.15 Ln(RPCE3) Robustness: All Treated Counties

Control Group Matching	Midwest			State		
	Best Match	RUCC	Pop Match	Best Match	Pop Match	RUCC
	lrpce3	lrpce3	lrpce3	lrpce3	lrpce3	lrpce3
Treated	0.0382	0.0268	0.0419	0.0358	0.0312	0.0445
	(0.0224)	(0.0231)	(0.0225)	(0.0232)	(0.0230)	(0.0232)
After	0.0573*	0.0426	0.0425	0.0724**	0.0577*	0.0535*
	(0.0246)	(0.0265)	(0.0254)	(0.0262)	(0.0283)	(0.0271)
ATOT (δ)	0.0750*	0.0897*	0.0898*	0.0599°	0.0746*	0.0789*
	(0.0378)	(0.0390)	(0.0383)	(0.0389)	(0.0403)	(0.0395)
_cons	9.934***	9.946***	9.930***	9.937***	9.941***	9.928***
	(0.0159)	(0.0169)	(0.0161)	(0.0170)	(0.0168)	(0.0171)
adj. R-sq	0.099	0.086	0.103	0.092	0.082	0.097
BIC	-185.8	-160.8	-174.9	-164.5	-136.3	-151.8
F	11.9	10.28	12.24	11.61	10.86	12.13
N	384	384	384	384	384	384

Standard errors in parentheses
° p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Note: One-sided test p-values for ATOT. All other variables are two-sided test p-values.

5.5 Figures in Chapter 5

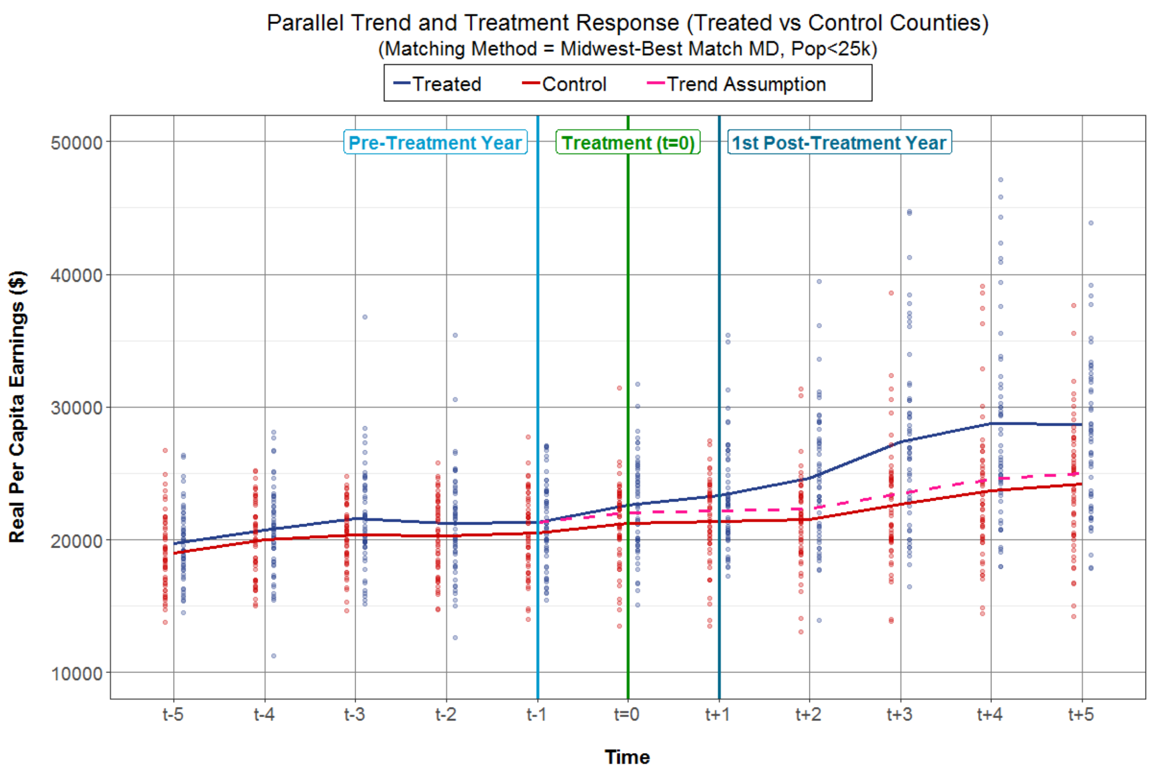

Figure 5.1 Parallel Tread and Treatment Response (MW-Best Match, Pop<25k)

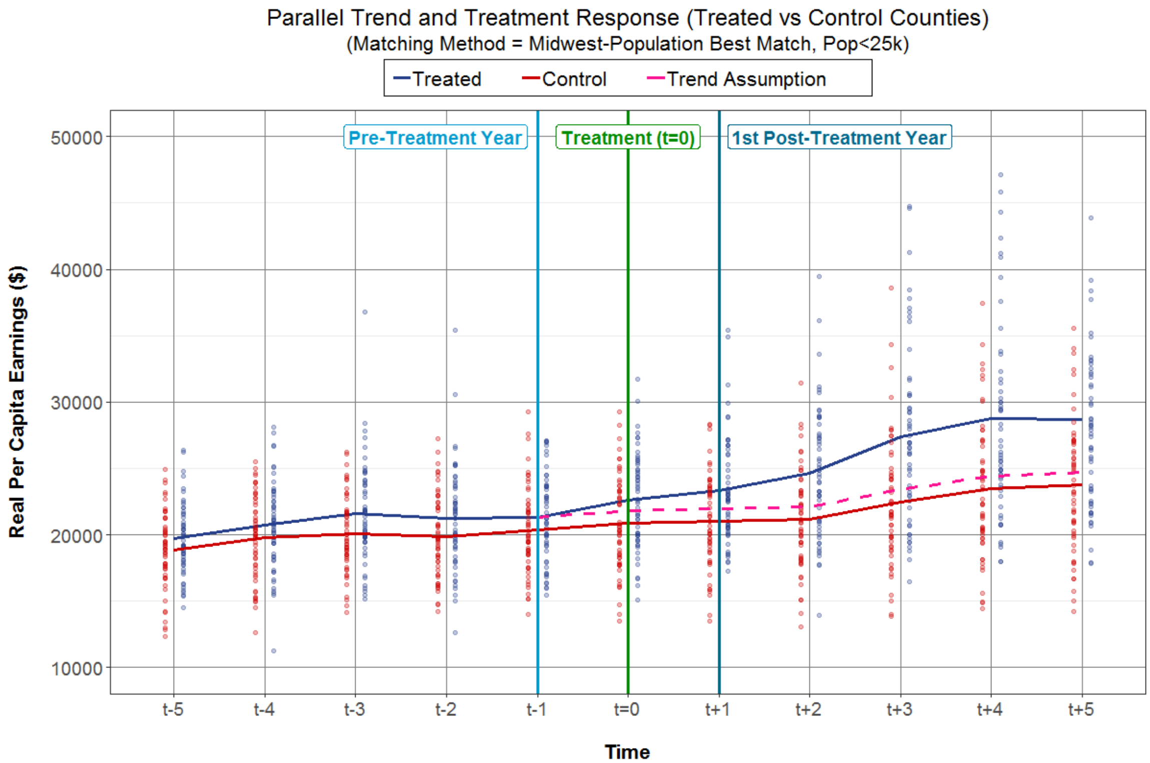

Figure 5.2 Parallel Tread and Treatment Response (MW-Pop Match, Pop<25k)

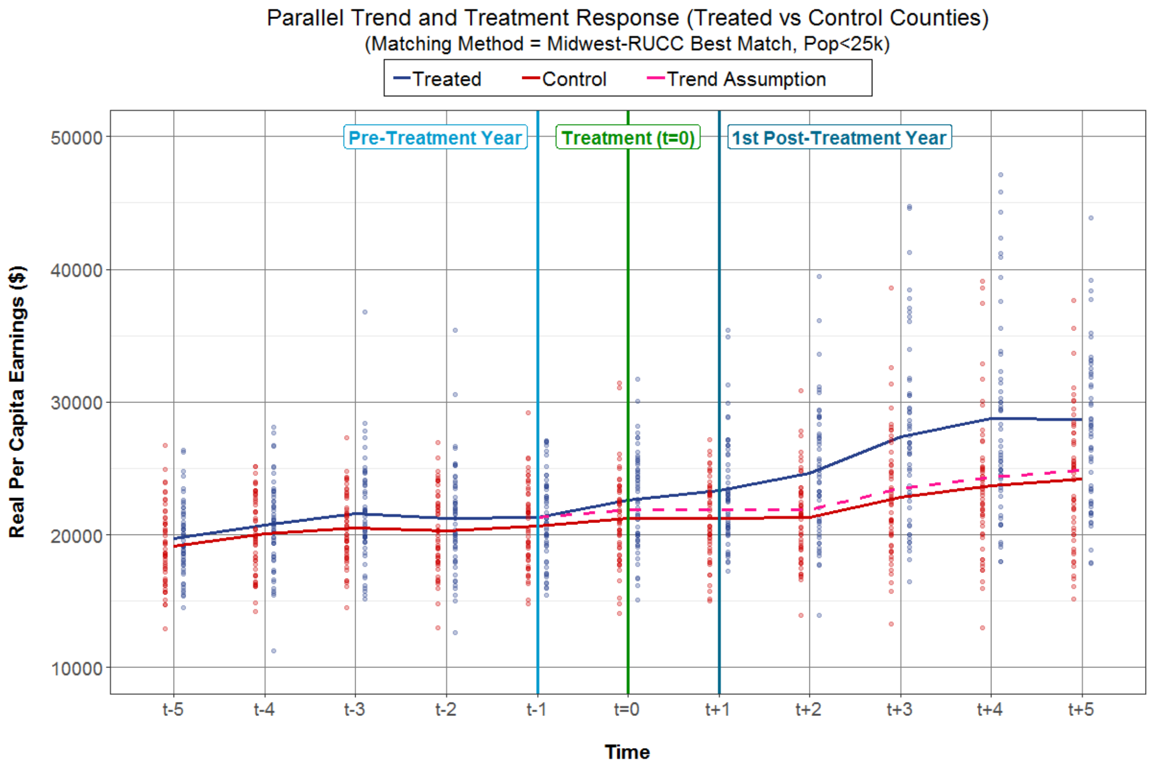

Figure 5.3 Parallel Tread and Treatment Response (MW-RUCC, Pop<25k)

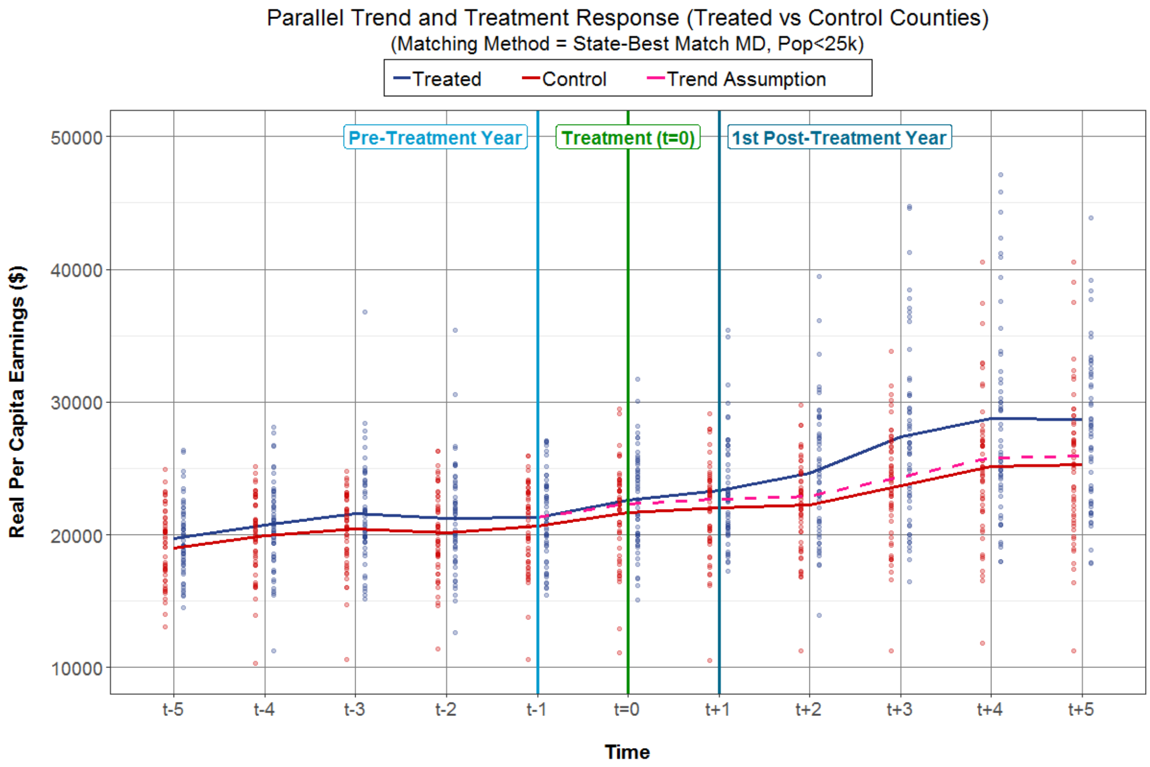

Figure 5.4 Parallel Tread and Treatment Response (State-Best Match, Pop<25k)

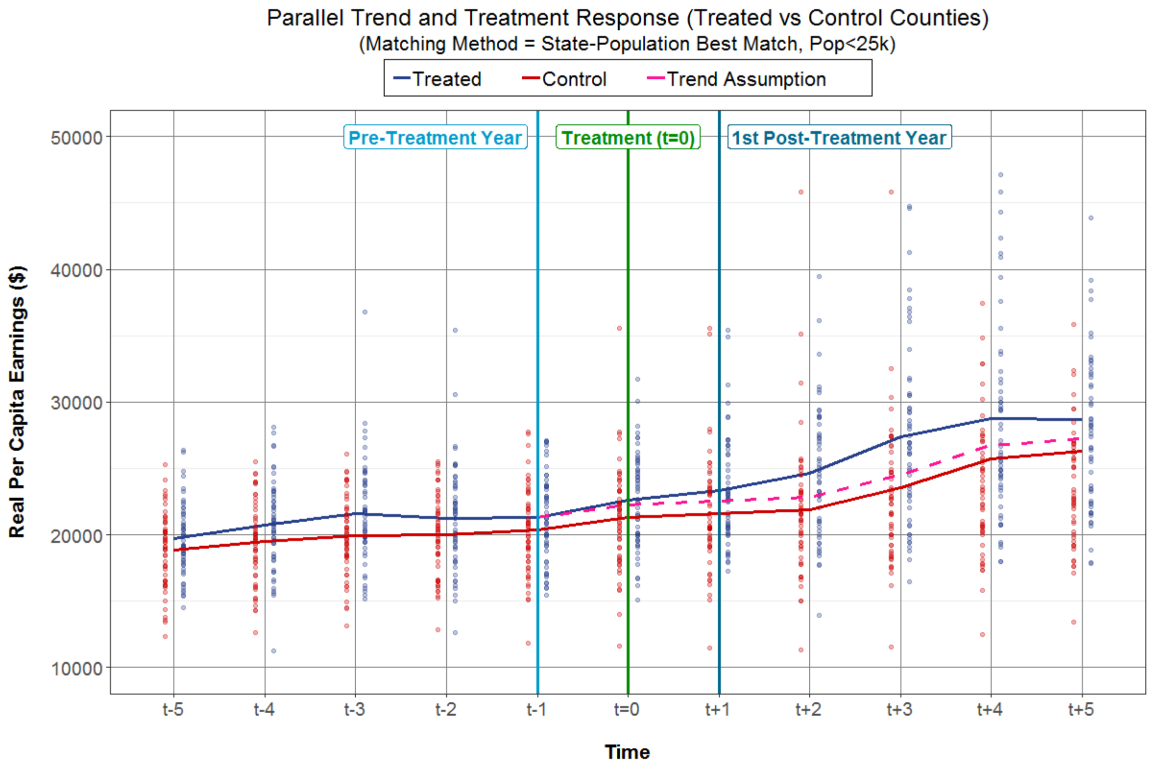

Figure 5.5 Parallel Tread and Treatment Response (State-Pop Match, Pop<25k)

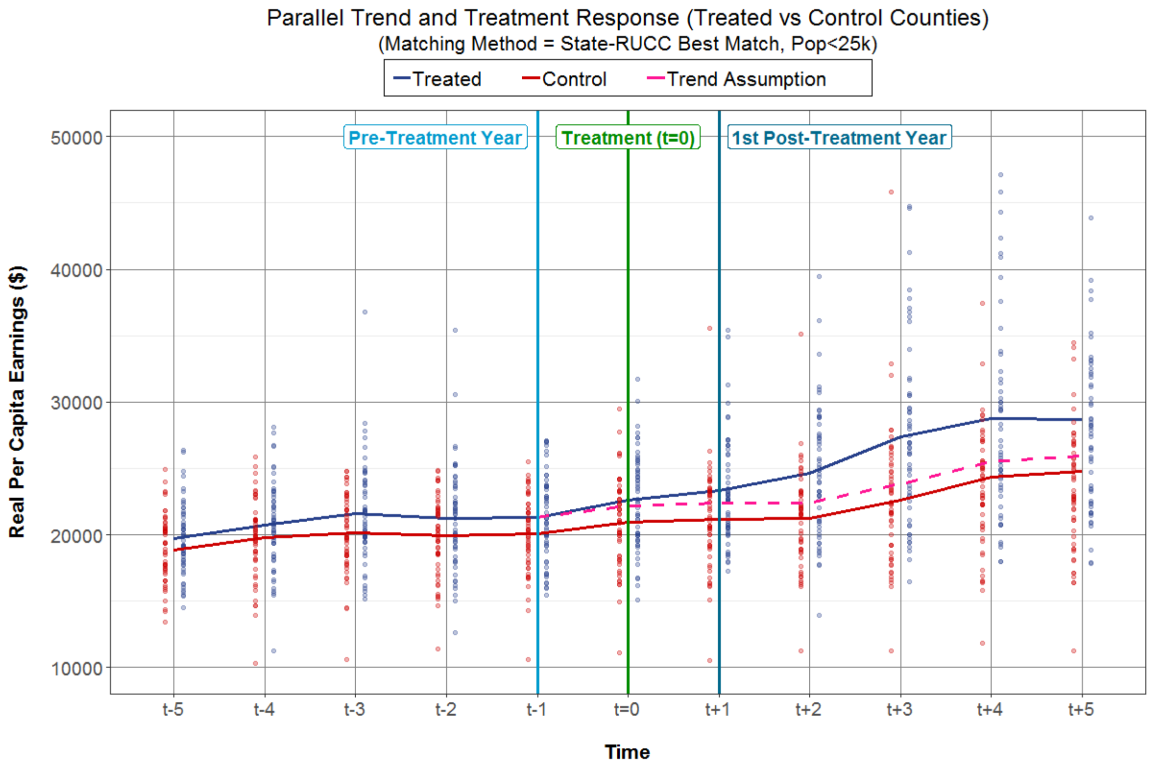

Figure 5.6 Parallel Tread and Treatment Response (State-RUCC, Pop<25k)

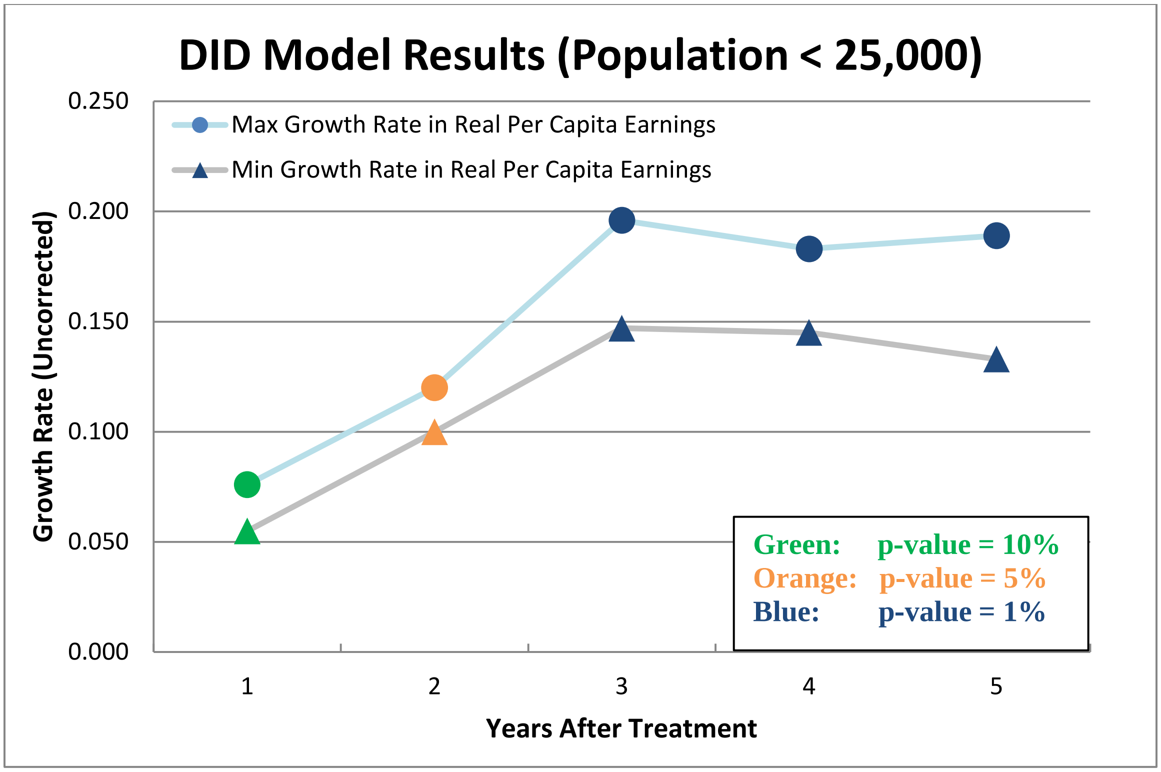

Figure 5.7 Max and Min Growth Rates (Period after Treatment)

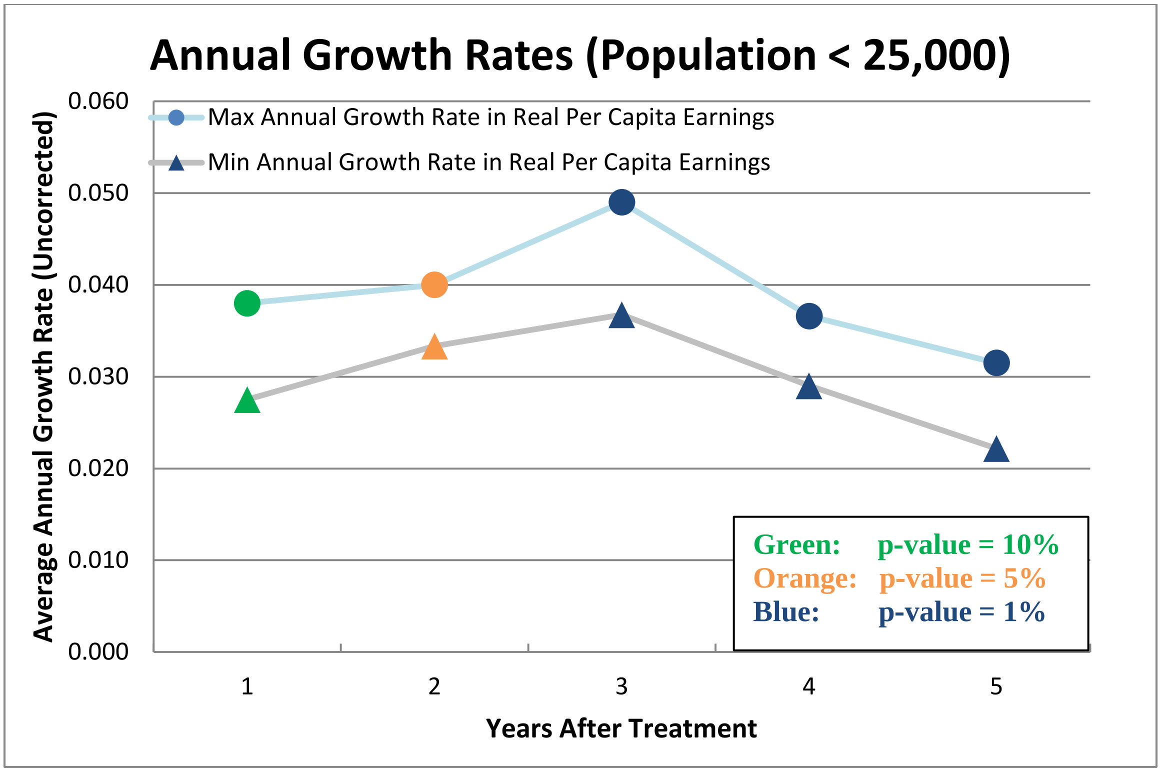

Figure 5.8 Max and Min Average Annual Growth Rates (Period after Treatment)

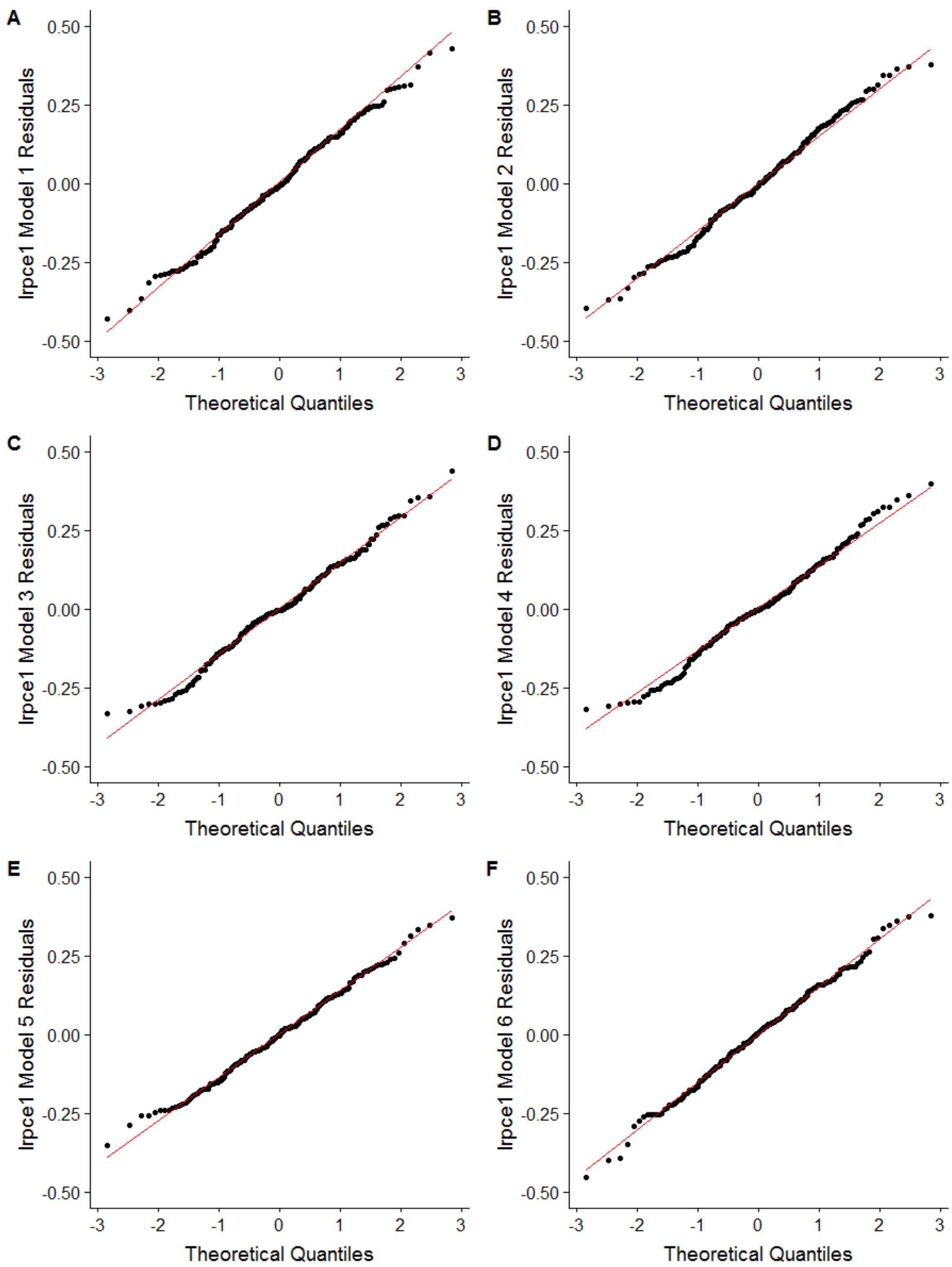

Figure 5.9 Normal QQ-Plot for lrpce1 DID Model Residuals

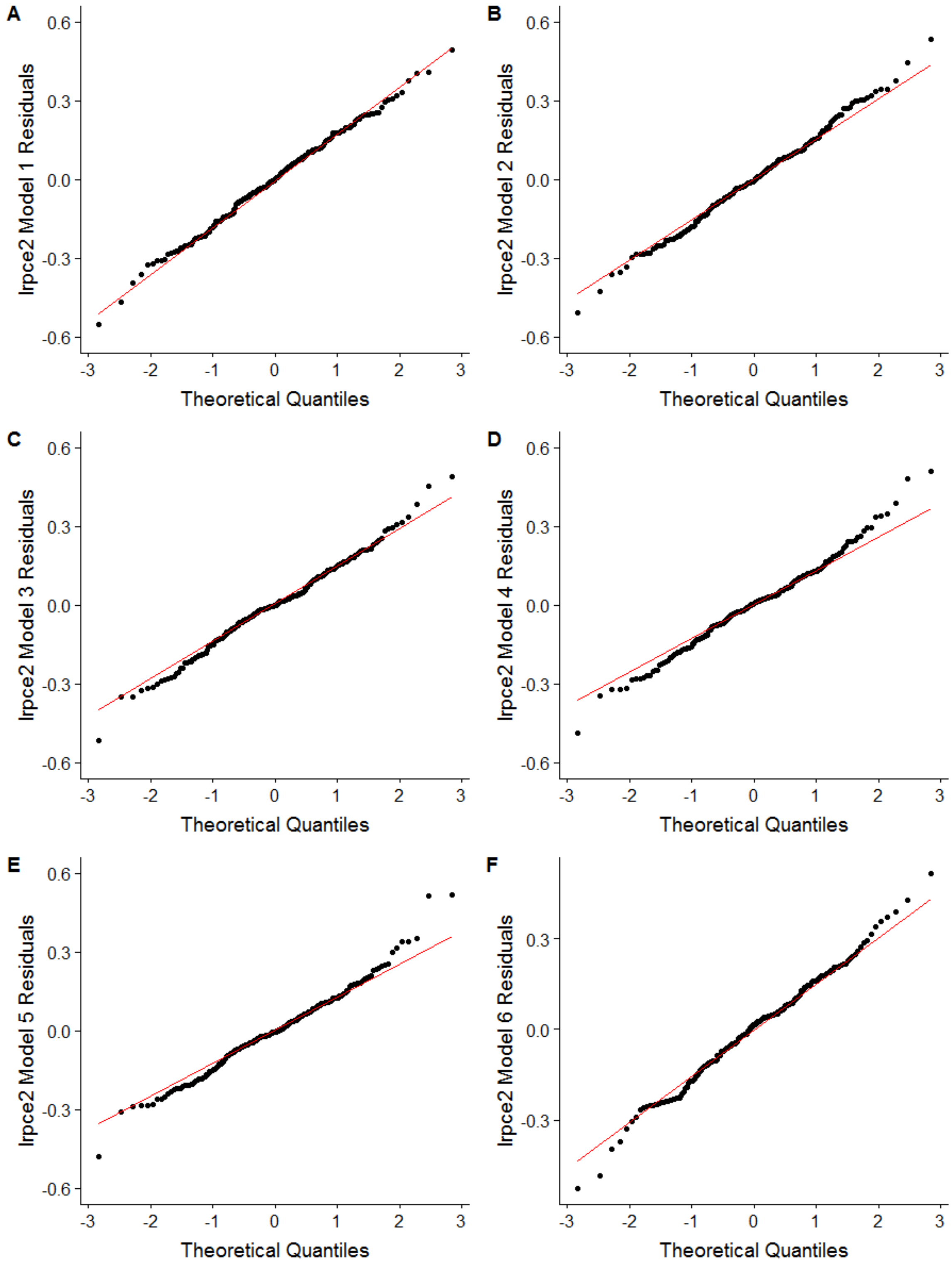

Figure 5.10 Normal QQ-Plot for lrpce2 DID Model Residuals

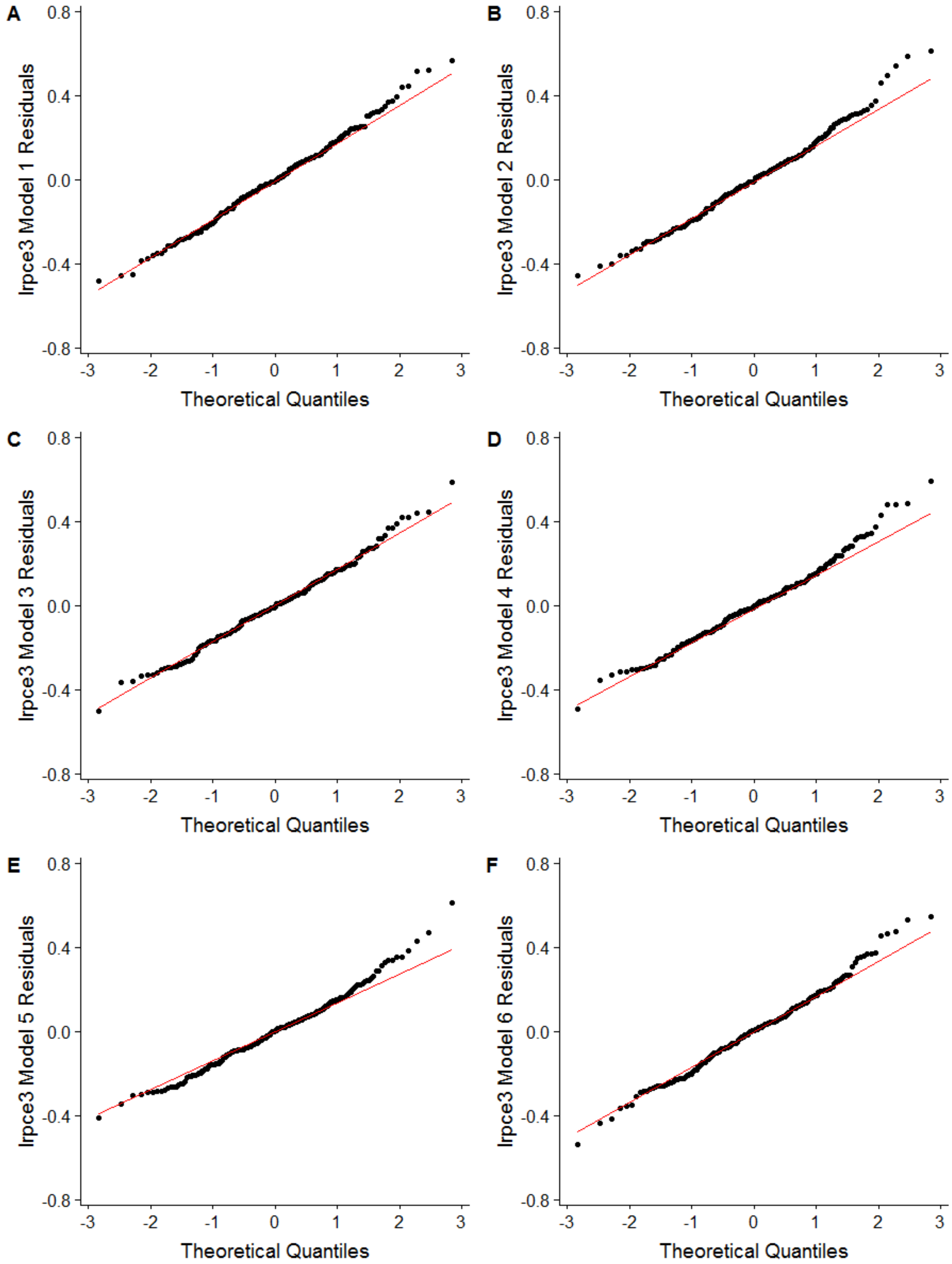

Figure 5.11 Normal QQ-Plot for lrpce3 DID Model Residuals

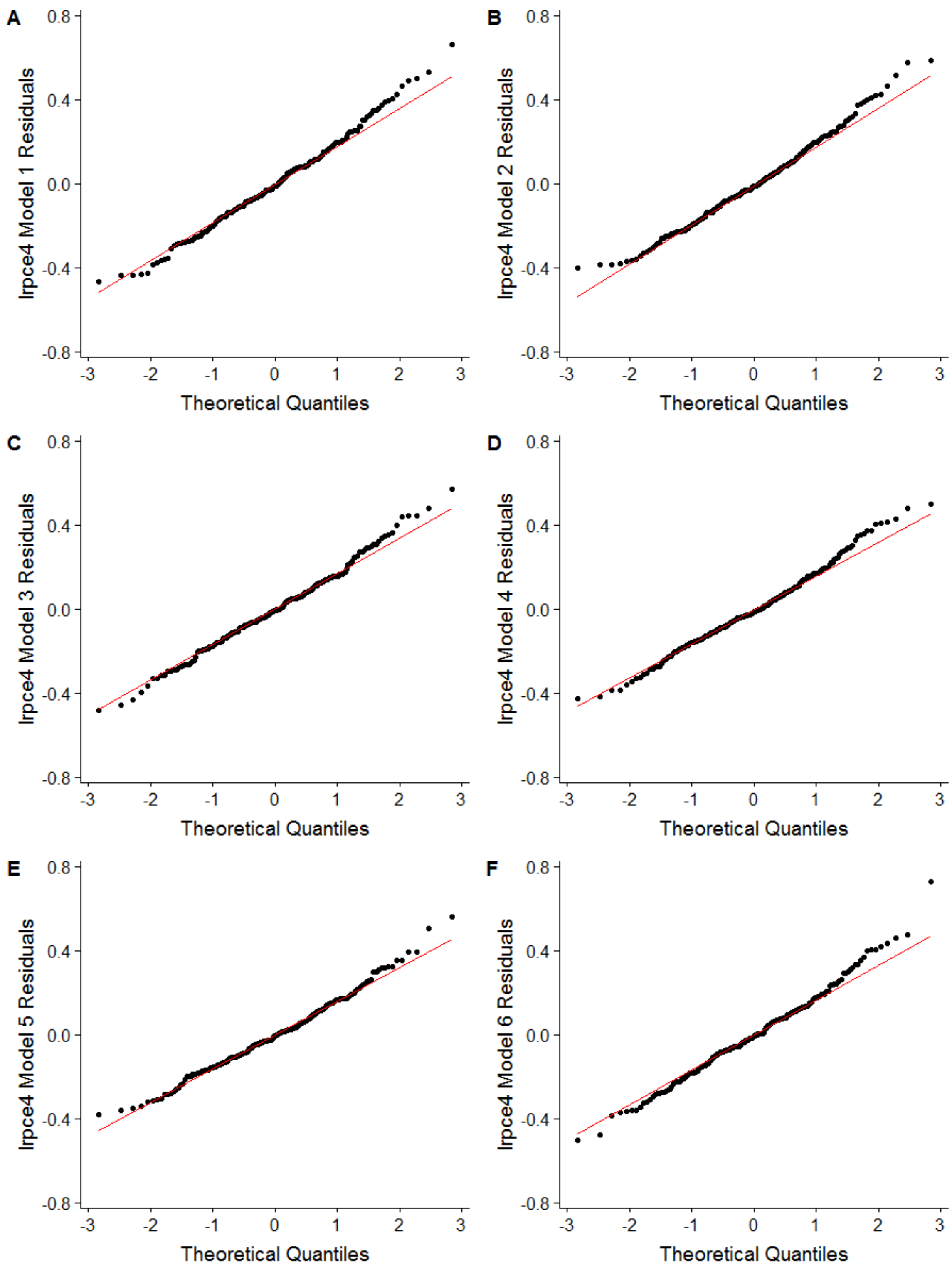

Figure 5.12 Normal QQ-Plot for lrpce4 DID Model Residuals

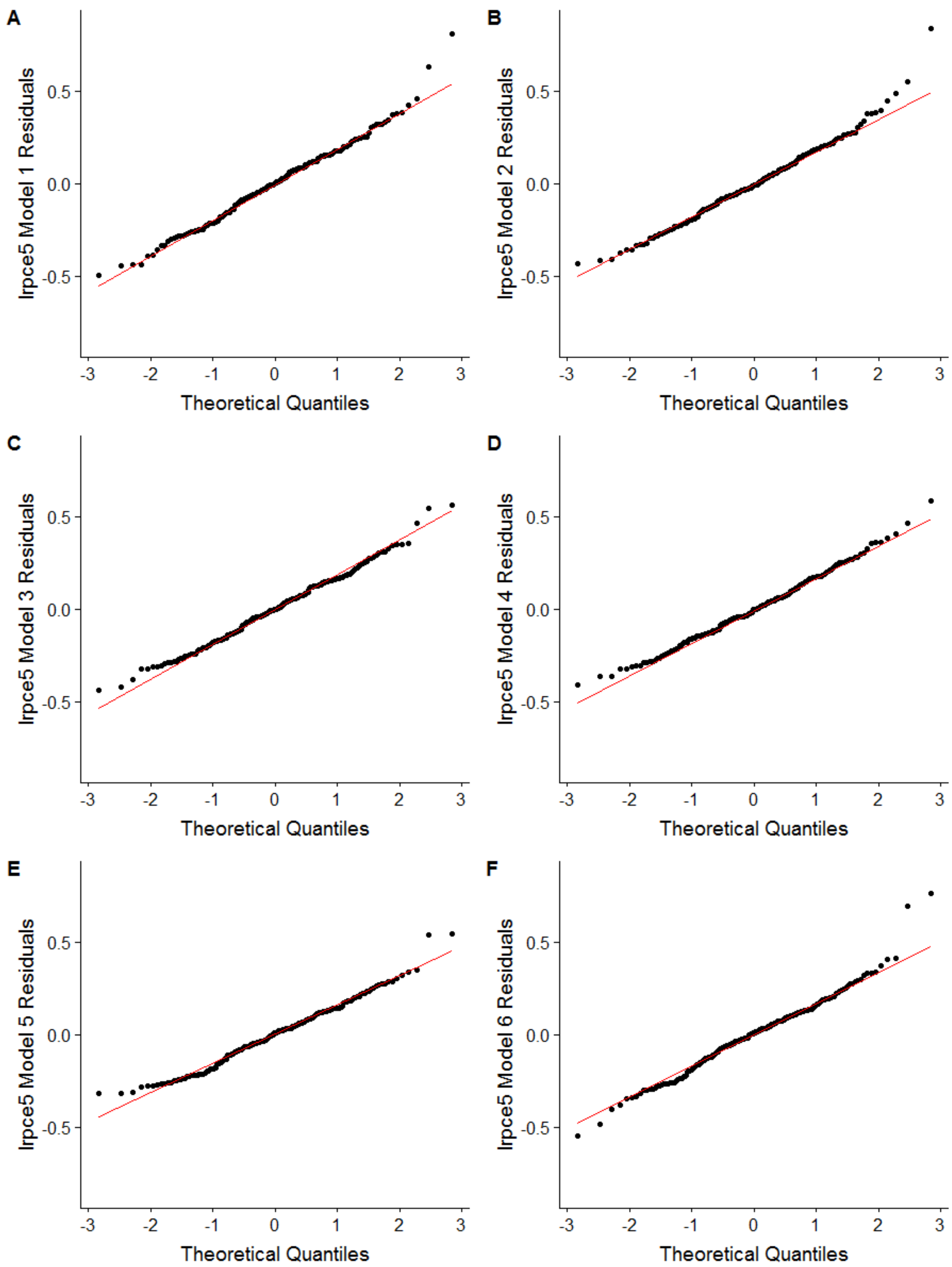

Figure 5.13 Normal QQ-Plot for lrpce5 DID Model Residuals

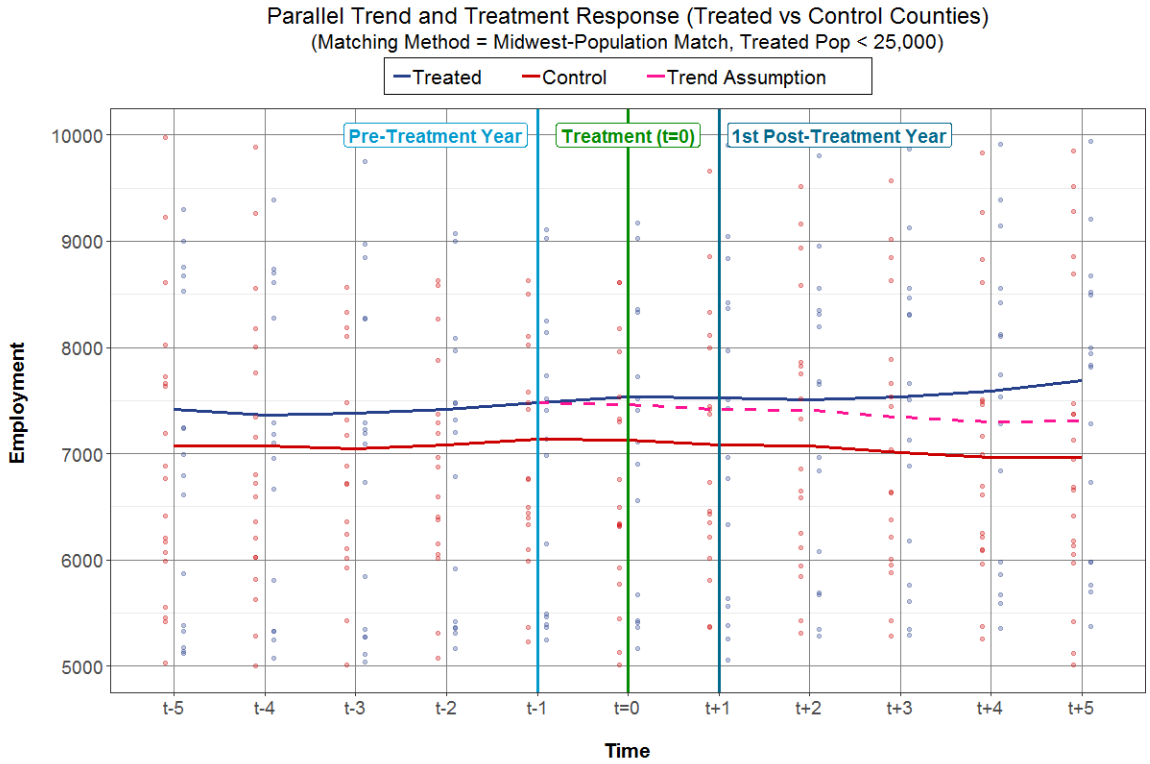

Figure 5.14 Employment Parallel Trend & Response (MW-Pop Match, Pop<25k)

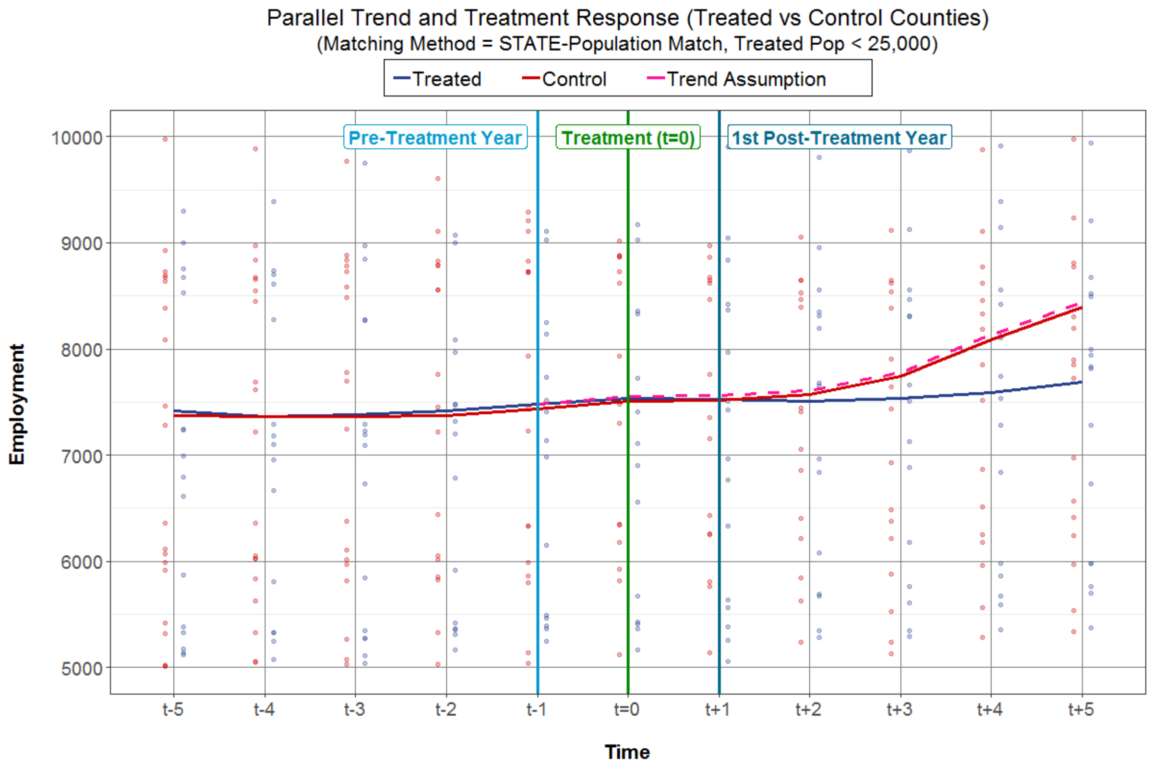

Figure 5.15 Employment Parallel Trend & Response (ST-Pop Match, Pop<25k)

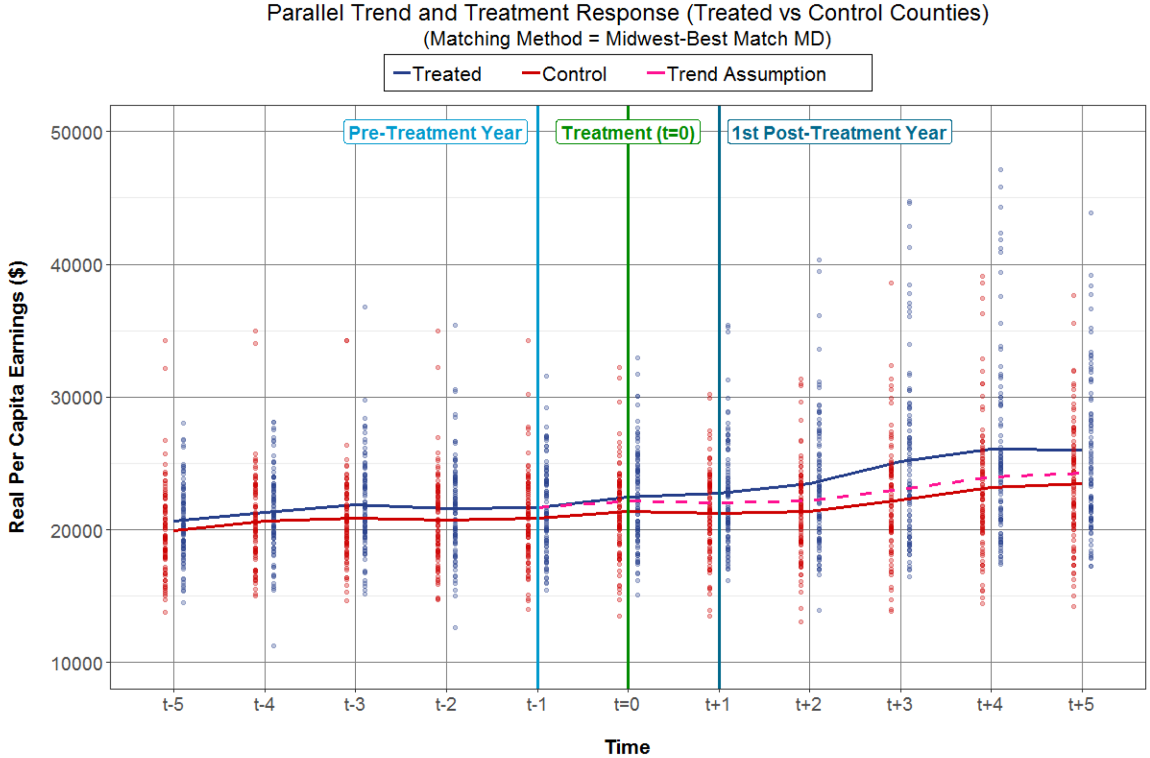

Figure 5.16 Parallel Trend and Treatment Response (Midwest-Best Match MD)

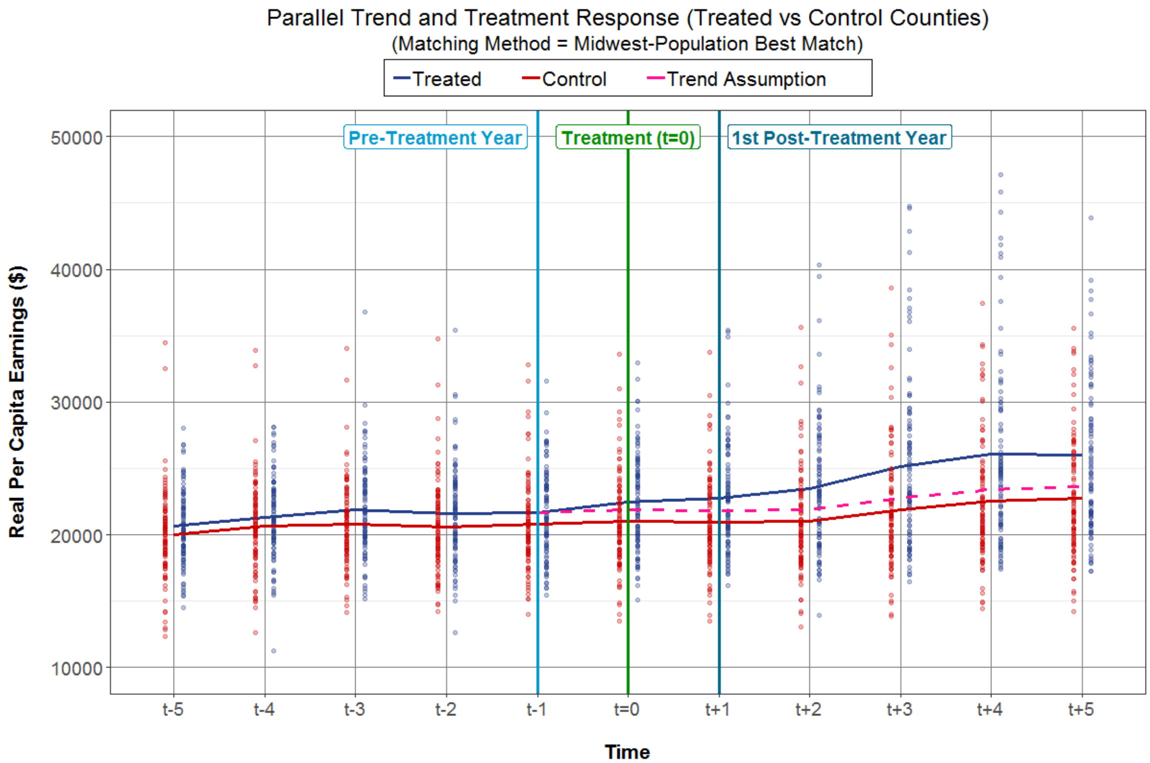

Figure 5.17 Parallel Trend and Treatment Response (Midwest-Population Match)

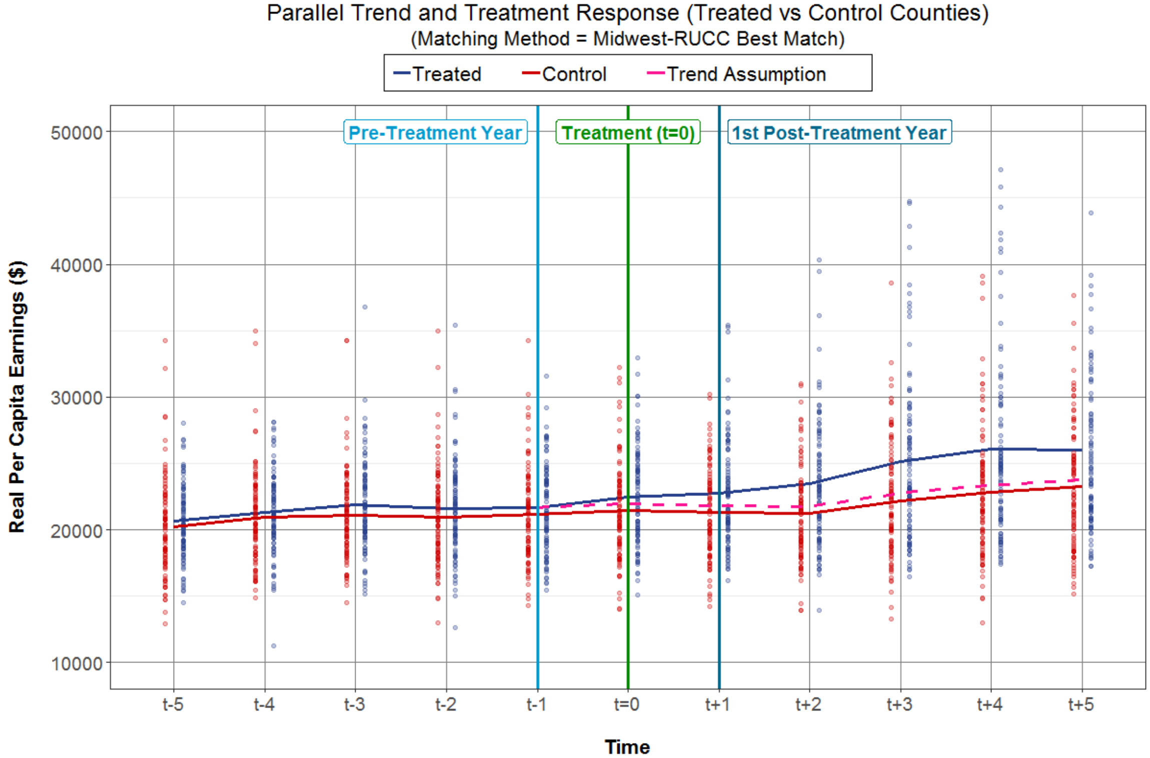

Figure 5.18 Parallel Trend and Treatment Response (Midwest-RUCC)

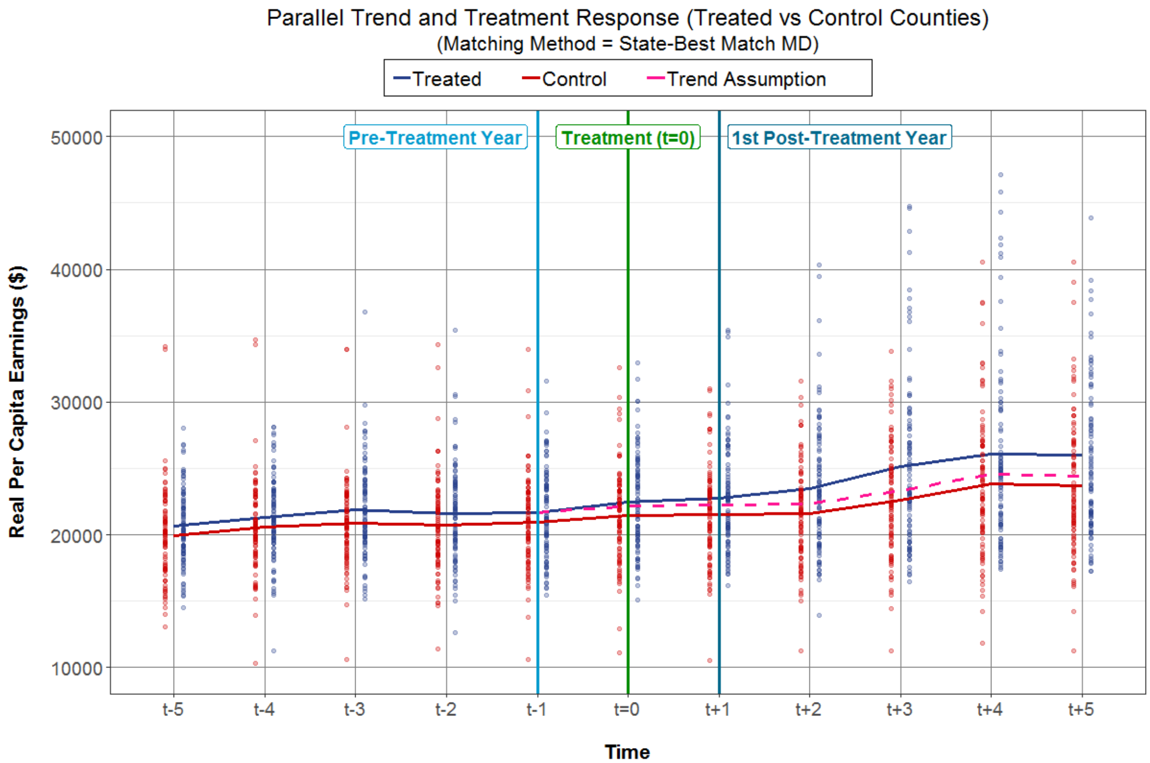

Figure 5.19 Parallel Trend and Treatment Response (State-Best Match MD)

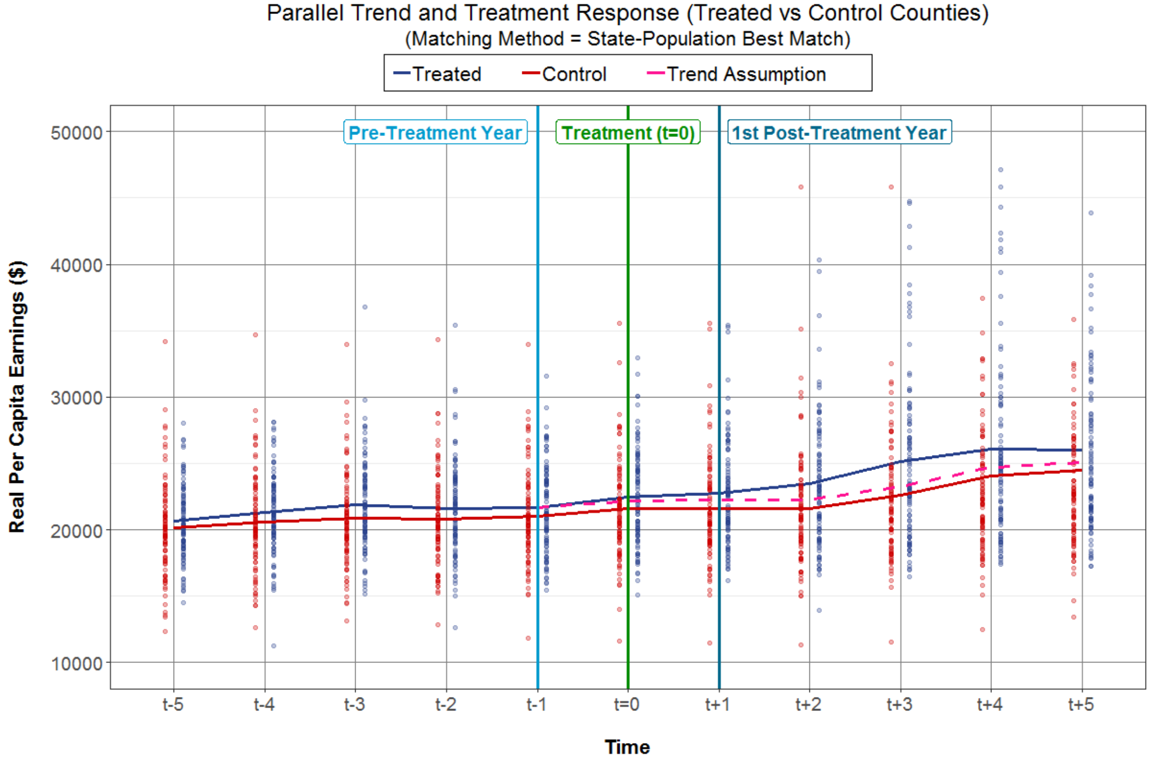

Figure 5.20 Parallel Trend and Treatment Response (State-Population Match)

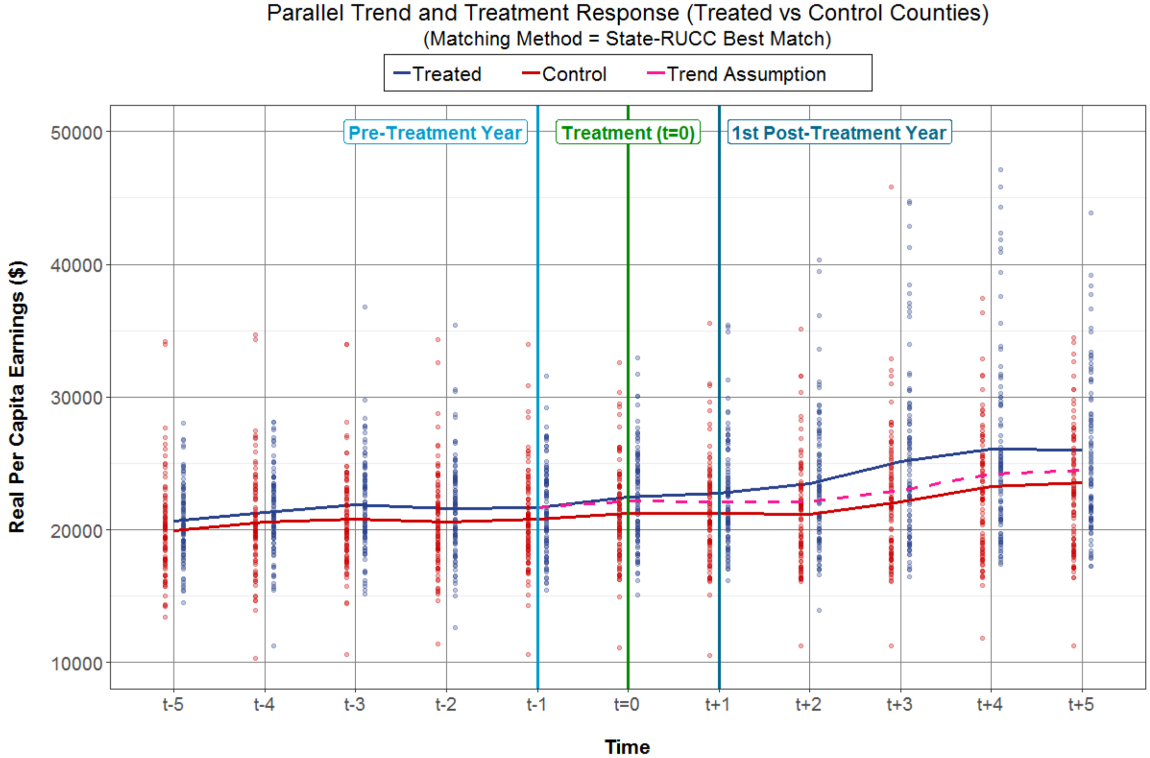

Figure 5.21 Parallel Trend and Treatment Response (State-RUCC)

lrpce1	Low Capacity	High Capacity
Low Population	p=0.1185	p=0.189
High Population	Not Significant	Not Significant

lrpce2	Low Capacity	High Capacity
Low Population	δ = 0.1191 se = 0.0564 p=0.0185	p=0.133
High Population	Not Significant	Not Significant

lrpce3	Low Capacity	High Capacity
Low Population	δ = 0.1994 se = 0.0643 p=0.001	p=0.1855
High Population	Not Significant	Not Significant

lrpce4	Low Capacity	High Capacity
Low Population	δ = 0.1664 se = 0.0695 p=0.009	δ = 0.111 se = 0.067 p=0.0515
High Population	Not Significant	Not Significant

lrpce5	Low Capacity	High Capacity
Low Population	δ = 0.159 se = 0.0678 p=0.0105	δ = 0.090 se = 0.0626 p=0.077
High Population	Not Significant	Not Significant

Note 1: Low population < 25,000
Note 2: 25,000 <= High Population <= 250,000
Note 3: Low capacity <= 55 mgy
Note 4: 55 mgy < High capacity <= 120 mgy
Note 5: δ = Average Treatment on the Treated (ATOT)
Note 6: Control Group = Midwest Population Match

Figure 5.22 Population versus Initial Capacity Diagrams

CHAPTER 6. CONCLUSIONS AND IDEAS FOR FUTURE RESEARCH

6.1 Conclusions

This dissertation examined the economic impact of ethanol biorefineries on rural counties and found that treated rural counties, on average, economically benefitted from significant growth in real per capita earnings that exceeded real per capita earnings growth for a several sets of control counties (counterfactuals). Six difference-in-differences (DID) models using a Midwest population match control group with different combinations of control variables (spillover effects, state fixed effects, and time fixed effects) produced mostly significant results for the average treatment on the treated (ATOT) coefficient over the period from one year after treatment to five years after treatment. A robustness check across all six control groups showed variations in the level and significance of growth, but only results for one year after treatment had multiple models that failed to reject the null hypothesis. The second and third years after treatment showed positive growth in real per capita earnings for the treated counties relative to the control groups. In the fourth and fifth years after treatment, the treated counties, on average, did not have real per capita earnings growth while the control group counties began to show signs of growth.

Treated county employment increased by 211 jobs, on average, over a five year period after treatment. In the Midwest population match control group, employment dropped, on average, 161 jobs over the same five year period after treatment. Thus, treated counties added new jobs and avoided the loss of jobs which effectively means that 380 jobs, on average, are attributable to an ethanol biorefinery being located in a rural county up to five years after treatment. This translates into an employment multiplier effect for treated counties of 1.46 in the year of treatment and rises to 7.6 five years after treatment assuming direct employment of 50 employees at the ethanol biorefinery. In a comparison with the State population match control group, there is an unexplained rise in the control group's employment numbers. Based on this unexplained rise in State control county employment, no conclusion can be drawn from these results. The employment rate in the rural treated counties increased by 2.2% over the five year period after treatment which is slightly better than the two control groups used in this comparison.

Exploratory research showed that ethanol biorefineries with initial production capacity of 55 mgy or less had greater economic impact on rural treated counties than ethanol biorefineries with higher initial capacities. A deep dive into the data found that 13 of the 34 low capacity ethanol biorefineries were locally owned. This could be one reason for the more significant rural impact, but a more thorough investigation is required to fully understand this situation.

A review of the threats to internal validity was presented. Though there are several violations of treatment assumptions for quasi-experiments (primarily, spillover effects), a logical evaluation of these violations seemed to indicate that any model evaluation would find an underestimate of the parameter of interest relative to the true value. Therefore, it can be debated on whether the spillover indicator variables properly captured the spillover effects and whether the significant ATOT results found in those models represents a close approximation of the true value. In some sense, it could still be an underestimation of the true value.

Overall, strong results were presented which clearly suggest that in two to the five years after treatment the null hypothesis can be rejected in favor of the alternative hypothesis. Though there was only growth in the first three years after treatment relative to the control groups, the gap in real per capita earnings only closed slightly with the control groups in the fourth and fifth years after treatment. Based on the slowed growth in the treated group and the rise in growth of the control group, there are potentially two interacting events working in the fourth and fifth year after treatment. For the treated group, it may be experiencing regression to the mean. The underlying factors that are forcing this regression to the mean could be ethanol biorefinery plant efficiency and excess corn production. Excess corn production is probably the major factor since it will drive down corn prices which will impact the local economy. For the control groups, they could be experiencing the benefit that a rising tide lifts all ships (eventually, but perhaps by not the same amount). This may be counterintuitive, but more of the economic benefits of the treated counties may start to flow in the direction of the control counties in the fourth and fifth years after treatment (lagged impact). Regardless, longer term research is required to more fully understand these events.

This dissertation demonstrated that positive rural economic development impacts did occur as part of the Renewable Fuel Standard (RFS) which drove the expansion of the ethanol fuel industry in the U.S. Midwest. Counties in the U.S. Midwest treated with an ethanol biorefinery experienced, on average, positive growth in real per capita earnings and increases in employment relative to a Midwest population match control group. Thus, the rural economic development impacts proposed by the RFS have been confirmed by the research performed in this dissertation. Though the initial economic impacts may seem to be short-lived, it is anticipated that the direct, indirect, and induced effects will drive the U.S. Midwest economy toward a new balanced-growth equilibrium.

6.2 Ideas for Future Research

6.2.1 Explore Employment Growth in State Control Counties

It was discovered that State control county matches (counterfactuals) for counties with populations less than 25,000 had a significant increase in employment relative to the treated counties. Since this result is counterintuitive especially when compared to the Midwest control counties, an investigation into this issue would help to provide insight into why this occurred.

6.2.2 Explore Large Capacity Ethanol Biorefineries in Rural Counties

The exploratory analysis on capacity versus treated county population showed that ethanol biorefineries with lower initial production capacities had a greater impact on treated rural counties than did ethanol biorefineries with large initial production capacities. This seems counterintuitive and a more thorough investigation should provide a greater understanding of this result.

6.2.3 Longer Term Study

A longer term study could reveal whether the gap between treated and control counties earnings continued to widen or whether it began to converge. The results in this analysis show that the treated county earnings growth began to slow dramatically in the fourth year after treatment and that control county earnings began to rise. A study of 10 to 15 years after treatment could reveal a new parallel gap forming between the treated and control groups.

6.2.4 Apply Panel Data Techniques

Panel data models might present another approach for determining the parameters of interest associated with the hypothesis in this research. This approach could potentially identify unique treatment effects based on the underlying characteristics of the treated county.

6.2.5 Synthetic Control Groups

Often, only one set of controls is used to assess whether a quasi-experimental analysis of the treatment effect was significant. This dissertation shows that the overall treatment effect can vary relative to the control groups depending on the criteria used in the matching process. Synthetic control groups attempt to form equivalent control groups as opposed to the non-equivalent control groups used in this dissertation. It would be interesting to see how well the synthetic control groups actually match the treated groups and whether by using a synthetic control group method it might reduce the impact associated with spillover effects at least in the control group.

6.2.6 Coarsened Exact Matching

Coarsened exact matching is another control group matching algorithm proposed by Iacus, King, and Porro (2008). It primarily reduces the imbalance in covariates between treated and control groups and as a result claims to improve estimation of causal effects.

6.2.7 Spatial Difference in Differences

Spatial analysis has been improving over recent years due to the continued improvement in computational speed. GeoDa software can run spatial econometric analysis on well-defined spatial problems. The problem often arises of how to handle the analysis when there is a temporal dimension. Work by Jean Dubé and Diègo Legros (2012, 2014b) and Jean Dubé et al. (2014a, 2017) are advancing spatial econometrics techniques which could enable this problem to be solved by those techniques in the future.

APPENDICES

A 1. MAHALANOBIS DISTANCE METRIC MIDWEST

```
################################################################################
#
#               Midwest Ethanol Biorefinery Analysis:
#               Mahalanobis Distance Calculations to
#               match Treatment with Control Counties
#               for all Midwest counties.
#
#                       By: Scott W. Hall
#                       January 4, 2019
#
#
################################################################################
# Load Required libraries

library(plyr)       # access to empty function
library(rJava)      # Used to read/write xlsx files
library(xlsx)       # Used to read/write xlsx files


# Set working directory to Selection Variables directory

path <- "C:/Users/ScottH/Documents/MidwestData/"
sv <- paste(path,"SV", sep="")
setwd(sv)

# Setup State arrays:  Used to extract state data, indexing and writing output.

state.abv <- c("IL", "IN", "IA", "KS", "MI", "MN", "MO", "NE", "ND", "OH", "SD", "WI")
state.full <- c("Illinois", "Indiana", "Iowa", "Kansas", "Michigan", "Minnesota",
        "Missouri", "Nebraska", "NorthDakota", "Ohio", "SouthDakota", "Wisconsin")

# All county data are aligned by state (and fips id) in all relevant data files
# State Breaks by row #:  Illinois[1,102], Indiana[103,194], Iowa[195, 293], Kansas[294,398],
#     Michigan[399,481], Minnesota[482,568], Missouri[569,683], Nebraska[684,776],
#     North Dakota[777,829], Ohio[830,917], South Dakota[918,983], Wisconsin[984,1055]
# state.breaks represent the start row of each state's data + last overall value to close out set

state.breaks <- c(1,103,195,294,399,482,569,684,777,830,918,984,1056)

# Set initial production years to be analyzed (also known as treatment years)

prod.year <- c(2001:2015)

# Load Midwest Datasets (13 csv files plus a Midwest counties file)

temp <- read.csv("Midwest_POP.csv", header=TRUE)
Pop <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_PopGrowthRate.csv", header=TRUE)
PopG <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_PopDensityLA.csv", header=TRUE)
PopDen <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_RPEG.csv", header=TRUE)
RPEG <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_RPCE.csv", header=TRUE)
RPCE <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_Corn.csv", header=TRUE)
```

```
corn <- as.matrix(temp[,-c(2:3)]); rm(temp)
temp <- read.csv("Midwest_Soy.csv", header=TRUE)
soy <- as.matrix(temp[,-c(2:3)]); rm(temp)
temp <- read.csv("Midwest_FRMSH.csv", header=TRUE)
frmsh <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_MFGSH.csv", header=TRUE)
mfgsh <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_RETSH.csv", header=TRUE)
retsh <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_FMRCR.csv", header=TRUE)
fmrcr <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_CornAcres.csv", header=TRUE)
cornAC <- as.matrix(temp[,-c(2:3)]); rm(temp)
temp <- read.csv("Midwest_EMPLY.csv", header=TRUE)
emply <- as.matrix(temp[,-c(2:4)]); rm(temp)
mw.counties <- read.csv("Midwest_Counties.csv", header=TRUE)
rucc <- read.csv("Midwest_RUCC.csv", header=TRUE)

# change working directory to output directory
output <- paste(path,"OutputMidwest", sep="")
setwd(output)

# Load Treated Counties and pre-2001 counties with biorefineries and other
# counties to be excluded from the control counties list
# treated.list defines path to Ethanol Biorefinery file; mw.treated reads file
treated.list <- paste(path, "Biorefinery/TreatedBiorefineryLocations_Final.csv", sep="")
mw.treated <- read.csv(treated.list, header=TRUE, sep=",")

# Extract selected columns from the ethanol biorefinery master list
mw.eth <- mw.treated[,c(1,4,10,5)]   # Columns extracted (Fips, County, State, ProdYear)
attach(mw.eth)
mw.eth <- mw.eth[order(Fips),]  # Order rows by Fips value
detach(mw.eth)

# qq is the master data matrix with all selection variable information
# for the pre-treatment period prior to first production.
# Dimension qq array (1055 = all counties, 29 = # of selection variables)
        qq = array(0, dim=c(1055,29))

# Dimension Variance-Covariance Matrix (s) & Inverse Matrix (sinv)
        s = array(0, dim=c(29,29))
        sinv = array(0, dim=c(29,29))

# Loop to fill qq-matrix for each time period (Treatment Year), calculate var-cov matrix and inverse
for (i in 1:15) {


 # eth.full = Extract Treated countes based on Treatment Year (initial production year)
 # Test if dataframe is empty (i.e. no Treated Counties); if so, next i (year).
 # If Statement is used to trap any year without a new biorefinery (empty set)
 # and then skip to next year.

 if (empty(subset(mw.eth, ProdYear == prod.year[i]))) {
   next
 }

 eth.full <- subset(mw.eth, ProdYear == prod.year[i])

 # Fill qq matrix with data from input files for a particular treatment year (prod.year[i])
 # The index i is used to step through the longitudinal data set for each treatment year.

        qq[,] = cbind(PopG[,(1+i):(5+i)],RPEG[,(1+i):(5+i)],Pop[,(6+i),drop=FALSE],
```

```
            PopDen[,(6+i),drop=FALSE],cornAC[,(6+i),drop=FALSE],corn[,(1+i):(5+i)],
            soy[,(1+i):(5+i)],RPCE[,(6+i),drop=FALSE],frmsh[,(6+i),drop=FALSE],
            mfgsh[,(6+i),drop=FALSE],retsh[,(6+i),drop=FALSE],fmrcr[,(6+i),drop=FALSE],
            emply[,(6+i),drop=FALSE])

# qq is a character matrix -> convert into a numeric matrix
            class(qq) <- "numeric"
# Population is in q[,11] for each county; find counties with pop>200,000 (these will be excluded)
            popgt200k <- which(qq[,11]>200000)

  # qq matrix can be ill-conditioned for var-cov inverse calculations
  # Indexing Corn & Soybean Production and Corn Arces resolves the inverse problem
            # where the Max value during the 5 year pre-conditioning period is set to be 100
            # and all other values are set relative to the Max value.
  # Convert Corn Production into an Index over the 5 year period [columns 14:18]
            cornmax = max(qq[,14:18])/100
            qq[,14:18] = qq[,14:18]/cornmax

  # Convert Soybean Production into an Index over the 5 year period [columns 19:23]
            soymax = max(qq[,19:23])/100
            qq[,19:23] = qq[,19:23]/soymax

  # Convert Corn Acres into an Index to rescale this variable [column 13]
            cornACmax = max(qq[,13])/100
            qq[,13] = qq[,13]/cornACmax


# Catch matrix ill-conditioning due to high population areas (eliminate pop > 200k)
# Only include data for counties with pop < 200k in qq matrix to form iq2 matrix.
# Calculate Variance-Covariance Matrix

            iq2 = array(0, dim=c(1055-length(popgt200k),29))
            iq2[,]=qq[qq[,11]<200000,]
            s[,] = var(iq2[,])

# Calculate Variance-Covariance Matrix Inverse
            sinv[,] = solve(s[,])

# Initialize Mahalanobis Distance Metric array for all counties in US Midwest

md=array(0,dim=c(1055,1055))

# Calculate full MD matrix for all Midwest counties against all Midwest counties
for (m in 1:1055) {
            for (n in 1:1055) {
                        x1=qq[m,]  # County m of state.abv[st.index] in prod.year[i]
                        x2=qq[n,]  # County n of state.abv[st.index] in prod.year[i]
                        xd=x1-x2   # Difference
                        md2= t(xd) %*% sinv[,] %*% xd    # Mahalanobis distance squared between Counties m and n
                        md[m,n] = sqrt(md2)              # Mahalanobis distance matrix (fill matrix)
            }
}

# Setup array of counties to index column for MD Control county extraction
# First, set county and state information as character to insure proper matching
# Reminder:  eth.full contains all treated counties for a particular prod.year[i]
eth.full$County <- as.character(eth.full$County)
eth.full$State <- as.character(eth.full$State)

# Extract county names for a particular state & treatment period defined by the MD matrix
md.names <- as.character(mw.counties$GeoName)
md.states <- as.character(mw.counties$State)
```

```
md.Fips <- mw.counties$GeoFIPS

# Create county indexed on midwest level data to extract column(s) from MD matrix
# First, extract treatment county names for in a particular year[i]
county.eth <- eth.full$Fips

# Second, assess number of counties and then dynamically dimension the county.index array
ci.length=max(1,length(eth.full$Fips))
county.index <- array(0,dim=c(ci.length))

# Loop to fill the county.index array with numerical location of treatment counties
# Explanation:  md.names contains all county names for a particular State[j] and is in the exact order
# as the county data in the data files.  By using the 'which' command on a treatment county
# against md.names, it extracts the numerical location of that treatment county within md.names
# which facilitates the extraction of that column (treatment county) from the MD matrix
# along with the MD calculations against all other counties in the state.
for (k in 1:length(county.eth)) {
county.index[k] <- which(county.eth[k] == md.Fips)
}

# Assign column and row names to facility sorting in Excel
colnames(md) <- paste(md.names,md.states,sep="_")
rownames(md) <- paste(md.names,md.states,sep="_")

# Extract all treated county columns in prod.year[i] from the Mahalanobis Distance Matrix
treat.county <- cbind(md[,county.index, drop=FALSE],rucc)

# Output all treated counties with their list of matches from the MD matrix into an Excel Workbook
# Sheets are organized by year of treatment (ie "Treated_2005")

filename <- "Treatment_Control_Midwest.xlsx"
sheet <- paste("Treated", prod.year[i],sep="_")
write.xlsx2(treat.county, filename, sheetName=sheet, col.names=TRUE,
        row.names=TRUE, append=TRUE, showNA=FALSE)


}                                    # Close bracket: i indexed for-loop (Treatment Year)
```

A 2. MAHALANOBIS DISTANCE METRIC IN-STATE

```
################################################################################
#
#             Midwest Ethanol Biorefinery Analysis:
#             Mahalanobis Distance Calculations to
#             match Treatment with Control Counties
#             within each State.
#
#                    By: Scott W. Hall
#                    January 4, 2019
#
################################################################################
# Load Required libraries

library(plyr)       # access to empty function
library(rJava)      # Used to read/write xlsx files
library(xlsx)       # Used to read/write xlsx files

# Set working directory to Selection Variables directory

path <- "C:/Users/ScottH/Documents/MidwestData/"
sv <- paste(path,"SV", sep="")
setwd(sv)

# Setup State arrays:  Used to extract state data, indexing and writing output.

state.abv <- c("IL", "IN", "IA", "KS", "MI", "MN", "MO", "NE", "ND", "OH", "SD", "WI")
state.full <- c("Illinois", "Indiana", "Iowa", "Kansas", "Michigan", "Minnesota",
        "Missouri", "Nebraska", "NorthDakota", "Ohio", "SouthDakota", "Wisconsin")

# All county data are aligned by state (and fips id) in all relevant data files
# State Breaks by row #:  Illinois[1,102], Indiana[103,194], Iowa[195, 293], Kansas[294,398],
#    Michigan[399,481], Minnesota[482,568], Missouri[569,683], Nebraska[684,776],
#    North Dakota[777,829], Ohio[830,917], South Dakota[918,983], Wisconsin[984,1055]
# state.breaks represent the start row of each state's data + last overall value to close out set

state.breaks <- c(1,103,195,294,399,482,569,684,777,830,918,984,1056)

# Set initial production years to be analyzed (also known as treatment years)

prod.year <- c(2001:2015)

# Load Midwest Datasets (13 csv files plus a Midwest counties file)

temp <- read.csv("Midwest_POP.csv", header=TRUE)
Pop <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_PopGrowthRate.csv", header=TRUE)
PopG <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_PopDensityLA.csv", header=TRUE)
PopDen <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_RPEG.csv", header=TRUE)
RPEG <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_RPCE.csv", header=TRUE)
RPCE <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_Corn.csv", header=TRUE)
corn <- as.matrix(temp[,-c(2:3)]); rm(temp)
temp <- read.csv("Midwest_Soy.csv", header=TRUE)
soy <- as.matrix(temp[,-c(2:3)]); rm(temp)
temp <- read.csv("Midwest_FRMSH.csv", header=TRUE)
frmsh <- as.matrix(temp[,-c(2:4)]); rm(temp)
```

```
temp <- read.csv("Midwest_MFGSH.csv", header=TRUE)
mfgsh <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_RETSH.csv", header=TRUE)
retsh <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_FMRCR.csv", header=TRUE)
fmrcr <- as.matrix(temp[,-c(2:4)]); rm(temp)
temp <- read.csv("Midwest_CornAcres.csv", header=TRUE)
cornAC <- as.matrix(temp[,-c(2:3)]); rm(temp)
temp <- read.csv("Midwest_EMPLY.csv", header=TRUE)
emply <- as.matrix(temp[,-c(2:4)]); rm(temp)
mw.counties <- read.csv("Midwest_Counties.csv", header=TRUE)
rucc <- read.csv("Midwest_RUCC.csv", header=TRUE)

# change working directory to output directory
output <- paste(path,"OutputState", sep="")
setwd(output)

# Load Treated Counties and pre-2001 counties with biorefineries and other
# counties to be excluded from the control counties list
# treated.list defines path to Ethanol Biorefinery file; mw.treated reads file
treated.list <- paste(path, "Biorefinery/TreatedBiorefineryLocations_Final.csv", sep="")
mw.treated <- read.csv(treated.list, header=TRUE, sep=",")

# Extract selected columns from the ethanol biorefinery master list
mw.eth <- mw.treated[,c(1,4,10,5)]   # Columns extracted (Fips, County, State, ProdYear)
attach(mw.eth)
mw.eth <- mw.eth[order(Fips),]  # Order rows by Fips value
detach(mw.eth)

# qq is the master data matrix with all selection variable information
# for an initial production year of a biorefinery
# Dimension qq array (1055 = all counties, 29 = # of selection variables)
        qq = array(0, dim=c(1055,29))

# Dimension Variance-Covariance Matrix (s) & Inverse Matrix (sinv)
        s = array(0, dim=c(29,29))
        sinv = array(0, dim=c(29,29))

# Loop to fill qq-matrix for each time period (Treatment Year), calculate var-cov matrix and inverse
for (i in 1:15) {

  # eth.full = Extract Treated countes/states based on Treatment Year (initial production year)
  # Test if dataframe is empty (i.e. no Treated Counties); if so, next i (year).
  # If Statement is used to trap any year without a new biorefinery (empty set)
  # and then skip to next year.

  if (empty(subset(mw.eth, ProdYear == prod.year[i]))) {
    next
  }

  eth.full <- subset(mw.eth, ProdYear == prod.year[i])

  # Fill qq matrix with data from input files for a particular treatment year (prod.year[i])
  # The index i is used to step through the longitudinal data set for each treatment year.

        qq[,] = cbind(PopG[,(1+i):(5+i)],RPEG[,(1+i):(5+i)],Pop[,(6+i),drop=FALSE],
        PopDen[,(6+i),drop=FALSE],cornAC[,(6+i),drop=FALSE],corn[,(1+i):(5+i)],
        soy[,(1+i):(5+i)],RPCE[,(6+i),drop=FALSE],frmsh[,(6+i),drop=FALSE],
        mfgsh[,(6+i),drop=FALSE],retsh[,(6+i),drop=FALSE],fmrcr[,(6+i),drop=FALSE],
        emply[,(6+i),drop=FALSE])

# qq is a character matrix -> convert into a numeric matrix
```

```
            class(qq) <- "numeric"
# Population is in q[,11] for each county; find counties with pop>200,000 (these will be excluded)
            popgt200k <- which(qq[,11]>=200000)

# st.unique = unique states vector for a particular treatment year (ProdYear[i])
# Or lists a state only once even if that state had multiple new biorefineries
#           in a particular treatment year
            st.unique <- unique(eth.full$State)

# st.level = length of unique state array and loop on this value
            st.levels <- length(st.unique)

            for (j in 1:st.levels) {


  # Use state index to extract data from qq matrix.  Since states are not aligned
  # in the st.unique array, must align referencing through State arrays which are
  # defined in alignment with the data.
            st.index <- which(st.unique[j]==state.abv)

  # md.index = number of counties in the state being analyzed (used to dimension MD matrix)
            md.index <- state.breaks[st.index+1]-state.breaks[st.index]

            # Dynamically define iq (number of counties in each state is different)
            # iq subsets the qq matrix (Midwest data) into state specific data
  iq = array(0, dim=c(md.index,29))

            iq[,] = as.numeric(qq[state.breaks[st.index]:(state.breaks[st.index+1]-1),])

  # iq matrix can be ill-conditioned for var-cov inverse calculations
  # Indexing Corn & Soybean Production and Corn Arces partially resolves the inverse problem
            # where the Max value during the 5 year pre-conditioning period is set to be 100
            # and all other values are set relative to the Max value.
  # Convert Corn Production into an Index over the 5 year period [columns 14:18]
            cornmax = max(iq[,14:18])/100
            iq[,14:18] = iq[,14:18]/cornmax

  # Convert Soybean Production into an Index over the 5 year period [columns 19:23]
            soymax = max(iq[,19:23])/100
            iq[,19:23] = iq[,19:23]/soymax

  # Convert Corn Acres into an Index to rescale this variable [column 13]
            cornACmax = max(iq[,13])/100
            iq[,13] = iq[,13]/cornACmax


# Eliminate high population areas greater than 200,000
# iq2 extracts all state-counties with population less than 200,000 from the iq (state matrix)
# Calculate Variance-Covariance Matrix

            iq2 = array(0, dim=c((md.index-length(which(iq[,11]>=200000))),29))
            iq2[,]=iq[iq[,11]<200000,]      # all counties with pop < 200,000
            s[,] = var(iq2[,])

# Calculate Variance-Covariance Matrix Inverse
            sinv[,] = solve(s[,])

# Dynamically dimension Mahalanobis Distance array
# reminder:  md.index <- state.breaks[st.index+1]-state.breaks[st.index]
md=array(0,dim=c(md.index,md.index))

# Calculate full MD matrix for within state counties against within state counties
```

```
for (m in 1:md.index) {
            for (n in 1:md.index) {
                        x1=iq[m,]  # County m of state.abv[st.index] in prod.year[i]
                        x2=iq[n,]  # County n of state.abv[st.index] in prod.year[i]
                        xd=x1-x2   # Difference
                        md2= t(xd) %*% sinv[,] %*% xd    # Mahalanobis distance squared between Counties m and n
                        md[m,n] = sqrt(md2)  }          # Mahalanobis distance matrix (fill matrix)
}

# Setup array of counties to index column for MD Control county extraction
# First, set county and state information as character to insure proper matching
# Reminder:  eth.full contains all treated counties for a particular prod.year[i]
eth.full$County <- as.character(eth.full$County)
eth.full$State <- as.character(eth.full$State)

# Extract county names for a particular state & treatment period defined by the MD matrix
md.names <- as.character(mw.counties$GeoName[state.breaks[st.index]:(state.breaks[st.index+1]-1)])

# eth.st = Counties with ethanol biorefineries for a particular state[j] in a particular year[i]
# Essentially, eth.st extracts only the treatment counties (rows) associated with a particular
# State (column array in eth.full) in prod.year[i]
eth.st <- subset(eth.full, State == st.unique[j])

# Create county indexed on state level data to extract column(s) from MD matrix
# First, extract treatment county names for particular state[j] in a particular year[i]
county.eth <- eth.st$County

# Second, assess number of counties and then dynamically dimension the county.index array
ci.length=max(1,length(county.eth))
county.index <- array(0,dim=c(ci.length))

# Loop to fill the county.index array with numerical location of treatment counties
# Explanation:  md.names contains all county names for a particular State[j] and is in the exact order
# as the county data in the data files.  By using the 'which' command on a treatment county
# against md.names, it extracts the numerical location of that treatment county within md.names
# which facilitates the extraction of that column (treatment county) from the MD matrix
# along with the MD calculations against all other counties in the state.
for (k in 1:length(county.eth)) {
county.index[k] <- which(county.eth[k] == md.names)
}

# Assign column and row names to facility sorting in Excel
colnames(md) <- md.names
rownames(md) <- md.names

# Extract RUCC codes by state
st.rucc <- rucc[state.breaks[st.index]:(state.breaks[st.index+1]-1),]

# Extract all treated county columns for state[j] in prod.year[i] from the Mahalanobis Distance Matrix
# cbind with RUCC information to assist with matching in post-processing
treat.county <- cbind(md[,county.index, drop=FALSE],st.rucc)

# Output all treated counties with their list of matches from the MD matrix into an Excel Workbook
# Sheets are organized by state and year of treatment (ie IA_2005)
filename <- "Treatment_Control_State.xlsx"
sheet <- paste(st.unique[j], prod.year[i],sep="_")
write.xlsx2(treat.county, filename, sheetName=sheet,col.names=TRUE,
      row.names=TRUE, append=TRUE, showNA=FALSE)

            }      # Close bracket:  j indexed for-loop (State)
}        # Close bracket:  i indexed for-loop (Treatment Year)
```

BIBLIOGRAPHY

Campbell, D. T., and J. C. Stanley (1963). *Experimental and Quasi-experimental Designs for Research*. Chicago, IL: Rand McNally

Card, David & Krueger, Alan B, (1994). "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," American Economic Review, American Economic Association, vol. 84(4), pages 772-793, September.

Cook, T. D., and D. T. Campbell (1979). *Quasi-experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Dubé, Jean, & Legros, Diègo. (2012). “Dealing with Spatial Data Pooled over Time in Statistical Models.” (Vol. 6). doi:10.1007/s12076-012-0082-3.

Dubé, Jean, Thériault, Marius, & Des Rosiers, François. (2013). “Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009.” Transportation Research Part A: Policy and Practice, 54, 49-66. doi: https://doi.org/10.1016/j.tra.2013.07.015

Dubé, Jean, Legros, Diègo, Thériault, Marius, & Des Rosiers, François. (2014a). “A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices.” Transportation Research Part B: Methodological, 64, 24-40. doi: https://doi.org/10.1016/j.trb.2014.02.007

Dubé, Jean, & Legros, Diègo. (2014b). “Spatial econometrics and the hedonic pricing model: what about the temporal dimension?” Journal of Property Research, 31(4), 333-359. doi: 10.1080/09599916.2014.913655

Dubé, Jean, Legros, Diègo, Thériault, Marius, & Des Rosiers, François. (2017). “Measuring and Interpreting Urban Externalities in Real-Estate Data: A Spatio-Temporal Difference-in-Differences (STDID) Estimator.” Buildings, 7(2). doi: 10.3390/buildings7020051

Edwards, Mary E. (2007). *Regional and Urban Economics and Economic Development: Theory and Methods*. Auerback Publications, Taylor & Francis Group.

Fatal, Yehushua S., and Walter Thurman (2014). “Response of Corn Acreage to Ethanol Plant Siting.” Journal of Agricultural and Applied Economics, 26(2), 157-171.

Feser, Edward. (2013). “Isserman’s Impact: Quasi-experimental Comparison Group Designs in Regional Research.” International Regional Science Review, 36(1), 44-68.

Greenstone, Michael, & Enrico Moretti. (2004). “Bidding for Industrial Plants: Does Winning a ‘Million Dollar Plant’ Increase Welfare.” NBER Working Paper No. 9844. doi: 10.3386/w9844 or https://eml.berkeley.edu//~moretti/milliondollarplant.pdf

Greenstone, Michael, Richard Hornbeck, & Enrico Moretti. (2010). “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings.” Journal of Political Economy, 118(3), 536-598. doi: 10.1086/653714

Haddad, Mônica A., Gary Taylor, & Francis Owusu. (2010). “Locational Choices of the Ethanol Industry in the Midwest Corn Belt.” Economic Development Quarterly, 24(1), 74-86. doi: 10.1177/0891242409347722

Higgins, Matthew J., Levy, Daniel, & Young, Andrew T. (2006). “Growth and Convergence across the United States: Evidence from County-Level Data.” The Review of Economics and Statistics, 88(4), 671-681. doi: 10.2307/40043027

Hill, R. Carter, William Griffiths, Guay Lim (2011). *Principles of Econometrics, 4th ed*. John Wiley & Sons, Inc.

Hochman, Gal, Sexton, Steven E, & Zilberman, David D. (2008). “The economics of biofuel policy and biotechnology.” Journal of Agricultural & Food Industrial Organization, 6(2).

Holland, Paul W. (1986). “Statistics and Causal Inference,” Journal of the American Statistical Association, 81:396, 945-960, DOI: 10.1080/01621459.1986.10478354

Iacus, Stefano Maria, Gary King, and Giuseppe Porro. (2008). “Matching for Causal Inference Without Balance Checking.” Available at SSRN: https://ssrn.com/abstract=1152391 or http://dx.doi.org/10.2139/ssrn.1152391

Iacus, Stefano M., Gary King, and Giuseppe Porro. (2012). “Causal Inference Without Balance Checking: Coarsened Exact Matching.” Political Analysis, 20, 1, Pp. 1--24.

Imbens, G., & Rubin, D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139025751

Isserman, A. M., and J. D. Merrifield. 1982. ‘‘The Use of Control Groups in Evaluating Regional Economic Policy.’’ Regional Science and Urban Economics 12:43–58.

Isserman, A. M., and J. D. Merrifield. 1987. ‘‘Quasi-experimental Control Group Methods for Regional Analysis: An Application to an Energy Boomtown and Growth Pole Theory.’’ Economic Geography 63:3–19.

Johnson, Richard, Dean Wichern (2014). *Applied Multivariate Statistical Analysis, 6th ed. International*. Pearson Education Limited.

Katchova, Ani L. (2009). “The Spatial Effect of Ethanol Biorefinery Locations on Local Corn Prices.” Selected Paper prepared for presentation at the Agricultural & Applied Economics Association’s 2009 AAEA &ACCI Joint Annual Meeting, Milwaukee, WI, July 26-28,2009.

Kiel, Katherine A., & McClain, Katherine T. (1995). “The Effect of an Incinerator Siting on Housing Appreciation Rates.” Journal of Urban Economics, 37(3), 311-323. doi: http://dx.doi.org/10.1006/juec.1995.1016

King, Gary, and Richard Nielsen. (2019). “Why Propensity Scores Should Not Be Used for Matching.” Forthcoming in Political Analysis, https://gking.harvard.edu/publications/why-propensity-scores-should-not-be-used-formatching

Koizumi, Tatsuji (2014). “Biofuels and Food Security: Biofuel Impact on Food Security in Brazil, Asia and Major Producing Countries.” Springer.

Low, S. A., and A. M. Isserman. 2009. ‘‘Ethanol and the Local Economy Industry Treads, Location Factors, Economic Impacts, and Risks.’’ Economic Development Quarterly 23(1):71–88.

McNew, Kevin, and Duane Griffith. "Measuring the impact of ethanol plants on local grain prices." Review of Agricultural Economics 27(2005): 164-180.

Monchuk, Daniel C., Miranowski, John A., Hayes, Dermot J., & Babcock, Bruce A. (2007). “An Analysis of Regional Economic Growth in the U.S. Midwest.” Applied Economic Perspectives and Policy, 29(1), 17-39. doi: 10.1111/j.1467-9353.2006.00328.x

National Commission on Terrorist Attacks upon the United States. (2004). The 9/11 Commission report: final report of the National Commission on Terrorist Attacks upon the United States. Authorized ed., 1st ed. New York: Norton.

OECD (2003). “The Sources of Economic Growth in OECD Countries.” Organisation for Economic Co-operation and Development.

OECD (2009a). “How Regions Grow: Trends and Analysis.” Organisation for Economic Co-operation and Development.

OECD (2009b). “Regions Matter: Economic Recovery, Innovation and Sustainable Growth.” Organisation for Economic Co-operation and Development.

Ona, Lucia Y., Agus Hudoyo, & David Freshwater (2007). “Economic Impact of Hospital Closure on Rural Communities in Three Southern States: A Quasi-Experimental Approach.” The Journal of Regional Analysis & Policy, 37(2), 155-164.

Ona, Lucia Y. (2007). “Economic Impact of Hospital Closures in Three Southern States (Georgia, Tennessee, and Texas): A Quasi-Experimental Approach.” Dissertation, University of Kentucky.

Renewable Fuels Association (2001, February). “Ethanol Industry Outlook 2001.” Accessed March 4, 2019 online: https://ethanolrfa.org/wp-content/uploads/2017/09/2001-RFA-Industry-Outlook.pdf

Renewable Fuels Association (2002, February). “Ethanol Industry Outlook 2002.” Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/outlook_2002.pdf

Renewable Fuels Association (2003, February). “Ethanol Industry Outlook 2003.” Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/outlook_2003.pdf

Renewable Fuels Association (2004, February). “Ethanol Industry Outlook 2004.” Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/outlook_2004.pdf

Renewable Fuels Association (2005, February). “Ethanol Industry Outlook 2005.” Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/outlook_2005.pdf

Renewable Fuels Association (2006, February). "Ethanol Industry Outlook 2006." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/outlook_2006.pdf

Renewable Fuels Association (2007, February). "Ethanol Industry Outlook 2007." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/RFA_Outlook_2007.pdf

Renewable Fuels Association (2008, February). "Ethanol Industry Outlook 2008." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/RFA_Outlook_2008.pdf

Renewable Fuels Association (2009, February). "2009 Ethanol Industry Outlook." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/RFA_Outlook_2009.pdf

Renewable Fuels Association (2010, February). "2010 Ethanol Industry Outlook." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/RFAoutlook2010_fin.pdf

Renewable Fuels Association (2011, February). "2011 Ethanol Industry Outlook." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/RFA_Outlook_2011.pdf

Renewable Fuels Association (2012, January). "2012 Ethanol Industry Outlook." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/2012-Ethanol-Industry-Outlook.pdf

Renewable Fuels Association (2013, January). "2013 Ethanol Industry Outlook." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/RFA-2013-Ethanol-Industry-Outlook1.pdf

Renewable Fuels Association (2014, January). "2014 Ethanol Industry Outlook." Accessed February 8, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/2014-Ethanol-Industry-Outlook1.pdf

Renewable Fuels Association (2015, January). "2015 Ethanol Industry Outlook." Accessed February 23, 2015 online: https://ethanolrfa.org/wp-content/uploads/2015/09/c5088b8e8e6b427bb3_cwm626ws2.pdf

Renewable Fuels Association (2016, January). "2016 Ethanol Industry Outlook." Accessed March 11, 2016 online: https://ethanolrfa.org/wp-content/uploads/2016/02/RFA_2016_full_final.pdf

Renewable Fuels Association (2017, February). "2017 Ethanol Industry Outlook." Accessed March 4, 2019 online: https://ethanolrfa.org/wp-content/uploads/2017/02/Ethanol-Industry-Outlook-2017.pdf

Renewable Fuels Association (2018, February). "2018 Ethanol Industry Outlook." Accessed March 4, 2019 online: https://ethanolrfa.org/wp-content/uploads/2018/02/NECfinalOutlook.pdf

Renewable Fuels Association (2019, February). "2019 Ethanol Industry Outlook." Accessed March 4, 2019 online https://ethanolrfa.org/wp-content/uploads/2019/02/RFA2019Outlook.pdf

Rephann, Terrence & Andrew Isserman (1994). "New Highways as Economic Development Tools: An Evaluation Using Quasi-Experimental Matching Methods." Regional Science and Urban Economics 24(6), 723-751.

Thanos, Sotirios, Dubé, Jean, & Legros, Diègo. (2016). "Putting time into space: the temporal coherence of spatial applications in the housing market." Regional Science and Urban Economics, 58, 78-88. doi: https://doi.org/10.1016/j.regsciurbeco.2016.03.001

Shadish, William R., Thomas D. Cook, and Donald T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin

Stock, James H., Mark W. Watson (2015). *Introduction to Econometrics, 3rd ed. Update*. Pearson Education Inc.

Swenson, David (2007). "Understanding Biofuels Economic Impact Claims." Iowa State University. Accessed March 31, 2017 online:

Swenson, David (2008). "The Economic Impact of Ethanol Production in Iowa." Iowa State University. Accessed March 31, 2017 online:

Taylor, Gary & Lisa Elliot (2012). "The Economic Impact of the Ethanol Industry on the South Dakota Economy in 2012." Accessed October 3, 2016 online: http://www.redfieldenergy.com/wp-content/uploads/2014/06/The_Economic_Impact_of_the_Ethanol_Industry_on_the_South_Dakota_Economy.pdf

Thome, Karen E. & C.-Y. Cynthia Lin Lawell (2015). "Investment in Corn-Ethanol Plants in the Midwestern United States." Accessed October 3, 2016 online: http://www.des.ucdavis.edu/faculty/Lin/ethanol_US_paper.pdf

Urbanchuk, J. (2007). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed October 3, 2016 online: http://www.ethanolrfa.org/objects/documents/2006_ethanol_economic_contribution.pdf

Urbanchuk, J. (2008). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed October 3, 2016 online: http://www.ethanolrfa.org/objects/documents/1537/2007_ethanol_economic_contribution.pdf

Urbanchuk, J. (2009). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: http://ethanolrfa.3cdn.net/cc35069ea568cde3ba_rnsmvy7rw.pdf

Urbanchuk, J. (2010). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: http://ethanolrfa.org/wp-content/uploads/2015/09/2009_ethanol_economic_contribution.pdf

Urbanchuk, J. (2011). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: https://web.archive.org/web/20121018144839/http://ethanolrfa.org/page/-/Ethanol%20Economic%20Contribution%202010%20Final%20Revised%20010411.pdf?nocdn=1

Urbanchuk, J. (2012). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: https://web.archive.org/save/_embed/http://ethanolrfa.3cdn.net/c0db7443e48926e95f_j7m6i6zi2.pdf?nocdn=1

Urbanchuk, J. (2013). "Contribution of the ethanol industry to the economy of the United States." Paper https://web.archive.org/web/20140408094847/http://ethanolrfa.org/page/-/PDFs/2012%20Ethanol%20Economic%20Impact.pdf?nocdn=1

Urbanchuk, J. (2014). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: https://web.archive.org/web/20140312003315/http://www.ethanolrfa.org/page/-/rfa-association-site/studies/ABF_Ethanol_Economic_Impact_US_2013.pdf?nocdn=1

Urbanchuk, J. (2015). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: http://ethanolrfa.3cdn.net/94596be2e72251b795_nkm6ii26n.pdf

Urbanchuk, J. (2016). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: http://ethanolrfa.org/wp-content/uploads/2016/02/Ethanol-Economic-Impact-for-2015.pdf

Urbanchuk, J. (2017). "Contribution of the ethanol industry to the economy of the United States." Paper prepared for the Renewable Fuels Association. Accessed April 6, 2017 online: http://ethanolrfa.org/wp-content/uploads/2017/02/Ethanol-Economic-Impact-for-2016.pdf

U.S. Department of Agriculture – Economic Research Service (USDA-ERS, 2018). "Recent Trends in GE Adoption." Updated by Seth J. Wechsler on July 16, 2018. Retrieved from: https://www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-us/recent-trends-in-ge-adoption/

U.S. Environmental Protection Agency (USEPA, 1996). "Phaseout of Leaded Gasoline". Document accessed online March 9, 2019, Retrieved from: https://archive.epa.gov/epa/aboutepa/epa-takes-final-step-phaseout-leaded-gasoline.html

U.S. Environmental Protection Agency (USEPA, 2007). "State Actions Banning MTBE (Statewide)". Document accessed online April 16, 2017, Retrieved from: https://nepis.epa.gov/Exe/ZyPDF.cgi/P1004KIR.PDF?Dockey=P1004KIR.PDF

U.S. Environmental Protection Agency (USEPA, 2008). "Chapter 13 MTBE". Document accessed online March 9, 2019, Retrieved from: https://www.epa.gov/sites/production/files/2014-09/documents/chapter_13_mtbe.pdf

U.S. Environmental Protection Agency (USEPA, 2015). "Regulation History of the E15 Partial Waivers Under the Clean Air Act." Document accessed online March 11, 2019, Retrieved from: https://www.epa.gov/sites/production/files/2015-09/documents/420f15044.pdf

U.S. Environmental Protection Agency (USEPA, 2017). "Final RFS for 2017". Document accessed online March 9, 2019, Retrieved from: https://www.epa.gov/renewable-fuel-standard-program/final-renewable-fuel-standards-2017-and-biomass-based-diesel-volume

U.S. Environmental Protection Agency (USEPA, 2018). "Reformulated Gasoline". Document accessed online January 29, 2019, Retrieved from: https://www.epa.gov/gasoline-standards/reformulated-gasoline

Wooldridge, Jeffrey M. (2016). *Introductory Econometrics: A Modern Approach, 6th Edition*. Boston: Cengage Learning

VITA

SCOTT W. HALL

PROFESSIONAL EXPERIENCE

University of Kentucky, Agricultural Economics, Lexington, KY
- Research Assistant — 2011 – 2015
- Teaching Assistant — 2013 – 2014

Quant Fusion LLC, Saline, MI — Aug. 2006 – Aug. 2010

AGC Automotive Americas R&D Inc., Ypsilanti, MI — Nov. 2001 – Oct. 2004

Ford Motor Company/ Visteon, Dearborn, MI — July 1991 – Nov. 2001

University of Kentucky, Electrical Engineering, Lexington, KY
- Research Assistant — 1988 – 1991
- Teaching Assistant — 1988 – 1990

EDUCATION

University of Kentucky
- Ph.D. in Agricultural Economics (expected) — Aug. 2019
- Master of Science in Economics — May 2014
- Master of Science in Electrical Engineering — May 1991
- Bachelor of Science in Electrical Engineering — May 1988

University of Michigan
- Master of Business Administration, High Distinction — May 2006

Wayne State University
- Continuing Education (Statistics, Operations Management) — 1993-1995

RESEARCH INTERESTS

Energy Economics, Regional & Local Economic Development, Agri-Business, Environmental Economics and International Trade

PUBLICATIONS

S.W. Hall, C.R. Paul, K.B. Hardin, and A.D. Nielsen, "Prediction of Crosstalk due to Showering Arcs at Switch Contacts", Proceedings of the 1991 IEEE International Symposium on Electromagnetic Compatibility, Cherry Hill, NJ, August 1991.

WORK IN PROGRESS

"US-Brazil Bilateral Fuel Ethanol Trade" (with Michael R. Reed)

"Economic Impact of Ethanol Biorefineries in the U.S. Midwest: A Quasi-Experimental Approach" (with David Freshwater)

AWARDS AND SCHOLARSHIPS

2018-2019 Dr. John C. Redman Memorial Scholarship
2017-2018 Dr. John C. Redman Memorial Scholarship
Patents: US 5,936,585; US 5,883,599; US 5,640,167
1995 Ford Motor Company Award for Customer Satisfaction Recipient
1995 Ford ACD Quality Award for Customer Satisfaction Recipient
1987 Ford Motor Company $100,000 Research Grant Recipient

PRESENTATIONS

2019 Southern Agricultural Economics Association (Paper & Poster)
2013 Agricultural & Applied Economics Association (Case Study)