THREE ESSAYS IN REGIONAL ECONOMIC DEVELOPMENT: FORECASTING, FIRM SIZES AND ETHANOL PLANTS

By

Timothy M. Komarek

A DISSERTATION

Submitted to
Michigan State University
partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Agricultural, Food and Resource Economics

2012

ABSTRACT

THREE ESSAYS IN REGIONAL ECONOMIC DEVELOPMENT: FORECASTING, FIRM SIZES AND ETHANOL PLANTS

By

Timothy Michael Komarek

This work examines several topics related to regional economic development and public policy. It consists of three essays. The first essay examines whether spatial linkages between regions and coincident and leading employment indices based on resident's perceptions of the local economy improve forecast performance. The state of Michigan is used as the study area and separated into 6 geographic regions. I compare forecast performance using both unrestricted vector autoregressive and Bayesian vector autoregressive models. The results show significant improvement in forecast accuracy for each forecasting technique by adding leading and coincident indicators of economic activity. The results also suggest that spatial linkages for Michigan regions aid in predicting future employment levels in 8-quarter ahead forecasts. Furthermore, I compare the forecast performance of the region-specific indices based on surveys of residents in each of the 6 Michigan regions with analogous indices based on a nationally representative survey. I find that the region-specific indices outperform the national indices, while the national indices improve forecast performance over each of the models without indices.

The second essay explores the effect of the business size distribution on per-capita income and employment growth. I estimate a growth model with U.S. county data from 1990-2000. The business size distribution is measured in two ways. First, the distribution is measured as the share of employees across nine establishment size categories that range from micro firms (1-4 employees) to large firms (1000+ employees). Second, I use several indices that include an index similar to a Gini coefficient and the Atkinson index of

inequality. The results show that business size distribution has a significant impact on county level growth patterns. Furthermore, the employment shares in small firms increase employment growth, but decreases per-capita income growth. The results have implications for national economic growth policies, such that emphasizing entrepreneurship and small firms is well suited in times of high unemployment, while in times of stable employment growth shifting policies toward large firms may spur income growth.

The third essay examines the effect of ethanol production facilities on the local labor market. Few studies examine this question with historical data, largely due to the data constraints related to local ethanol production. Using a difference-in-difference identification strategy, I use a data set containing the timing of ethanol plant construction and production start dates in 12 states from 1990-2011 to estimate the net employment effect. Furthermore, I add leads and lags to the start of ethanol production to examine the dynamic response of an ethanol plant on a local economy. When using non-urban, high corn counties, the results suggest a positive and statistically significant employment multiplier, with the overall average local employment impact of approximately 125-200 jobs. Conversely, the analysis shows little evidence of a positive economic effect resulting from the construction of an ethanol plant. The dynamic estimates suggest that the local employment multiplier grows over the first several years of ethanol production, and yields a long-run local employment impact of approximately 275 jobs per production facility.

Dei gratia, ad maiorem Dei gloriam

ACKNOWLEDGEMENTS

I would like to thank my dissertation committee, namely Scott Loveridge, Mark Skidmore, Soren Anderson, Charles Ballard and Steve Miller for their guidance and support throughout my dissertation research. This work would not have been possible without their contributions. In particular, I am deeply indebted to my major professor Scott Loveridge for his guidance, mentoring, support, and friendship throughout my doctoral studies. I hope to become someone that my committee members are proud to call a colleague and a friend.

I would like to thank the Center for Economic Analysis, the Glenn and Sandy Johnson Dissertation Enhancement Fellowship, and the College of Agriculture and Natural Recourses Dissertation Competition Fellowship for financial support of my dissertation research.

I would like to thank all of the graduate students and faculty that I have worked with at Michigan State, and who have been so helpful and supportive. A short list could not possibly due justice to all of the people at MSU that have enriched my life. I am particularly grateful to my roommates, Nate Brandsen, Eric VanderVeen, Mark Medenblik, and Jeff Uitvlugt that I have had at various times at MSU. Your friendship over the years means the world to me.

Finally, I would like to thank all of the people closest to my heart, my family, and friends that I consider family. Your love and support are a testament to grace and have shaped the person I am today and the person I strive to become.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Forecasting Regional Employment Using a BVAR With Leading and Coincident Indicators: The Case of Michigan

1.1 Introduction

Policy makers and businesses often desire economic forecasts that are conducted at local geographic levels to inform their decision-making (Miller, 1998). Nonetheless, forecasts are most often conducted at the national level (Engemann et al., 2008). National forecasts typically focus on aggregate economic metrics, such as gross domestic product. They also often take advantage of extensive time-series data, and are informed by broad leading and coincident indicators of economic activity (Lahiri and Moore, 1993). Modeling and forecasting the aggregate economy can be important to understanding forthcoming local economic activity. However, Dua and Miller (1996a) note that aggregate forecasts are only useful to local officials and firms if regional economic activity mimics the larger national economy. Sub-national regions often differ from the nation in terms of the proportion of economic activity in a given sector. Moreover, some sectors are much more cyclical than others (Petersen and Strongin, 1996; Tan and Mathews, 2010). It follows that state or region business cycles could differ substantially from the national business cycle (Guha and Banerji, 1998/1999). Also, state policies, specific management strategies by major employers, or the structure of the industry may result in differing growth patterns across regions, even when the industry mix is broadly similar.

Despite the traditional focus on national forecasts, state- and region-specific forecasting models are playing an increasingly important role. Local forecasts can assist state and local governments and small businesses in formulating policy and designing initiatives. However, they have been considerably less tested in the literature than their national counterparts (Engemann et al., 2008). To increase forecast performance, regional models have started using information from spatial linkages (Rey 1998, 2000). In particular, several studies have shown that employment forecasts can be improved by incorporating spatial information (LeSage and Krivelyova, 1999; Hernandez-Murillo and Owyang, 2006; Rickman et al., 2009).

Another technique for increasing forecast performance is including leading and coincident indicators of economic activity. However, relatively few studies have extended the notion of leading indicators used in national forecasts[1] to the regional level. Forecasting research for the state of Connecticut provides several exceptions. Specifically, Dua and Miller (1996a) develop a state-level index of leading indicators that uses the average manufacturing workweek and measures of unemployment to forecast a coincident index of economic activity, employment, and the unemployment rate. Banerji et al. (2006) revise the leading economic indicator index developed by Dua and Miller (1996a) and test its out-of-sample forecast performance. Both studies show increased forecast performance when including leading economic indicators. Similarly, leading indicators were also found to improve the accuracy of forecasts for home sales for both Connecticut (Dua and Miller, 1996b) and the U.S. (Dua et al., 1999).

The objective of this study is to examine the accuracy of competing techniques for forecasting regional employment, using regions in the state of Michigan as a case study. The state of Michigan is separated into 6 geographic regions, comprised of groups of counties. I focus on two forecasting techniques. First, I examine whether leading and coincident indices of economic activity increase local forecasting performance. I develop two regional economic indices based on resident's perceptions of the local economy. The survey questions and methodology used to create the local indices are analogous to the nationally representative

[1] See the United States Department of Commerce for examples of pertinent leading indicators, and Koch and Rasche (1988) for an examination of their approach.

Thomson Reuters/University of Michigan Surveys of Consumers[2] (SOC). Furthermore, I examine the relative merits of the local and national indices. Second, I examine whether spatial linkages connecting the 6 Michigan regions improve forecast performance. Spatial linkages for the 6 regions are taken into account by imbedding a spatial contiguity matrix into a Bayesian vector autoregressive framework, akin to LeSage and Pan (1995).

The three types of models tested are an unrestricted vector autoregressive model (UVAR), Bayesian vector autoregressive model with Minnesota prior specification (MN BVAR), and Bayesian vector autoregressive model with an informative prior based on the spatial linkages of the 6 regions (CONT BVAR). I start by examining each model with and without region-specific leading and coincident indices, resulting in six competing forecasting techniques. This is followed by estimates of each model using the national indices from the SOC. To assess the performance of the competing models, a scoring procedure is used, based on the mean absolute percent error of out-of-sample performance. The scoring procedure calculates the number of wins against each competing technique for 6- and 8-quarter out-of-sample forecasts.

This research contributes to the forecasting literature in several ways. First, I use a regional parallel to the Consumer Sentiment Indices produced by the Thomson Reuters/University of Michigan Surveys of Consumers (SOC). The SOC has a long history for predicting future economic activity in the macroeconomic literature (e.g. Howrey, 2001). Yet, to the best of my knowledge, residents' perceptions concerning current and future economic conditions have not been examined in a regional context. I also compare the forecasting performance when using the national, SOC indices, to the region-specific leading and coincident indices. Second, I incorporate and test the effectiveness of spatial linkages for regions within a

[2] See http://www.sca.isr.umich.edu/ for more details on the Thomson Reuters/University of Michigan Surveys of Consumers.

state in a Bayesian vector autoregressive framework. Spatial linkages are introduced into the model through the variances of the Bayesian priors. Examining regions within a state contrasts with several studies that use states (Magura, 1997; LeSage and Krivelyova, 1999; Engemann et al., 2008) and metropolitan areas (Rickman et al., 2009). Third, this study uses a relatively short time series, with quarterly data from 2001q1 to 2010q4. A major constraint to regional forecasting is the scarcity of usable regional economic data. Many variables at the regional level do not have lengthy histories, high frequencies, or high levels of industry detail. Thus, this research contributes to the forecasting literature by examining whether techniques, such as leading and coincident indicators, are useful where estimates are constrained by data.

I find that adding leading and coincident indicators of economic activity leads to improved forecast accuracy for each forecasting technique. Furthermore, the results show that the region-specific indices are more effective in forecasting local employment than their national counterparts. The results also suggest that spatial linkages among Michigan regions aid in predicting future employment levels over MN BVAR in 8-quarter ahead forecasts. The next section outlines the two data sets that are used in the analysis and the creation of the region-specific indices. The competing forecasting models are described in section 1.3. In section 1.4, I describe the forecasting experiments and implementation, as well as statistics for evaluating forecast accuracy. The results for the competing forecasting models follow in section 1.5, while the final section provides a summary and conclusion.

1.2 Data

Data on each region's employment level as well as perceptions of the economy are used to conduct the forecasting model proposed above. The data used in this forecasting experiment come from three sources, and are described in detail below.

1.2.1 Employment data

First, employment data at the county level come from the Quarterly Census of Employment and Wages (QCEW) collected by the Bureau of Labor Statistics. The QCEW collects employment data at industry classification levels down to the 6-digit NAICS level, covering 98% of U.S. jobs for each county in the United States (QCEW, 2006). I add the total private employment levels across counties in each of the six regions within the state. The regions have previously been designated by Michigan State University Extension and are shown in figure 1.1.

Normalized levels of seasonally adjusted employment from 2001q1 are plotted for each region in figure 1.2. The employment level was seasonally adjusted using the U.S. Census Bureau's X-12 Seasonal Adjustment Program (U.S. Census Bureau, 2011). Figure 1.2 depicts a downward trend for each Michigan region. According the National Bureau of Economic Research, the United States experienced two recessions over the sample time period (NBER, 2012). The first recession lasted 8 months from March 2001 to November 2001. The second recession occurred between December 2007 and June 2009. The recession of the late 2000s, which has come to be known as "The Great Recession," dramatically weakened the U.S. labor market (Farber, 2011). Figure 1.2 shows the consequences of the recession of the late 2000s on the level of employment in Michigan regions. The recession of the late 2000s negatively affected all six Michigan regions. Figure 1.2 shows that it was particularly hard on South East Lower Michigan, which is home to the city of Detroit.

Summary statistics for each region and the state total are presented in Table 1.1. Excluding the Northern Lower Peninsula, the regions share 2001q1 as the time period with the highest employment level. Similarly, the lowest employment for all six regions falls at the end of the recession of the late 2000s. It is particularly noteworthy that the average quarterly growth rate for all six regions is negative over the sample period. The negative quarterly growth rate, shown in table 1.1 for each region, confirms the negative trend apparent in total private employment in figure 1.2. To prevent spurious regression in the forecasting models from the state-wide negative trend, the employment data were de-trended for each region.

1.2.2 Survey data

The following data sources were used to develop indicators of attitudes regarding economic activity. First, I use data collected by the Thomson Reuters/University of Michigan Surveys of Consumers (SOC). Since 1978 the SOC has conducted a nationally representative survey each month using a minimum of 500 telephone interviews. Three indices, the Index of Consumer Sentiment, the Index of Consumer Expectations and the Index of Current Economic Conditions, are created from five SOC questions. The Index of Consumer Expectations is included in the Leading Indicator Composite Index published by the U.S. Department of Commerce.

Second, survey data collected by the Institute of Public Policy and Social Research for the State of the State Survey (SOSS)[3] was used to create two local indices of economic activity. SOSS was initiated in 1994, and involves quarterly interviews with approximately 1,000

---

[3] These data were collected by the Office for Survey Research of the Institute for Public Policy and Social Research (IPPSR, various years) at Michigan State University. While IPPSR accepts responsibility for the quality of the data, the interpretations and conclusions presented are solely those of the author. More detail on the data and sampling procedures can be found in the full report of the data collection methods, available at http://www.ippsr.msu.edu/SOSS.

Michigan residents. The survey collects information on each respondent's county and ZIP code. Information on the geographic location of respondents was used to allocate their responses to one of the regions depicted in figure 1.1. The SOSS is given in fairly regular intervals (Spring, Summer, Winter, Fall) for the sample period of 2001 to 2010. However, the actual interview dates vary within each year. Table 1.A1 in the Appendix shows how the time period (quarter and year) are matched with the SOSS wave numbers. This table also shows the interview dates and sample sizes in each wave. The SOSS consistently asks several questions on key economic conditions in the respondent's household and local area in its core questionnaire. The core economic questionnaire in the SOSS mimics several key questions used in the Table 1.2 shows a comparison of the economic questions from both the SOC and SOSS.

The core economic questions from both the SOC and SOSS are quite similar in phrasing. Nonetheless, several differences are noteworthy. First, the SOSS asks residents about future business conditions in their community, while the SOC asks about business conditions in the entire U.S. This is beneficial to the current study, which focuses on forecasting regions within a state rather than the whole U.S. Both surveys ask about the future unemployment situation for the entire U.S. However, the SOSS maintains a focus on the near-term (12 month) unemployment situation, while the SOC turns to the long-term labor market (5 years). The SOSS allows for several additional response options to characterize each household's present financial situation than the SOC. Finally, the SOC includes a question on large consumer purchases (not included in the SOSS), which is used in the Index of Current Economic Conditions.

1.2.3 Leading and coincident indices

Following the SOC methodology[4] I calculate two indices based on the SOSS questions in Table 1.2. The first index, the Index of Michigan Current Economic Conditions (IMCEC), uses the present financial situation (EI4) SOSS question. The IMCEC is similar to the SOC's Index of Current Economic Conditions. The Index of Current Economic Conditions also includes question EI5 on buying major household items, which is not available in the SOSS. The Index of Michigan Resident Economic Expectations (IMREE) uses SOSS questions EI1, E12, and EI3. The IMREE is analogous to the SOC Index of Consumer Expectations.

I use the same procedure to calculate both the IMCEC and IMREE. First, the relative scores (the percent giving a favorable reply minus the percent giving an unfavorable reply, plus 100) were calculated for the 4 SOSS questions in each region. For question EI4 "not so good" and "poor" were coded as unfavorable while "excellent" and "good" were considered favorable. The IMCEC was calculated using the following formula:

$$IMCEC_{i,t} = \frac{EI4_{i,t}}{EI4_{i,2001q1}} \tag{1.1}$$

$EI4_{i,t}$ is the percent answering favorably minus the percent answering unfavorably, plus 100, for question *EI4* in region *i* at time *t*. $EI4_{i,2001q1}$ denotes the base-period total at time 2001q1 for region *i*. Thus, the IMCEC is equal to 100 for the first quarter of 2001. The IMREE was calculated in a similar fashion using equation (1.2):

$$IMREE_{i,t} = \frac{EI1_{i,t} + EI2_{i,t} + EI3_{i,t}}{EI1_{i,2001q1} + EI2_{i,2001q1} + EI3_{i,2001q1}} \tag{1.2}$$

Using equation (1.1) and (1.2), two indices were created with a base time period of 2001q1 for each of the six regions in the state of Michigan.

[4] See http://www.sca.isr.umich.edu/ for more details on index calculations

Figure 1.3 shows a graph of the Michigan state seasonally adjusted employment level, and IMCEC, while figure 1.4 displays the employment level and the IMREE. The left vertical axis depics the state employment level, while the right axis shows the index of interest. In general, the graphs show a correlation between the indicator variables of interest and the state employment level. However, the visual depictions of the relationship between the employment level and indices do not provide precise evidence of whether each index leads, lags, or is coincident with the employment level.

Intuitively I expect the IMCEC, which uses each resident's current household financial situation, to serve as a coincident economic index. Similarly, since each question in the IMREE pertains to a year from the interview date, it is thought *a priori* to lead by 4 quarters. To determine objectively the lead or lag structure of each index, J-tests[5] were performed to distinguish among competing models (Davidson and Mackinnon, 1981). It is possible the results of the J-tests for each of the 6 regions could give contradictory recommendations on the lead and lag structure of the indices. To remain consistent across regions, I examine indices and employment at the state total. The results suggest that IMCEC is a coincident index of employment and IMREE leads employment by 4 quarters.[6]

---

[5] J-tests provide one method of choosing among non-nested models. Non-nested models occur when neither model can be expressed as a restricted version of the other. The intuition behind the J-test is that if one model is "correct," then the fitted values of competing models should not have explanatory power when included in the "correct" model. Each J-test is performed by regressing the seasonally adjusted Michigan total private employment level on an index lead/lag and the fitted values of a competing lead/lag structure. For example, statistically insignificant fitted values of the competing lead/lag structure suggest we can fail to reject (not reject) the "correct" model.

[6] See Appendix table A2 and table A3 for results on the J-tests of competing lead and lag structures for each index. Each table displays P-values for the coefficient on the <u>fitted values</u> of each competing lead/lag of the index examined.

1.3 Forecasting models

The models tested in the forecasting experiment are an unrestricted vector autoregressive model (UVAR), a Bayesian vector autoregressive model with a Minnesota prior specification (MN BVAR) and a Bayesian vector autoregressive model with an informative prior based on spatial contiguity of the 6 regions (CONT BVAR). Furthermore, I incorporate leading and coincident indices of economic activity in each of the aforementioned models as exogenous regressors. The forecasting experiment examines the performance of a total of 6 different model techniques.

1.3.1 Unrestricted vector autoregression model (UVAR)

The unrestricted VAR (UVAR) is a relatively low cost technique, in comparison to its structural simultaneous equations counterpart, which is used extensively in forecasting. A UVAR is an 'atheoretic' approach that uses historic data for all of the variables to forecast future values. Despite its 'atheoretic' foundation it has been suggested that UVAR models approximate the reduced form of a structural simultaneous equations model (Zellner, 1979; Zellner and Palm, 1974). The UVAR model first proposed by Sims (1980) is:

$$y_t = C + A(L)y_t + \varepsilon_t \qquad (1.3)$$

where, y_t is an (n x 1) vector of endogenous explanatory variables to be forecast, $A(L)$ is an (n x n) polynomial matrix in lag operator L with lag length p such that $A_1L + A_2L^2 + . . . + A_pL^p$, C is a (n x 1) vector of constant terms, and ε_t is the error term with the usual normal distribution.

This model is estimated with OLS, with the same number of lags for each variable. A common drawback for UVAR models is they can quickly become over parameterized, especially with relatively short time series. For example, lag length *p* results in an equation that has ($n \times p$) $+1$ coefficients to be estimated.

1.3.2 Minnesota prior bayesian vector autoregression model (MN BVAR)

Bayesian techniques differ from the classical regression used in UVARs when estimating coefficients. Bayesian estimation combines time-series data with prior expectations of each parameter. The Bayesian vector autoregression model (BVAR) was first developed by Litterman (1980) in an attempt to circumvent the over parameterization problem with VAR models. One technique for reducing the number of parameters in UVAR models is to eliminate lagged explanatory variables that are statistically insignificant from the model. In contrast, Bayesian methods impose prior restrictions on variables, but allow the data to override the prior assumption. Specifically, Dua and Miller (1996a) emphasize that adding a Bayesian prior restriction on a parameter also effectively increases the number of observations by one. This mitigates the degrees-of-freedom issues that are inherent in adding a parameter to a VAR model. Sims et al. (1990) note that Bayesian models do not need to account for nonstationarity. The BVAR uses the mixed estimation procedure for each equation developed by Theil and Goldberger (1961):

$$B = \left(X'X + \sigma_u^2 R'\psi R \right)^{-1} \left(X'Y + \sigma_u^2 R'\psi^{-1} r \right) \tag{1.4}$$

In Theil and Goldberger’s (1961) mixed estimator, **Y** is a vector of own region private employment levels, and **X** is a matrix of lagged own- and cross-equation private employment levels. I keep the notation consistent, by continuing to use *n* to represent the number of equations and *p* the number of lags. The distributions of the Bayesian prior enter the estimate with **r** as a ($n \times p$) $\times 1$ vector of prior means and ψ as a ($n \times m$) $\times$ ($n \times m$) matrix of prior variances. Finally, **R** is a (n x m) x (n x m) identity matrix and σ_u^2 is the estimated variance from the unrestricted VAR. The coefficient estimates then intuitively are weighted between the actual past data values and the information from the Bayesian prior distribution.[7]

Since the usefulness of the BVAR model comes from the researcher's ability to set prior means and variances on parameters, it is helpful to illustrate the role of the prior distribution. Following the discussion of Rickman et al. (2009), I explain the use of hyper parameters of the variance of the prior distribution $\lambda^2(i,j,p)$ for variable j, equation i, and lag length p such that:

$$\lambda^2(i,j,p) = \left[\theta f(i,j) g(p) \left({S_i}/{S_j} \right)\right]^2 \tag{1.5}$$

Within this framework, θ represents the overall tightness of the prior distribution, $f(i,j)$ is the tightness of the parameter for the jth variable in the ith equation, and $g(p)$ is the tightness around the pth lag. The ratio $\left({S_i}/{S_j} \right)$ is a scalar that corrects for differences in the magnitudes between variables i and j by using the standard errors in the respective AR(1) models.

The BVAR forecasting procedure allows the researcher to specify the prior means (**r**) and variances (ψ) for the estimated coefficients. The Minnesota prior attributed to Litterman (1980) is one of the first specifications for the distribution of the Bayesian prior. The prior mean for each variable's own first lag was specified as unity and the lags of all the other variables were assumed to be zero. This specification is consistent with the notion that the times series follows an AR(1) random walk with drift (Todd, 1984). Doan et al. (1984) extended this specification by

---

[7] See Brikes and Dodge (1993) for a more detailed discussion of the weighting arrangement between the prior distribution and the historical data.

including specifications for the hyper parameters as well, where $\theta = .1$, $f(i,i) = 1$, and $f(i,j) = .5$. This suggests that other regions receive half of the weight relative to the equation's own lagged private employment.

1.3.3 Spatial contiguity prior bayesian vector autoregression model (CONT BVAR)

We specify the spatial employment spillovers based on the contiguity of regions in geographic space. The spatial relationships are incorporated into the basic Bayesian vector autoregressive model through the specification of the prior variance matrix. Where a MN BVAR assumes that the underlying structure of each region follows an AR(1) process, in contrast the CONT BVAR assumes contiguous regional neighbors influence regional employment levels. Thus, the geographic relationships, shown in figure 1.1, are directly translated to the spatial contiguity weight matrix of the prior means and variances found in Table 1.3. Specifically, the main diagonal of the matrix remains $f(i,i) = 1$, constituting the same AR(1) process as the MN BVAR. However, the off-diagonal elements (the prior co-variances) $f(i,j)$ are dictated by the spatial relationships.

1.3.4 Exogenous regressors

Exogenous explanatory variables are added to each of the models in a straightforward way. Each equation, which represents total private employment for one of the six regions, is matched with the IMCEC and IMREE for its own region. For example, in the UVAR for the first region, the explanatory variables are its own lagged employment levels, the lagged employment levels of the 5 other regions, and the IMCEC and IMREE for the first region. Within the Bayesian estimation procedure, the exogenous leading and coincident indicators are specified with a diffuse prior, such that identification of each coefficient comes solely from the observed data.

1.4 Forecast experiment and implementation

I run a competition among the competing forecasting models to assess the relative forecast performance of each model, using quarterly data from the first quarter of 2001 through the fourth quarter of 2010. Testing each of the models in all 6 regions provides robust evidence for the consistent performance of one model above the others. The optimal Lag lengths were determined by a system-wide likelihood ratio test (Enders, 1995). The test suggested that the optimal lag length was 2 quarters, which is used in each equation for all regions.

The forecast accuracy of each model is tested against the observed out-of-sample values using 6 and 8 quarters rolling forecasts. We measure the out-of-sample forecast accuracy for the competing models with the mean absolute percentage error (MAPE), which is defined as:

$$MAPE = \frac{1}{T}\sum\frac{A_{t+f} - F_{t+f}}{A_{t+f}} \tag{1.6}$$

where A_{t+f} is the actual data for the time period ($t + f$), t is the last time period used by the forecasting model and f is the number of periods forecasted ahead. Similarly, F_{t+f} is the forecasted values made at time t for period(s) ($t + f$).

Rolling window out-of-sample forecasts are produced for both 6-quarters and 8-quarters ahead. In a rolling window forecasting scheme, the window of observed data is fixed in size. For each new period forecasted, one new observation is added and one observation from the beginning of the sample is dropped. For every ‘roll’ ahead the model is re-estimated. For the first forecast, the last historical data in the sample are from the 4th quarter of 2007. This procedure yields seven 6-quarter forecasts and four 8-quarter forecasts.

1.5 Results

First, I start by examining the various forecasting techniques both with and without region-specific (based on the SOSS) leading and coincident indices. This is followed by a comparison of the forecast performance of the region-specific indices (SOSS) and the national indices (SOC). Mean absolute percent errors (MAPEs) for private sector employment are shown for 6-quarter ahead forecasts in table 1.4, and in table 1.5 for 8-quarter ahead forecasts. There are several broadly consistent features from the forecasting experiment. First, including leading and coincident indicators benefits forecasting performance. The three models that utilized the IMREE and IMCEC had the three lowest average MAPE's across the six regions in 8-quarter ahead forecasts. The IMREE and IMCEC models had a lower average MAPE than their non-leading and coincident indicator counterparts for the 6-quarter forecast ahead forecast. Second, forecasting models that use Bayesian priors generally outperform their UVAR counterparts.

The best-performing model for the 6-quarter ahead forecast, measured by the average MAPE across the six regions, is the MN BVAR with the leading and coincident indices. The performance of the CONT BVAR using the indices follows closely behind. Conversely, for the 8-quarter ahead forecast, the UVAR with leading and coincident indices had the lowest MAPE averaged across the six regions. However, the MN BVAR and CONT BVAR models using the indices performed only slightly worse than the UVAR in the 8-quarter forecasts. Finally, none of the forecasting techniques had dramatically inaccurate forecasts for any of the individual regions. The most accurately forecasted region for all techniques 6-quarter ahead was the Northern LP, while the West Central region performed the best in the 8-quarter forecasts.

To examine the contribution of the Bayesian forecasting techniques, the models are compared to a UVAR model of the same lag length. The results suggest that imposing prior information on the Bayesian means and variances is beneficial to model performance. The increased forecast performance of the Bayesian estimators is due in part to their shrinkage-like properties (Birkes and Dodge, 1993; Vinod, 1978). The shrinkage-like properties of Bayesian estimators (e.g. MN BVAR) come from specifying same- and cross-lag priors as zero, thus shrinking these parameters toward zero. Table 1.4 shows that in the 6-quarter ahead forecast without indices, the average MAPE drops by approximately 0.6 when using a Bayesian prior. There is a similar decline in the average MAPE when including the leading and coincident indices to the 6-quarter forecast. For the models without indices the average MAPE decline was twice as large for 8-quarter forecasts as it was for 6-quarter forecasts. The 8-quarter ahead forecasts using indices are the only models that deviate from this pattern. The two models with Bayesian prior (MN BVAR and CONT BVAR) yield very similar average MAPEs.

Based on the MAPEs shown in Tables 1.4 and 1.5, pairwise forecast model comparisions are calculated and shown in Tables 1.6 and 1.7. Table 1.6 shows the pairwise comparison of each model for the 6-quarter ahead forecast, and similarly Table 1.7 for 8-quarter ahead forecast. In the pairwise comparisons each element represents the number of wins the row model has over the corresponding column model across the six regions. Thus, there is a maximum of 6 wins among every pair of models. The maximum total number of wins for each model is 30. There is a symmetry inherent in each table, where the element in row *i*, column *j* is the additive inverse of the element in row *j*, column *i*. Using the pairwise comparison method allows us to consider the forecasting results across the 6 regions and avoids the influence of any outliers on the average MAPEs.

Table 1.6 shows that the MN BVAR and CONT BVAR using leading and coincident indices have the most head-to-head wins against the other models for the 6-quarter ahead forecasts. In comparison, the UVAR model using indices has one more head-to-head matchup win than the CONT BVAR for the 8-quarter ahead forecasts. We also see further evidence of the dominance of the leading and coincident indices in each of the pairwise-comparison tables. The win counts in the lower left corners of each table show the number of wins for models with leading indicators over their non-leading indicator counterparts. The models with indices win between 4 – 6 matchups against each of the non-index models. The pairwise comparisons reiterate the poor performance of the UVAR model. The UVAR only wins in 2 regions against the competing models. Interestingly, the UVAR using the leading and coincident indices performs quite well, winning 23 of its 8-quarter matchups. This suggests that the public's perceptions of economic activity add considerable information for the atheoretic UVAR models. Finally, the CONT BVAR model performs well in the head-to-head matchups. This occurs despite the fact that the spatial linkages in the BVAR are artifacts of the clustering of counties into geographic regions. For example, the Bayesian technique that incorporates spatial linkages sweeps the UVAR and BVAR when the leading and coincident indicators are not included in the 8-quarter forecasts. This complements LeSage and Krivelyova (1999) and Rickman et al. (2009), who show that spatial priors outperform non-informative priors.

The preceding results show that the region-specific indices from the SOSS increase forecast performance in comparison to the models without the leading and coincident indices of economic activity. Next, I move on to comparing the forecast performance of the region-specific indices constructed from the SOSS with their national counterparts from the SOC. As noted in section 1.2.2, the SOSS and the SOC share many similarities. However, they differ in their population of interest (U.S. versus Michigan). The data and indices are publically available for both the SOSS and the SOC. Nonetheless, the nationally representative SOC could be more useful to other regions of the country than the Michigan specific indices constructed from the SOSS. Therefore, it is useful to examine if there is a similar forecasting benefit from the national indices. The leading and coincident indices from the SOC (i.e. the Index of Current Economic Conditions and the Index of Consumer Expectations) are incorporated in the forecasting models in the same was as the IMCEC and IMREE.

The MAPEs for 6- and 8-quarter ahead forecasts that use the national leading and coincident indices from the SOC are shown in table 1.8. The results share some broad similarities with the previous results. The forecasting models that use Bayesian priors outperform the UVAR models. The Bayesian prior models have lower MAPEs in each of the 6 regions than the UVAR for both 6- and 8-quarter ahead forecasts. On average, the 8-quarter ahead model also out performs the 6-quarter ahead forecasting model.

Furthermore, a striking result emerges when comparing the average MAPEs for models without indices, with the regional-specific indices, and the national indices (the final column in tables 1.4,1.5, and 1.8). The models that use the regional-specific indices have the lowest average MAPEs for each of the three forecasting techniques (UVAR, MN BVAR and CONT BVAR). The models that use the national indices from the SOC have slightly higher average MAPEs than the aforementioned SOSS indices. However, they outperform the models without indices. Thus, when measuring forecast performance by the average MAPE across the six regions, the results show that leading and coincident indices of economic activity (national or region-specific) aid employment forecasting. While the region-specific indices outperform their national counterparts, the magnitudes of this difference appear small. These results suggest it is beneficial for regions to use the nationally representative leading and coincident indicators from the SOC in the absence of region-specific information on the local economy.

1.6 Summary and conclusion

This research assesses the performance of including leading and coincident indices and spatial information in several forecasting models. The state of Michigan is broken into six regions and used as the study area. The forecasting models examined are an unrestricted vector autoregressive model (UVAR), a Bayesian vector autoregressive model with Minnesota prior specification (MN BVAR) and Bayesian vector autoregressive model with an informative prior based on spatial contiguity of the 6 regions (CONT BVAR). This research uniquely uses sub-state employment data from the QCEW, the nationally representative Thomson Reuters/University of Michigan Surveys of Consumers a national survey, and a survey of Michigan residents.

The results support the view that including leading and coincident indicators improves forecast performance. For example, the marginal contribution of including the region-specific indices is higher for 8-quarter ahead forecasts than for 6-quarter ahead forecasts. This is particularly apparent when adding IMREE and IMCEC to UVAR models. Adding the indices marginally improved forecast performance for the 6-quarter ahead forecast, but significantly increases performance for the 8-quarter ahead forecast. This may suggest that the public's perceptions of the local economy are more correlated with longer-run outcomes rather than shorter-run.

In the short-term 6-quarter forecasts, the BVARs marginally increase forecast performance over the UVAR, but substantially improve performance in the 8-quarter ahead forecasts. This suggests that the shrinkage estimators contribute more to long-term forecasts. Spatial linkages offer marginal improvement over MN BVAR in 8-quarter ahead forecasts. This suggests that the shrinkage properties of BVARs utilizing spatial information outweigh the MN BVAR, and that this becomes more important in longer-term forecasts.

Table 1.1. State and Region Summary Statistics

	State Total	Upper Peninsula	Northern Lower Peninsula	West Central Lower Peninsula	East Central Lower Peninsula	South West Lower Peninsula	South East Lower Peninsula
Level (thousands)							
Mean	3449.37	85.42	144.09	560.23	235.02	455.15	1969.56
Variance	47066.28	8.29	54.39	612.18	95.62	467.19	22051.05
Maximum	3752.10	88.48	150.67	595.27	251.48	484.24	2192.61
Date	2001Q2	2001Q1	2004Q4	2001Q1	2001Q1	2001Q1	2001Q1
Minimum	2964.98	79.09	128.28	506.31	215.69	410.44	1689.45
Date	2010Q1	2009Q4	2009Q4	2009Q3	2009Q3	2009Q3	2009Q4
Growth Rate (percent)							
Mean	-0.43	-0.26	-0.36	-0.34	-0.36	-0.36	-0.61
Variance	7.82	0.71	0.90	0.95	0.51	0.68	0.97
Maximum	4.02	1.47	1.52	1.05	0.74	0.81	0.96
Date	2010Q2	2010Q2	2004Q1	2010Q2	2010Q2	2010Q2	2010Q2
Minimum	-8.54	-3.48	-3.62	-4.49	-3.15	-4.00	-4.93
Date	2009Q1	2009Q1	2009Q1	2009Q1	2009Q1	2009Q1	2009Q1

Table 1.2. Comparison of Questions from the Thomson Reuters/University of Michigan Surveys of Consumers (SOC) and the State of the State Survey (SOSS)

Economic Indicator	SOC	SOSS
Future Business Conditions (EI1)	Now turning to business conditions in the country as a whole--do you think that during the next twelve months we'll have good times financially, or bad times, or what?	Now turning to business conditions in your community, do you think that during the next twelve months your community will have good times financially, or bad times, or what?
Future Unemployment (EI2)	Looking ahead, which would you say is more likely--that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?	Twelve months from now, do you expect the unemployment situation in this country to be better than, worse than, or about the same as it was in the last 12 months?
Future Financial Situation (EI3)	Now looking ahead--do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?"	Now looking ahead, do you think that a year from now, you and your family living there will be better off financially or worse off or just about the same as now?
Present Financial Situation (EI4)	We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?	How would you rate your household's overall financial situation these days? Would you say it is excellent, good, just fair, not so good, or poor?
Future Purchases (EI5)	About the big things people buy for their homes--such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?	Not Available

* Underling for emphasis added by SOC survey publication and by the author for the SOSS

Table 1.3. Spatial Contiguity Based Weight Matrix

	UP	Northern LP	West Central	East Central	South West	South East
UP	1	0.5	0.5	0.5	0.5	0.5
Northern LP	0.5	1	1	1	0.5	0.5
West Central	0.5	1	1	1	1	0.5
East Central	0.5	1	1	1	1	1
South West	0.5	0.5	1	1	1	1
South East	0.5	0.5	0.5	1	1	1

Table 1.4. Aggregate Private Sector Employment Forecast MAPE with Region Specific Indices: 6 quarter

	UP	Northern LP	West Central	East Central	South West	South East	Average
UVAR	1.306	1.871	2.027	1.920	1.652	2.267	1.841
MN BVAR	1.337	0.965	1.497	1.331	1.308	1.043	1.247
Contiguity BVAR	1.302	0.898	1.605	1.426	1.275	1.005	1.252
UVAR with indicators	1.555	1.213	1.736	1.703	1.398	1.182	1.465
MN BVAR with indicators	1.066	0.919	1.470	0.981	1.186	1.231	1.142
Contiguity BVAR with indicators	1.041	0.868	1.541	1.021	1.147	1.295	1.152

Note: MAPEs are average mean absolute percent error of seven 6-quarter rolling forecasts.

Table 1.5. Aggregate Private Sector Employment Forecast MAPE with Region Specific Indices: 8 quarter

	UP	Northern LP	West Central	East Central	South West	South East	Average
UVAR	1.206	1.912	1.416	1.756	1.199	1.859	1.558
MN BVAR	0.691	0.657	0.739	0.803	0.935	0.896	0.787
Contiguity BVAR	0.653	0.620	0.661	0.747	0.918	0.838	0.739
UVAR with indicators	0.748	0.572	0.342	0.895	0.723	0.380	0.610
MN BVAR with indicators	0.801	0.544	0.494	0.966	0.760	0.386	0.659
Contiguity BVAR with indicators	0.784	0.496	0.487	0.928	0.750	0.320	0.628

Note: MAPEs are average mean absolute percent error of four 8-quarter rolling forecasts.

Table 1.6. Aggregate Pairwise Comparison of 6 Quarters Forecast with Region Specific Indices: Number of Wins (of row forecasting method)

	UVAR	MN BVAR	Contiguity BVAR	UVAR with indicators	MN BVAR with indicators	Contiguity BVAR with indicators	Total Wins
UVAR	NA	1	0	1	0	0	2
MN BVAR	5	NA	2	6	1	2	16
Contiguity BVAR	6	4	NA	6	2	1	19
UVAR with indicators	5	0	0	NA	1	1	7
MN BVAR with indicators	6	5	4	5	NA	3	23
Contiguity BVAR with indicators	6	4	5	5	3	NA	23

Table 1.7. Aggregate Pairwise Comparison of 8 Quarters Forecast with Region Specific Indices: Number of Wins (of row forecasting method)

	UVAR	MN BVAR	Contiguity BVAR	UVAR with indicators	MN BVAR with indicators	Contiguity BVAR with indicators	Total Wins
UVAR	NA	0	0	0	0	0	0
MN BVAR	6	NA	0	2	2	2	12
Contiguity BVAR	6	6	NA	2	2	2	18
UVAR with indicators	6	4	4	NA	5	4	23
MN BVAR with indicators	6	4	4	1	NA	0	15
Contiguity BVAR with indicators	6	4	4	2	6	NA	22

Table 1.8. Aggregate Private Sector Employment Forecast MAPE: National Leading and Coincident Indicators

	UP	Northern LP	West Central	East Central	South West	South East	Average
6 Quarter Forecast							
UVAR with indicators	1.567	1.834	2.023	1.786	1.569	1.822	1.767
MN BVAR with indicators	1.364	0.869	1.347	1.304	1.334	1.064	1.214
Contiguity BVAR with indicators	1.211	0.835	1.493	1.312	1.301	0.925	1.179
8 Quarter Forecast							
UVAR with indicators	1.530	1.104	1.247	1.079	1.394	0.935	1.215
MN BVAR with indicators	0.670	0.665	0.723	0.916	0.815	0.716	0.751
Contiguity BVAR with indicators	0.577	0.615	0.620	0.845	0.855	0.628	0.690

Figure 1.1. Map of 6 Michigan Regions

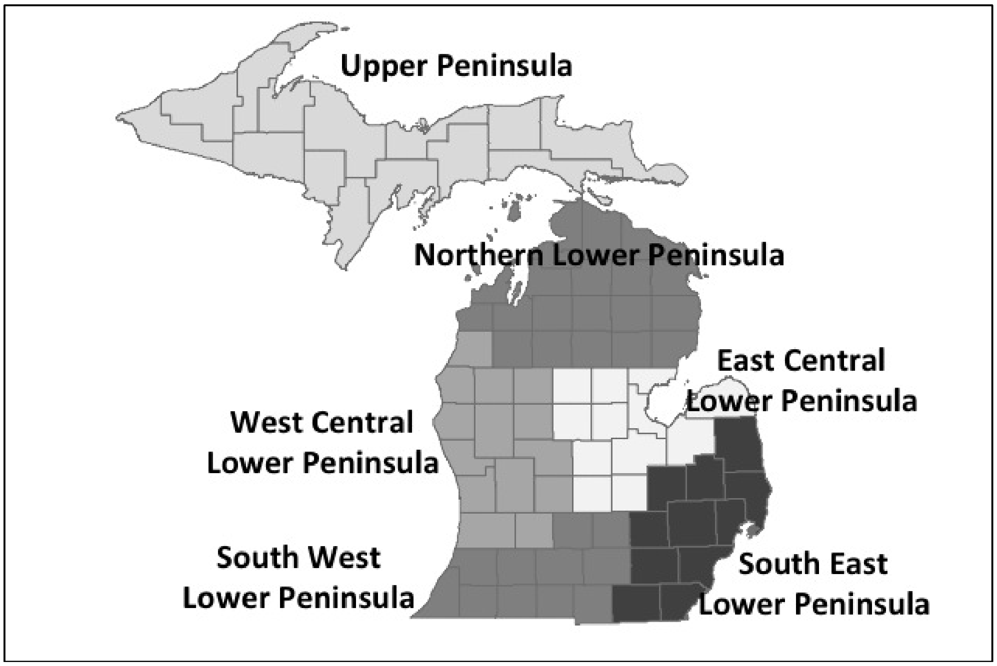
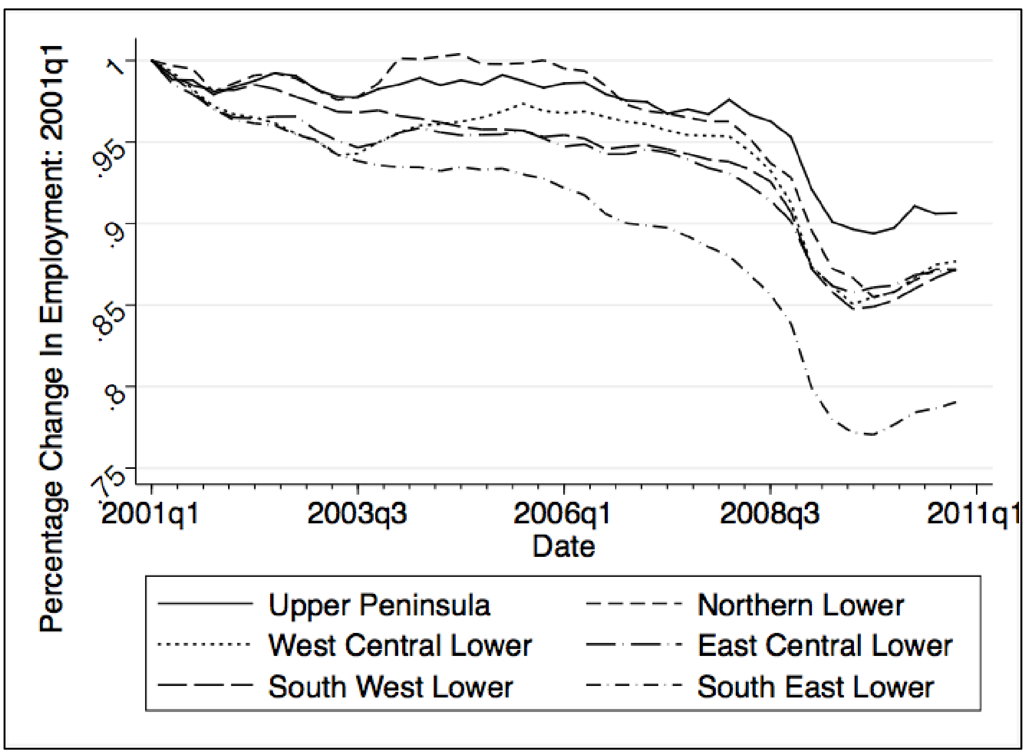


Figure 1.2. Normalized Total Private Employment, by Region



Note: The employment series for each region is seasonally adjusted using the Census X-12 Seasonal Adjustment Program. See http://www.census.gov/srd/www/x12a/ for details.

Figure 1.3. Michigan Total Private Employment Level by Index of Current Financial Situation

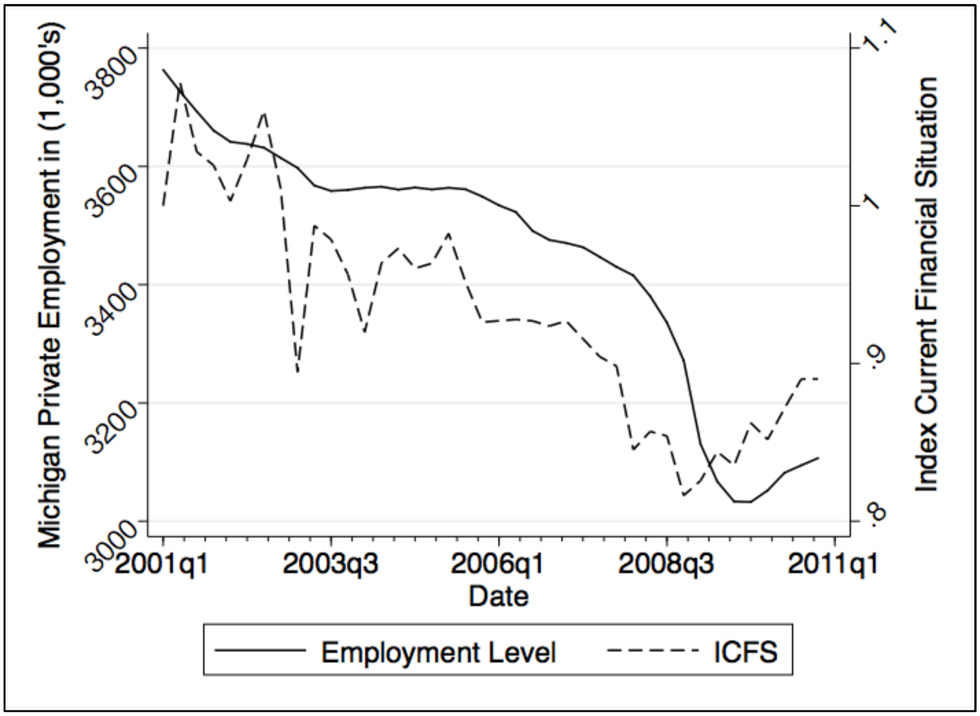
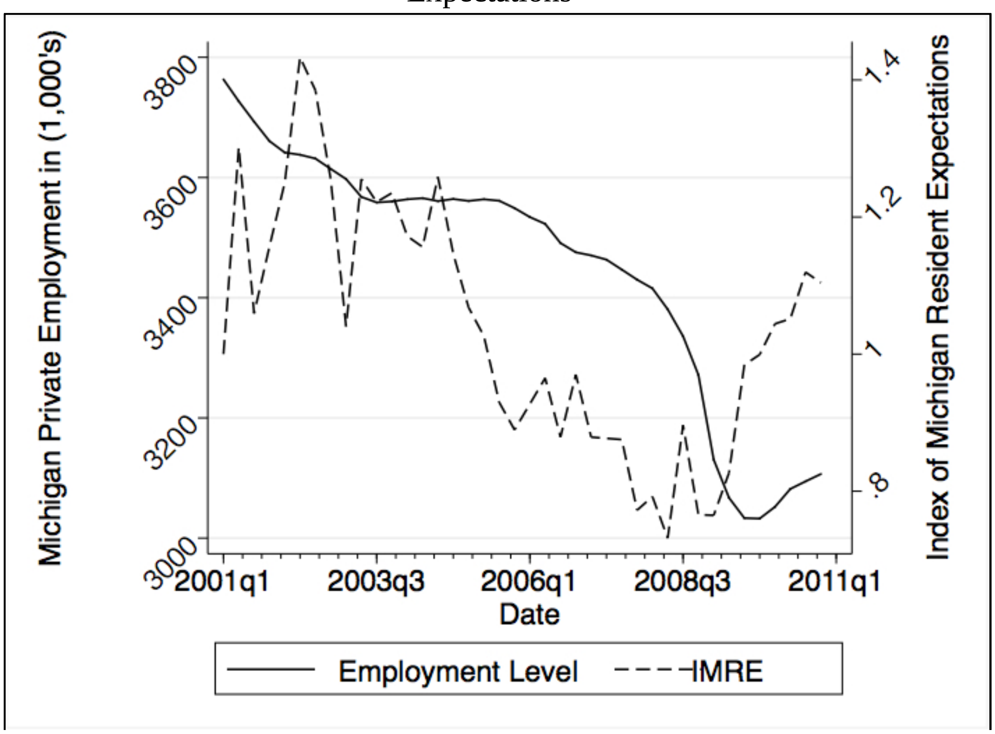


Figure 1.4. Michigan Total Private Employment Level by Index of Michigan Resident Expectations



APPENDIX

APPENDIX

Table 1.A1. State of the State Survey Interview Dates

Time Period	SOSS Number	SOSS Interview Dates	Total Sample Size
2001q1	20	1/14/01 - 02/27/01	954
2001q2	21	5/30/01 - 7/12/01	958
2001q3	22	9/04/01 - 10/31/01	978
2001q4	23	10/26/01 - 12/05/01	1,001
2002q1	24	1/11/02 - 2/19/02	1,012
2002q2	25	3/08/02 - 4/29/02	951
2002q3	26	5/21/02 - 7/16/02	933
2002q4	27 & 28	10/19/02 - 12/31/02; 8/15/02 - 10/26/02	989; 945
2003q1	29	1/21/03 - 3/10/03	1,017
2003q2	30	4/25/03 - 6/16/03	975
2003q3	31	6/27/03 - 8/12/03	965
2003q4	32	10/27/03 - 12/1/03	990
2004q1	33	1/31/04 - 3/14/04	940
2004q2	34	4/19/04 - 6/15/04	962
2004q3	35	9/1/04 - 11/15/04	1,000
2004q4	36	12/6/04 - 02/01/05	965
2005q1	37	2/15/05 - 04/04/05	965
2005q2	38	5/28/05 - 7/18/05	949
2005q3	39	8/10/05 - 9/26/05	988
2005q4	40	10/28/05 - 12/13/05	1,003
2006q1	-	-	-
2006q2	41	2/16/06 - 4/05/06	1,023
2006q3	42	6/16/06 - 8/9/06	959
2006q4	43	8/10/06 - 10/21/06	993
2007q1	44	2/20/07 - 4/23/07	958
2007q2	-	-	-
2007q3	45	7/10/07 - 9/22/07	880
2007q4	46	10/10/07 - 11/26/07	1,001
2008q1	47	1/24/08 - 3/20/08	1,012
2008q2	48	5/4/08 - 7/6/08	1,006
2008q3	49	7/20/08 - 9/14/08	1,010
2008q4	50	10/13/08 - 12/29/08	953
2009q1	51	1/28/09 - 3/25/09	1,001
2009q2	52	5/26/09 - 6/30/09	1,036
2009q3	53	7/23/09 - 8/27/09	1,022
2009q4	54	11/11/09 - 1/12/10	991
2010q1	55	2/3/10 - 4/30/10	1,969
2010q2	56	5/18/10 - 7/13/10	979
2010q3	57	10/25/10 - 12/30/10	1,003
2010q4	58	1/31/2011 - 3/29/11	981

Note: Survey data for 2006q1 and 2007q2 were imputed using linear interpolation.

Table 1.A2. IMREE Diagnostic Tests for Lead/Lag Structures: J-tests and R-squared

Model Tested	Fitted Values of Competing Model				
	2 Quarter Lead	3 Quarter Lead	4 Quarter Lead	5 Quarter Lead	R-squared
2 Quarter Lead	-	0.064	0.000	0.000	0.2896
3 Quarter Lead	0.326	-	0.001	0.001	0.3662
4 Quarter Lead	0.874	0.662	-	0.023	0.5707
5 Quarter Lead	0.557	0.431	0.033	-	0.5529

* Values provided are p-values of the fitted values of the competing model.

Table 1.A3. IMCEC: Diagnostic Tests for Lead/Lag Structures: J-tests and R-squared

Model Tested	Fitted Values of Alternative Model				
	1 Quarter Lag	Coincident	1 Quarter Lead	2 Quarter Lead	R-squared
1 Quarter Lag	-	0.000	0.000	0.000	0.5181
Coincident	0.911	-	0.011	0.005	0.7045
1 Quarter Lead	0.313	0.009	-	0.033	0.6935
2 Quarter Lead	0.375	0.001	0.007	-	0.6597

* Values provided are p-values for the fitted values of the competing model.

REFERENCES

REFERENCES

Banerji, A., Dua P., and Miller S.M. 2006. "Performance Evaluation of the New Connecticut Leading Employment Index Using Lead Profiles and BVAR Models" *Journal of Forecasting*, 25(6), 415-437.

Birkes, David, and Yadolah Dodge. *Alternative Methods of Regression*. New York: John Wiley and Sons, 1993.

Davidson, R. and MacKinnon, J. G.. 1981, "Several Tests for Model Specification in the Presence of Alternative Hypotheses" *Econometrica*, 49, 781-793.

Doan, T., Litterman, R.B. and Sims, C. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1-100.

Dua, P. and S. M. Miller. 1996a. "Forecasting and Analyzing Economic Activity with Coincident and Leading Indexes: The Case of Connecticut." *Journal of Forecasting*, 15(7), 509-26.

Dua, P. and S. M. Miller. 1996b. "Forecasting Connecticut Home Sales in a Bvar Framework Using Coincident and Leading Indexes." *Journal of Real Estate Finance and Economics*, 13(3), 219-235.

Dua, P., S. M. Miller and D. J. Smyth. 1999. "Using Leading Indicators to Forecast Us Home Sales in a Bayesian Vector Autoregressive Framework." *Journal of Real Estate Finance and Economics*, 18(2), 191-205.

Enders W. *Applied Econometric Time Series*. New York: John Wiley and Sons, 1995.

Engemann, K.M., R. Hernandez-Murillo, and M.T. Owyang. 2008. "Regional Aggregation in Forecasting: An Application to the Federal Reserve's Eighth District" *Federal Reserve Bank of St. Louis Regional Economy Development*, 4(1), 15-29.

Farber, H.S. 2011. "Job Loss in the Great Recession: Historical Perspective from the Displaced Workers Survey, 1984-2010" (NBER Working Paper No. 17040).

Guha, D., and A. Banerji. 1998/1999. "Testing For Regional Cycles: A Markov-Switching Approach" *Journal of Economic and Social Measurement,* 25, 163-182.

Hernandez-Murillo, R. and M.T. Owyang. 2006. "The Information Content of Regional Employment Data for Forecasting Aggregate Conditions." *Economics Letters* 90, 335-339.

Howrey, E.P. 2001. "The Predictive Power of the Index of Consumer Sentiment." *Brookings Papers on Economic Activity*, 1:2001, 175-216.

Institute for Public Policy and Social Research. Various Years. State of the State Survey. Michigan State University. East Lansing, Mich. Available on World Wide Web: http://www.ippsr.msu.edu/SOSS.

Koch, P. D. and R. H. Rasche. 1988. "An Examination of the Commerce Department Leading-Indicator Approach." *Journal of Business & Economic Statistics*, 6(2), 167-187.

Lahiri, K. and G.H. Moore. 1993. “Introduction.” In K. Lahiri and G.H. Moore (ed.), *Leading Indicators: New approaches and forecasting records*. Cambridge: Cambridge University Press.

LeSage, J.P., and A. Krivelyova. 1999. “A Spatial Prior for Bayesian Vector Autoregressive Models.” *Journal of Regional Science*, 39, pp. 297-317.

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

Litterman, R.B. 1980. “A Bayesian Procedure for Forecasting With Vector Autoregression,” working paper, Massachusetts Institute of Technology, Department of Economics.

Magura, M. 1998. “IO and Spatial Information as Bayesian Priors in an Employment Forecasting Model.” *Annals of Regional Science*, 32, pp. 495-503.

Miller, J.R. 1998. “Spatial Aggregation and Regional Economic Forecasting.” *Annals of Regional Science*, 32, pp. 253-266.

National Bureau of Economic Research. “US Business Cycle Expansions and Contractions.” Accessed at <http://www.nber.org/cycles.html> January, 2012.

Petersen, B. and S. Strongin. 1996. "Why Are Some Industries More Cyclical Than Others?" *Journal of Business & Economic Statistics*, 14(2), pp. 189-198.

Quarterly Census of Employment and Wages (2006). Technical Notes and References: Bureau of Labor Statistics.

Rey, S.J. 2000. “Integrated Regional Econometric Input-output Modeling: Issues and Opportunities.” *Papers in Regional Science*, 79, pp. 271-292.

Rey, S.J. 1998. “The Performance of Alternative Integration Strategies for Combining Regional Econometric and Input-output Models.” *International Regional Science Review*, 21, pp. 1-36

Rickman, D. S.; S. R. Miller and R. McKenzie. 2009. "Spatial and Sectoral Linkages in Regional Models: A Bayesian Vector Autoregression Forecast Evaluation." *Papers in Regional Science*, 88(1), 29-41.

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

Sims, C. A.; J. H. Stock and M. W. Watson. 1990. "Inference in Linear Time-Series Models with Some Unit Roots." *Econometrica*, 58(1), 113-144.

Tan, H. and J. A. Mathews. 2010. "Identification and Analysis of Industry Cycles." *Journal of Business Research*, 63(5), 454-462.

Theil, H. and A. S. Goldberger. 1961. "On Pure and Mixed Statistical Estimation in Economics." *International Economic Review*, 2(1), 65-78.

Todd, R.M. 1984. “Improving Economic Forecasting with Bayesian Vector Autoregression,” *Federal Reserve Bank of Minneapolis Quarterly Review*, 8, 18-29.

U.S. Census Bureau. 2011. X-12-ARIMA Reference Manual (Version 0.3), Washington, D.C.

Vinod H.D. 1978. “A Survey of Ridge Regression and Related Techniques for Improvements Over Ordinary Least Squares” *Review of Economics and Statistics,* 60, 121-131.

Zellner, A. 1979. “Statistical Analysis of Econometric Models,” *Journal of American Statistical Association*, 74, 628-664.

Zellner, A., and F. Palm. 1974. “Time Series Analysis and Simultaneous Equation Econometric Models,” *Journal of Econometrics*, 2, 17-54.

CHAPTER 2: The Distribution of Firm Size and Regional Economic Growth: Examining the Entrepreneurial Pipeline Theory for the United States

2.1 Introduction

A frequently heard debate in community-level economic development planning sessions is whether to invest more effort in recruiting large firms or to focus on assistance to smaller local businesses. This civic debate has its parallels in the academic literature. Successful job creation strategies can mitigate poverty, raise incomes, and increase employment. These goals have been thought of as key aspects of economic growth and have often been a priority to economic development practitioners (Shaffer, Deller, and Marcouiller, 2006). Yet persistent poverty remains a critical social issue facing policymakers in the United States.

Job creation policies are an often-debated economic development strategy among policymakers. This debate centers on several waves of economic development practices that have been adopted over the years in an attempt to improve economic outcomes. The main focus of the first wave is industrial recruitment of firms from outside the region with financial incentives. This wave remains popular (Hodge, 2011) despite various criticisms (Loveridge, 1996). Industrial recruitment tactics often include government subsidies and tax breaks, as part of what some have called "smoke-stack chasing" (Bradshaw and Blakely, 1999).

Attempts to recruit firms from outside the region involve financial burdens. In addition, regions have experienced increasing competition for recruitment, as more and more regions have adopted these methods. Thus, the second wave moved past the zero-sum game of attracting outside firms, to retention and expansion of existing firms within the region (Morse, 1990; Allanach and Loveridge, 1998). However, globalization and structural adjustment have led many to reevaluate the effectiveness of these traditional economic development approaches and the determinants of regional economic growth (Drabenstott, 2006).

Regional development practitioners and researchers have shifted the focus of their work toward innovation and entrepreneurship (Von Bargen et al., 2003; Olfert and Partridge, 2010). This has renewed interest in the promotion of small businesses as an economic development strategy (Aquilina et al., 2006). Emphasis on the role of small businesses and entrepreneurship dates back to the work of Schumpeter (1942, 1961), who highlights the role of the innovative entrepreneur in economic growth. Both Deller and McConnon (2009) and Shaffer (2006) lay out the theoretical arguments for the role of small firms in regional economic growth. In addition to Schumpeterian innovation, another argument for the importance of small firms is the flexibility of microenterprises in changing environments, especially in manufacturing. Small firms are also often relatively labor intensive. Finally, the entrance of small firms into a market may enhance competition, resulting in greater efficiencies for the existing firms, as well as the new firms.

These theoretical arguments have prompted a considerable amount of empirical investigation. Birch (1979, 1981, 1987) provided early empirical support for the idea that small businesses are net job creators. Using state-level data, Robbins et al. (2000) found that gross state product and productivity have grown faster in states in which a large proportion of the workforce is in small firms. Similarly, the results of Shaffer (2006) suggest that, in most cases, smaller establishments are associated with faster rates of employment growth, both within and across sectors. These findings have unleashed a large array of new strategies for encouraging entrepreneurs. (For an overview, see Walzer, 2007.)

Lichtenstein and Lyons (2006, 2010) offer a subtle critique of business-related economic development policies that focus on either small businesses or industrial recruitment of large firms. They argue that entrepreneurship from small businesses alone does not act as the engine of economic growth. Instead, they posit that entrepreneurs come from all business sizes, and that it is the distribution of businesses across size classes that matter for economic growth. In this framework, businesses of all sizes act interdependently within a regional economy. Through growth, smaller firms can replace larger firms that have ceased to be economically viable due to changes in the economy. Loveridge and Nizalov (2007) note that interdependence between businesses from different size classes can come from local production and consumption linkages, along with externalities, such as amenities from small firms and a critical mass for services from large firms. Thus, the so-called "entrepreneurial pipeline theory," makes a case that agglomeration economies and diversification based on firm size influence local economic growth.

The objective of this study is to examine the effect of the business size distribution on income and employment growth in U.S. counties from 1990 to 2000. Specifically, I ask whether a given distribution of firm sizes (the best available proxy for a community's pipeline of entrepreneurs from Lichtenstein and Lyons, 2006, 2010) in time t leads to higher income and employment growth in time $t + i$. To examine empirically the link between the distribution of business sizes and economic growth, I estimate a growth model using a generalized method of moments (GMM) procedure that accounts for spatial spillovers. The growth model is grounded in regional economic theory, and controls for a wide range of covariates thought to influence economic growth. The firm size distribution is measured in several ways that include the employment share in nine firm-size categories, business distribution indices somewhat analogous to a Gini coefficient and the Atkinson index of inequality. Using several different measures of the firm-size distribution provides a more robust empirical assessment of the entrepreneurial pipeline theory. The results show statistically significant relationships between the distribution of employment across size classes and county-level employment and per-capita income growth.

2.2 The entrepreneurial pipeline theory and the size distribution of firms

Lichtenstein and Lyons (2006, 2010) propose a conceptual model, which they refer to as the entrepreneurial pipeline theory. The Lichtenstein and Lyons framework was developed over many years of hands-on regional and business development assistance. The authors argue that firms of all size classes can have entrepreneurial characteristics. This is a view consistent with the mainstream business literature. Lichtenstein and Lyons make a key contribution to the business literature by arguing that the full distribution of firm sizes may act as an important determinant of economic growth. This happens due to potential agglomeration economies across firm size categories. The framework argues for a tailored policy that looks for gaps in the pipeline, instead of a one-size-fits-all policy favoring smaller or larger businesses.

The interdependence between business size classes may result from production and consumption linkages of firms locating in close proximity. Regional economics has documented the theoretical and empirical aspects of the positive externalities from agglomeration economies (Blair and Premus, 1987; Carlino, 1982; Hansen, 1990; Krugman, 1991; Devereux, Griffith, and Simpson, 2007). Agglomeration economies have traditionally been viewed as providing impacts in a region that are business-scale neutral. However, the presence of firms of different sizes contributing to regional growth and resilience is a newer concept associated with the entrepreneurial pipeline theory of Lichtenstein and Lyons. Differences in firm sizes may play a role in regional productivity gains. For example, Komarek and Loveridge (under review) suggest that small firms could act as a training ground for workers who go on to take higher paying positions in larger firms. This is particularly beneficial for skill sets that are useful across industries. The training relationship may appear one sided, where large firms benefit from smaller firms finding and training workers. Small firms may benefit from the relationship as well. Workers might be willing to accept lower wages in a small firm with the hopes of moving up the career ladder towards the larger, higher paying firm. Small firms may also benefit from contacts within the larger firm via its former employees, yielding orders or vital industry intelligence. Finally, small firms may provide services or worker amenities that reduce the costs of large firms, while large firms provide a market and role modeling for the smaller firms.

At the community or region-wide level, firms located across the size distribution continuum tend to diversify industry concentration in the local economy. Firms across different sizes might act as a buffer in the regional economy. In particular, a region with small firms with the ability to grow may be more apt to weather the death or departure of a large anchor firm. Furthermore, the presence of a set of larger firms in the region tends to raise overall pay levels, creating income that allows for other kinds of investments (e.g. service sector business, education) in the region.

Lichenstein and Lyons do not test the entrepreneurial pipeline theory empirically. Loveridge and Nizalov (2007) explore the validity of the entrepreneurial pipeline theory for the state of Michigan. They use fixed-effects generalized least squares estimation for 12 years of county-level data. They find strong links between a county's business size distribution and its job and income growth in Michigan. On the other hand, Komarek and Loveridge (under review) extend Michigan entrepreneurial pipeline results by Loveridge and Nizalov (2007) to four multi-state high-poverty regions as well as the continental U.S. The results suggest a connection between employment growth and the distribution of firms across size categories for Appalachia, the Plantation Belt, and the continental U.S. The authors find no statistically significant relationship for the Great Plains or the Borderlands.

Fotopoulos (forthcoming) makes an important point about the role of entrepreneurship in economic growth that supports Lichtenstein and Lyons, at least with respect to the smaller end of the business size continuum. Using self-employment as a proxy for entrepreneurship, he examines the relationship between self-employment rates and per-capita income growth in 197 European regions across fifteen countries. While using self-employment rates as a proxy for entrepreneurship has its critics (Parker, 2004), Fotopoulos finds an L-shaped relationship between self-employment rates and per capita income growth. This indicates that in some EU regions, growth might be faster with less self-employment. The Fotopolous result provides an important caveat to those who promote entrepreneurship as an economic development strategy: it may not be appropriate for all regions.

To examine the distribution of firm sizes at the county level, I follow a modified version of the techniques employed by Loveridge and Nizalov (2007). I use data from the U.S. Census Bureau's County Business Patterns, which provides information on the number of establishments (firms) in nine employment categories. The smallest employment category is firms with 1-4 employees, while the largest range consists of firms with 1,000 or more employees. Figure 2.1 shows the distribution of establishments across employment size ranges for U.S. counties in 1990. Figure 2.1 clearly shows that business numbers are concentrated in smaller employee size categories. Firms with 1-4 employees make up almost 60% of business establishments. Section A of table 2.1 provides further details on the establishment share distribution across each employment range.

While the distribution of establishments across several employee size ranges provides a useful insight into the general structure of the local business environment, it does not reflect the role that large businesses with several hundred employees can have on the local economy. To take into account the potentially disproportionate impact of large firms, I weight the number of establishments in each size category by employment size. In weighting the number of establishments in each range by the number of people employed by those establishments, I provide a more accurate picture of the local business climate. Figure 2.2 illustrates the share of employment by business size category in 1990. The new employment weighted establishment distribution in figure 2.2 is much more uniform in shape. The employment share percentages for each of the size categories are presented in Section B of table 2.1.

Following Loveridge and Nizalov (2007), I use the employment-weighted establishment distribution across the nine business size categories as a starting point for exploring the role of firm size on economic development, thereby exploring the validity of the entrepreneurial pipeline theory. Empirically, I regress aspects of economic development (employment growth and per-capita household income growth) on the employment shares across business size classes.

It is also helpful to represent the employment share distribution in a more parsimonious way. To accomplish this I use several well-known indices from the economics literature that measure the inequality of a distribution. First, I construct an index similar to a Gini coefficient for the employment share distribution. The index shares similarities to a Gini coefficient, because it is a bounded measure of deviation from a uniform distribution, and is agnostic to the optimal distribution. The index is represented in equation (2.1a) as one half the sum of absolute deviations of employment shares in each category from a uniform distribution (11.1%).

$$\textit{Uniform Index} = \frac{1}{2}\sum_{i=1}^{9}\left|x_i - 11.1\%\right| \tag{2.1a}$$

To interpret the index within an empirical framework, it is important to note that the index represents the extent to which a county deviates from a uniform distribution of employment across business size categories. Thus, a larger value for the index indicates that the current distribution of employment across size categories is more unequal. The index is useful for examining whether a uniform distribution is optimal for county level economic growth. If a uniform distribution is optimal, then a larger value of the index would correspond negatively to economic growth, and *vice versa*. Section C of table 2.1 shows that, for U.S. counties in 1990, the average index is 23.2%. The value of the index ranges from 12.7% to 67.1%.

The index described in equation (2.1a) sets up a straw-man hypothesis that the growth enhancing distribution of employment across business size categories is uniform. Examining the coefficient on the uniform index in an empirical model then provides a test of this hypothesis. There are several shortcomings of the straw-man hypothesis of uniformity. First, the index treats all deviations (positive and negative) the same. Second, it does not shed further light on the debate between small versus large firms and economic growth. To contribute to this debate I modify the index proposed in equation (2.1a). Specifically, I consider a hypothesis that a higher share of employment in large firms is growth enhancing. I also create a similar modified index for small firm sizes. The index in equation (2.1a) is modified to equations (2.1b) and (2.1c) in the following way:

$$\textit{High Uniform Index} = \frac{1}{2}\sum_{i=3}^{9}\left|x_i - 14.3\%\right| + \frac{1}{2}\sum_{i=1}^{2}\left|x_i - 0\right| \tag{2.1b}$$

$$\textit{Low Uniform Index} = \frac{1}{2}\sum_{i=1}^{7}\left|x_i - 14.3\%\right| + \frac{1}{2}\sum_{i=8}^{9}\left|x_i - 0\right| \tag{2.1c}$$

where the High Uniform Index is the summation of deviations from a uniform distribution for the seven largest employment size categories. Similarly, the Low Uniform Index is the summation of deviations from a uniform distribution for the seven smallest employment size categories. Thus, the High Uniform Index assumes that firms in the smallest size categories have little influence on economic growth, and *vice versa*. The distribution described in equations (2.1a) – (2.1c) are shown graphically in figure 2.3.

Furthermore, I test the robustness of the Gini-coefficient inspired indices by also using the Atkinson index of inequality (Atkinson, 1970). The Atkinson index is defined by equation (2.1d),

$$Atkinson\ Index = 1 - \left[\frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{\bar{x}}^{1-\varepsilon} \right]^{1/1-\varepsilon} \tag{2.1d}$$

where N is the sample size, x_i is the employment share in category i, $\bar{x}$ is the average employment share size, and ε is a inequality preference parameter. The Atkinson index was first developed as an alternative to the Gini-coefficient to measure income inequality. One key difference between the two indices is the parameter ε, which measures aversion to inequality. A larger value for the parameter ε suggests a higher degree of aversion to an unequal distribution. Similar to the alternative indices described above, a larger (smaller) value of the Atkinson index implies the employment share distribution is more (less) unequal. In the analysis that follows, I use two inequality aversion parameters suggested by the literature (eg. see Decancq and Decoster, 2009): .5 and 2.

2.3 Regional growth model

I use a simple model of regional economic growth to frame the productivity gains suggested by the entrepreneurial pipeline theory on local employment and labor income. The model focuses on employment and income, because of their importance to policymakers

(Hammond and Tosun, 2011), and to keep with the previous literature on the entrepreneurial pipeline theory. In the next section, on the empirical specification, I discuss some of the limitations of the following theoretical model and empirical techniques to account for them. The regional economic growth model follows Glaeser, Scheinkman and Shleifer (1995). Let total output in county *i* at time *t* equal Y_{it}, which is a function of county technology A_{it} and employment L_{it}:

$$Y_{it} = A_{it} L_{it}^{\alpha} \tag{2.2}$$

The Cobb-Douglas production function in equation (2.2), where α<1, is the same across all counties. The marginal product of labor yields worker's labor income, $w_{it} = \alpha A_{it} L_{it}^{\alpha - 1}$. The utility for households in county *i* at time *t* is the product of labor income and a quality of life index denoted $L_{it}^{-\lambda} Z_{it}$. $\lambda > 0$ represents the negative consequences of increasing county size, such as higher housing prices and congestion externalities, among other factors. The utility is described by the following equation:

$$U(.) = \alpha A_{it} Z_{it} L_{it}^{\alpha - \lambda - 1} \tag{2.3}$$

Households are allowed to move freely across counties, such that in equilibrium utility will be constant across space at any point in time. Thus, the utility in each county equals to the reservation utility U_{it}^{r} at time *t*. It is convenient to take logs and differences of equation (2.3) to reveal growth rates.

$$\ln U_{it-1}^{r} - \ln U_{it}^{r} = \left(\ln A_{it-1} - \ln A_{it}\right) + \left(\ln Z_{it-1} - \ln Z_{it}\right) + \left(\alpha - \lambda - 1\right)\left(\ln L_{it+1} - \ln L_{it}\right) \tag{2.4}$$

The growth in utility is made up of productivity growth, growth in regional quality of life and employment growth. Furthermore, assume that growth in quality of life, and productivity growth are determined by observable county level characteristics at the baseline time *t* denoted X_{it} and unobservable county characteristics, ψ_{it+1} and ζ_{it+1} respectively such that:

$$\ln A_{it+1} - \ln A_{it} = X'_{it}\gamma + \psi_{it+1} \tag{2.5}$$

$$\ln Z_{it+1} - \ln Z_{it} = X'_{it}\theta + \zeta_{it+1} \tag{2.6}$$

The coefficient γ is the effect of observed county characteristics (from the baseline time period *t*), such as the firm size distribution, on productivity growth in county *i* between time periods *t* and *t+1*. By substituting equations (2.5) and (2.6) into the reservation utility equation (4) and rearranging terms, it is possible to obtain an equation describing employment and labor income growth as a function of county characteristics:

$$\ln L_{it+1} - \ln L_{it} = \left(X'_{it}(\gamma+\theta)\right)/(1+\lambda-\alpha) + \eta_{it+1} \tag{2.7}$$

$$\ln w_{it+1} - \ln w_{it} = \left(X'_{it}(\lambda\gamma+\alpha\theta-\theta)\right)/(1+\lambda-\alpha) + v_{it+1} \tag{2.8}$$

where η_{it+1} and v_{it+1} are error terms uncorrelated with county characteristics[8]. More generally changes in county employment and labor income can be thought of as a function of the county characteristics:

$$\ln L_{it+1} - \ln L_{it} = X'_{it}\beta + \varepsilon_{it+1} \tag{2.9}$$

$$\ln w_{it+1} - \ln w_{it} = X'_{it}\beta + \varepsilon_{it+1} \tag{2.10}$$

---

[8] The error terms in equation (2.7) and (2.8) are defined as:

$$\eta_{it+1} = \frac{1}{(1+\lambda-\alpha)}\left(\psi_{it+1} + \zeta_{it+1} - \left(\ln U^r_{it+1} - \ln U^r_{it}\right)\right)$$

and

$$v_{it+1} = \frac{1}{(1+\lambda-\alpha)}\left(\lambda\psi_{it+1} + (\alpha-1)\zeta_{it+1} + (1-\alpha)\left(\ln U^r_{it+1} - \ln U^r_{it}\right)\right)$$

In each equation, β is a vector of coefficients to estimate that include determinants of productivity, and ε_{it+1} is an error term. Equations (2.9) and (2.10) show the conventional cross-section growth model, where the dependent variable is a growth rate and the control variables represent the initial conditions or baseline time period. Equations (2.9) and (2.10) serve as the basis for the empirical models to follow.

2.4 Empirical specification

The regional growth model presented above is used primarily to motivate the connection between county-level productivity gains and growth in employment and income. This connection is particularly useful considering the agglomeration benefits suggested by the entrepreneurial pipeline theory. Equations (2.9) and (2.10) are the basis for the empirical work that follow. The models are also consistent with commonly used reduced form regional growth models stemming from Steinnes and Fisher (1974) and Carlino and Mills (1987). Traditionally, the Steinnes and Fisher and Carlino and Mills models have focused on employment growth. However, Deller et al. (2001) expanded the reduced form version of the Carlino and Mills (1987) model to include income. The authors argue that income helps to trace out the regional growth process and capture job quality (measured by income levels).

Both the empirical literature and economic theory suggest that employment and income adjust to equilibrium levels with a substantial lag (e.g., Mills and Price, 1984; Carlino and Mills, 1987; Boarnet, 1994; Duffy, 1994; Duffy-Deno, 1998; Henry et al., 1999; Aronsson et al., 2001; Deller et al., 2001; Edmiston, 2004; Barro and Sala-i-Martin, 1991, 1992; Higgins et al., 2009). For example, neoclassical growth theory suggests that income growth rates are negatively related to the initial per-capita income levels, due to decreasing returns to capital. Similarly, disequilibrium in employment growth rates tend to be persistent due to frictions in mobility of households and firms. To account for this I incorporate a partial adjustment process by including the initial employment and income levels (i.e. time *t*) in the county-level observed characteristics X_{it}. To mitigate endogeneity issues for all the explanatory variables, I use initial time period values (i.e. time *t*) for the explanatory variables. This has become common practice in regional growth models. Finally, X_{it} includes the same explanatory variables for both income and employment growth, which incorporates industry structure variables for examining the entrepreneurial pipeline theory.

OLS estimates are based on the assumption that the error terms from different counties are independent. However, the regional factors that affect household and firm decision-making along with income growth are likely to display spatial autocorrelation (Anselin, 2003, 1988). For example, unobserved factors in neighboring locations could be correlated, because political boundaries often do not necessarily correspond to economic regions. Thus, assuming that the error terms in the spatially organized data are independent may be overly restrictive. I allow the disturbances to be spatially autocorrelated by using a spatial error model (LeSage and Pace, 2009). I assume that the disturbances are generated by the spatial processes in equation (2.11):

$$\varepsilon_t = \rho W \varepsilon_t + u_t \tag{2.11}$$

W is a minmax-normalized spatial first-order contiguity matrix of typical element w_{ij}. Normalization is accomplished by dividing by the minimum of the largest row sum and column sum of matrix **W**. ρ is a vector of parameters to be estimated, and the vector of error terms u_t, is assumed to be identically and independently distributed with mean zero and variance σ^2. To estimate equations (2.9) and (2.10) with spatially autocorrelated errors generated by equation

(2.11), I use a Method of Moments estimator detailed in Drukker et al. (2010). Furthermore, a two-stage least squares procedure also outlined in Drukker et al. (2010) is used for the employment growth equation.

2.5 Data

Regional economic theory suggests that firm location decisions are influenced by local business conditions, the supply of inputs, government policies, and distance or availability of markets for goods and services (Deller et al., 2001). These broad groups of explanatory variables are also among the explanatory variables thought to influence regional income growth (e.g. Higgins et al., 2006, Higgins et al., 2009, Rupasingha et al., 2002). Thus, to remain consistent across the regional income and employment growth models, I use the same explanatory variables in each. The data come from the County City Data Book, Economic Research Service of the U.S. Department of Agriculture, and the County Business Patterns. After removing counties with missing data, 3,046 usable observations for the continental U.S. were left. Table 2.2 provides descriptive statistics and the source for each variable used in the analysis.

2.5.1 Dependent variables

The dependent variables in the empirical analysis are the per-capita income growth from 1990-2000 and the employment growth from 1990-2000. Both of the dependent variables are measured as the natural log difference between the two time periods. The consumer price index from the Bureau of Labor Statistics is used to adjust income levels to year 1990 prices.

2.5.2 Independent Variables

The independent variables in the empirical analysis include local business conditions, the supply of inputs, the local government, and the availability of markets for goods and services. I use the initial conditions (i.e. year 1990 values) for the independent variables. It could be argued that some of the control variables are not truly independent. Nonetheless, using initial conditions reduces the endogeneity problem for analyzing employment and income growth. The control variables and categories used here are in line with those used in several studies of the regional determinants of economic growth (e.g. Deller et al., 2001).

Local Business Conditions

A wide variety of variables are used to proxy the overall local business conditions in a region. Among these variables are aspects of the firm size distribution, which are proxies for the entrepreneurial pipeline theory as well as variables controlling for industry structure and agglomeration.

1. Firm Size Distribution Indices 1990
2. Employment Shares by Establishment Size Categories 1990
3. Percent Employment in Agriculture 1990 (AG EMP)
4. Percent Employment in Manufacturing 1990 (MAN EMP)
5. Percent Employment in Services 1990 (SERVICE EMP)
6. Percent Employment in Trade 1990 (TRADE EMP)
7. Establishment Density 1990 (EST DENSITY)

Supply of Inputs

The supply of inputs is intended to capture the ability of the regional market to produce goods and services. This category contains variables that measure the level of human capital, and local labor market conditions.

1. Percent Population 25-44 1990 (POP 25-44)
2. Percent College Degree or higher 1990 (COLL PLUS)
3. Unemployment Rate 1990 (UNEMP)

Government

Local governments use taxes to finance local infrastructure and public services. High personal and business taxes are often thought to be detrimental to local economic growth. However, the services that governments provide from tax revenue are often thought of as growth enhancing. I use two variables to represent these alternate views of the effect of local government on economic growth.

1. Property Taxes Per Capita 1992 (PROP TAX)
2. Total Government Expenditure Per Capita 1990 (TOTAL GOVT EXP)

Markets

The variables in this category are used to capture the factors that influence the demand side of regional markets. In particular, I use demographics to control for the regions' consumption ability and demand for goods and services.

1. Population Density 1990 (POP DENSITY)
2. Percent Urban 1990 (URBAN)
2. Percent Population White 1990 (POP WHITE)
3. Percent Population Over 65 1990 (POP 65+)
4. Percent Population 5-17 1990 (POP 5-17)

It is common practice in the literature (and for ease of interpretation) to use control variables that are in percentage or per-capita terms. Thus, each of the independent control variables listed above enter the empirical model as levels, and in turn can often be interpreted akin to elasticities. The exceptions to this are the initial conditions for employment and per-capita income. The initial conditions of employment and per-capita income enter the empirical model as natural logs.

2.6 Reduced form estimation results

Table 2.3 presents estimates of the reduced-form employment and income growth models in equations (2.9) and (2.10) when using the shares of employment by business-size categories. Tables 2.4 and 2.5 displays estimates of the 3 Gini coefficient inspired indices described by equations (2.1a) – (2.1c), and table 2.6 shows the Atkinson index from equation (2.1d). Overall, the results suggest that the business size distribution is a significant determinant of employment and per-capita income growth. Many of the estimated coefficients in each equation are statistically significant, and the measures of the distribution are jointly significant in all specifications. Furthermore, the models allow for spatial autocorrelation in the disturbances through the parameter ρ. The results in all specifications reveal positive and statistically significant values for ρ. A positive estimate for ρ indicates that random shocks to both employment and income growth in county *i* is also correlated with a shock to contiguous counties.

The effect of the business size distribution on economic growth is evaluated in complementary ways in tables 2.3, 2.4, 2.5 and 2.6. First, table 2.3 shows estimates where the employment shares are included as an explanatory variable. Columns 1 and 2 display the effects on employment growth, and column 3 and 4 considers per-capita income growth. Using each employment share category as an explanatory variable allows me to examine the role of the broader distribution. I am restricted to using the size categories reported in the U.S. Census County Business Patterns. Although these categories do not provide the complete size distribution of firms, they nevertheless are sufficiently detailed and represent a good approximation of the full range of business sizes in the U.S.

In table 2.3, the first two columns display estimates for the reduced form model proposed by equation (2.9). The estimates for the employment shares show the effect of a marginal change of an employment share in comparison to the omitted category (firms with 1,000+ employees). The results in table 2.3 verify previous research on the importance of small businesses and micro-enterprises for employment growth (e.g., Deller and McConnon, 2009; Robbins et al., 2000). In column 2, when adding the control variables, all of the coefficients of the employment shares are statistically significant for employment growth. A common feature for columns 1 and 2 is that magnitudes of the effect are the smallest at the largest firm sizes. On the other hand, the estimates shown in columns 3 and 4 suggest that the small firm size categories have little or negative effect on per-capita income growth. Thus, on average, a county increasing the share of its employment accounted for by small businesses may spur employment growth, but have a minimal influence on per-capita income growth.

The estimates in table 2.3 suggest an important difference between employment growth and per-capita income growth. The employment growth model shows that increasing the share of small firms, on average, will provide the largest benefits for employment growth. This suggests that the distribution with a higher proportion of small employers (equation 2.1c) might be growth-enhancing for employment. On the other hand, regional income growth benefits the most from medium sized firms (10 – 19 to 100 - 249 employees). This result is consistent with the literature that suggests that small firms create jobs, but the jobs created in small firms receive relatively low wages. Both sets of results suggest that large firms are less effective for employment and income growth.

In tables 2.4 and 2.5, I present results for the specification in which the business size distribution is represented by the Gini coefficient indices shown in equation (2.1a) – (2.1c). The results for the control variables are qualitatively and quantitatively similar across models. Therefore, in tables 2.4 and 2.5 I suppress the values of covariates to focus on the estimates for the three indices. The estimates for the control variables for each model using an index can be found in table 2.A1 of the appendix. Following Loveridge and Nizalov (2007) and Komarek and Loveridge (under review) I use each index and the index squared as explanatory variables. A positive (negative) coefficient for an index variable implies that deviating from (heading towards) the specified distribution of employment share categories are growth enhancing. That is, a positive coefficient suggests that the initial hypothesis is not optimal for economic growth. The quadratic index term suggests whether the effect of the deviation from the uniform distribution is increasing or decreasing.

Using the indices outlined in equations (2.1a) – (2.1c) can help to shed light on the role of the business size distribution. Furthermore, the indices weigh in on the debate between policies advocating for either small or large firms. In the reduced-form employment growth model, the coefficients for the indices hypothesizing a uniform distribution (equation 2.1a) and high proportion of large firms (equation 2.1b) are positive and statistically significant. This suggests that employment growth increases as the employee share distribution moves away from either a large firm dominant or uniform distribution. However, the low uniform index (equation 2.1c) suggests the opposite result for employment growth, with an estimated coefficient of -.6. This suggests employment growth increases with an employment share distribution concentrated with firms in the small size categories. Furthermore, the low uniform index-squared term is negative and statistically significant, which means that the benefits to economic growth from moving towards a small firm size skewed distribution are increasing. The per-capita income equation suggests a similar effect as the uniform and high employment results. In the income growth equations, the estimates are positive and statistically significant for the uniform and high-firm-skewed distribution. Again, this suggests that deviations from these distributions increase per-capita income growth.

The Atkinson index provides an alternative measure of the inequality of a distribution, and is useful for testing the robustness of the entrepreneurial pipeline theory. The results for the Atkinson index are shown in table 2.6. Similar to the previous Gini coefficient results, a higher value of the Atkinson index relates to a more unequal employment share distribution. The Atkinson index also benefits from having an inequality aversion parameter. While the Atkinson index does not provide further insights into the employment and income growth implications of large and small firms, nonetheless, it does allow for varying levels of inequality. Larger values of the inequality aversion parameter mean the Atkinson index is more averse, or sensitive to an unequal distribution. The results in table 2.6 use two different levels of aversion to inequality. Columns 1 and 2 use an inequality aversion parameter of .5, while columns 3 and 4 use an Atkinson index that is relatively more averse to an unequal firm size distribution with a parameter of 2. The results in table 2.6 support the previous Gini coefficient results. In particular, they show that employment growth increases with a more unequal employment share distribution for both types of inequality aversion parameters.

By comparison, in a study limited to the state of Michigan, Loveridge and Nizalov (2007) find a positive coefficient for a similar employee-share index for both income and employment growth, suggesting that deviating from the uniform distribution is optimal. This may be a result of Michigan's special character as a state with massive concentration in large-scale but declining manufacturing firms. The Loveridge and Nizalov study differs from the current research in several ways. First, the current study examines the continental U.S., while previous work examined a specific state. Second, they consider annual growth rates, which measures the short-run relationship between the size distribution and growth. This study considers economic growth over a longer time period (1 decade). Third, the current study allows for a wide range of covariates to control for heterogeneity in economic growth.

Next, I move on to examining the control variables in columns 2 and 4. In both the employment and income growth models, the initial conditions (LN EMP and LN INC) were both negative and statistically significant. This suggests that for income and employment, higher levels in 1990 tend to result in lower rates of growth from 1990 to 2000, all else equal. These results are consistent with neoclassical growth theory, which predicts that regions with initially lower incomes “catch-up” by having higher income growth rates overtime.

The firm size distribution, discussed above, in part controls for the local industry structure. I also included the proportion of employment in several types of industries, along with the establishment density, to control for the local labor market and industry structure. The estimates show that larger initial shares of employment in the service and trade industries yield elevated employment growth, while manufacturing and service industries provide the lion’s share of influence on per capita income growth. The estimate for the establishment density variable did not show a statistically significant effect on employment, yet it appears to influence per-capita income growth positively. It is possible that, on average, marginally increasing the establishment density could increase efficiencies through competition and agglomeration spillovers. Thus, the productivity gains would be likely to go to employees by driving up incomes, but would not affect employment growth.

I control for the supply of inputs and human capital by including the percentage of the population that is of young working age, the percent of population with a bachelor’s degree or higher, and the unemployment rate. Higher levels of both young and college-educated population are associated with increased levels of employment and income growth. However, the unemployment rate in 1990 does not seem to influence either income or employment growth over time. Finally, the remaining demographic variables control for the ability of the labor market to produce goods as well as earn income. The results show that urban counties and counties with a higher percentage of white residents tend to have higher growth rates of employment and per-capita income.

2.7 Employment growth robustness check

Regional economic theory highlights an important consideration for employment growth that is understated in the theoretical model displayed in section 2.3. The theory is based on the interplay between households choosing to locate in pursuit of utility, and firms based on profits. It suggests that population and employment are interdependent due to the mobility of people and firms. This results in the classic problem of whether "people follow jobs" or "jobs follow people." To account for this regional theory suggests using a simultaneous equations model for population and employment. In practice, the regional adjustment framework of Carlino and Mills (1987) has become a workhorse in the regional economic growth literature, and has generated a variety of extensions (see Rickman (2010) for an overview of the approach, critique, and extensions.)

I estimate reduced-form versions of the growth models in equations (2.9) and (2.10). Nonetheless, since regional economic theory suggests that employment growth and population growth are simultaneously determined, I also use a two-stage least squares GMM estimation procedure to estimate employment growth. Furthermore, the estimation results allow the disturbances to be spatially autocorrelated following equation (2.11). Regional economic theory suggests that valid instruments for population growth should be based on the household utility and subsequent location decisions, as well as on population growth factors. I use several instrumental variables to identify population growth in the employment growth equation. In particular the variables used to identify population growth are the number of births per 1,000 residents, a natural amenity index, median monthly homeownership costs, the percent of residents in poverty, per-capita local expenditures on fire and police, and per-capita local expenditures on education. More details about each of these variables can be found in table 2.2.

Table 2.7 presents results that take into account the potential simultaneity between employment growth and income growth. The first column shows OLS estimates of the first stage (population growth) regression, while the second column displays results using a 2SLS GMM estimation procedure. Both sets of results in table 2.5 use the employment shares by business-size categories as explanatory variables. First, the results in column 1 show that several of the instrumental variables are statistically significant at the 1% level in explaining population growth. In addition, since there are more instruments than needed to identify the employment growth equation, I use a statistic following Hausman (1983), to test the validity of the instruments. The results of the test suggest that the employment equation is appropriately identified, because the orthogonality assumption cannot be rejected[9].

The estimated coefficient for population growth variable in column 2, which uses a 2SLS estimation procedure, is positive and statistically significant. This result is consistent with economic theory for regional adjustment models. Controlling for simultaneity between

---

[9] This Hausman test statistic is obtained as NRu^2, where N is the sample size and Ru^2 is the usual R-squared of the regression of residuals from the second-stage estimation on all included and excluded instruments. The statistic has a limiting chi-squared distribution with degree of freedom equal to the number of over-identifying restrictions, under the assumed specification of the model.

population and employment growth does not change the estimated coefficients for the control variables in a meaningful way. The pattern for the employment share variables is consistent with those shown in table 2.3. The positive effect on employment growth of marginally increasing an employment share class decreases, as firm sizes get larger. However, the magnitudes of the employment share coefficients for the 2SLS model are smaller than the reduced form model.

2.8 Conclusion

The effect of the business size distribution on per-capita income and employment growth is estimated using cross-sectional growth models with U.S. county data from 1990-2000. The business size distribution is measured in several ways. First, the distribution is measured as the share of employees across nine establishment size categories that range from micro firms (1-4 employees) to large firms (1000+ employees). Second, I use several indices to parsimoniously include indices into the empirical model. I use three indices (similar to a Gini coefficient) are used that measure deviation from a specified distribution. I also use the Atkinson index of inequality. The results show that the business size distribution has a significant impact on county level growth patterns. Most notably, the employment shares in small firms increase employment growth, but have no influence on income growth.

As one might expect, the implications for policy vary depending on the objectives of the policy maker. The current emphasis on entrepreneurship seems well placed for national policy, especially in view of the high level of unemployment. Programs to foster more business start-ups, or to encourage succession from non-employer to small employer, may create more jobs than strategies that focus on enhancing the performance of larger-scale enterprises. Policies to enhance small-scale enterprises should be matched with increased emphasis on evaluating the effectiveness of the various competing techniques and programs designed to foster start-ups or growth of early-stage firms. These activities can increase employment, thereby reducing the costs of income support programs, and set the stage for the next phase of the country's growth.

After national employment recovers, it may be more effective to shift the emphasis of policy away from entrepreneurship, toward fostering greater efficiency or international market share in large-scale enterprises, to meet policy goals of income growth. Income growth in this phase can help supply start-up capital for new firms when the national economy enters another period of higher unemployment. I should emphasize that I am articulating the need for flexibility in emphasis, rather than an exclusive focus on one set of policies over the other. If techniques for assisting small firms are ignored in times of low unemployment, it will be difficult to reconstitute them when they are needed. Similarly, completely ignoring the needs of larger firms in times of high unemployment is likely to deepen an economic downturn.

Finally, this study represents a national perspective. States or regions may find that their structure varies substantially from the national allocation of employment across business size classes. A local structure that is different from the nation will likely lead to different policy conclusions. As noted earlier, the prior work by Loveridge and Nizalov (2007) indicated that both income and employment would be well served through an increased emphasis on smaller businesses. This conclusion flows logically from Michigan's history of great success in large-scale manufacturing. Large-scale manufacturing drew talent away from other types of enterprises. As employment in large-scale manufacturing declined, counties with a stronger entrepreneurial base were likely able to recover employment and income more quickly than counties whose employment base was so tied to big companies. State and local policy makers are therefore advised to replicate this analysis for their own region, before adjusting their economic development strategies to take full advantage of my contribution.

Table 2.1. Descriptive Statistics of the Business Size Distribution

Size Groups (Employees)	Mean	Std. Dev.	Min	Max	Source*
A. Establishment Share (%)					
1-4	59.67	6.68	40.91	100.00	CBP
5-9	19.83	3.06	0.00	50.00	CBP
10-19	10.82	2.55	0.00	33.33	CBP
20-49	6.17	2.20	0.00	13.33	CBP
50-99	2.00	1.02	0.00	11.11	CBP
100-249	1.07	0.74	0.00	8.11	CBP
250-499	0.29	0.32	0.00	2.78	CBP
500-999	0.11	0.18	0.00	2.44	CBP
1,000 +	0.04	0.09	0.00	0.87	CBP
A. Establishment Share (%)					
1-4	6.67	4.31	1.45	61.54	CBP and CCDB
5-9	10.37	4.74	0.00	62.50	CBP and CCDB
10-19	10.88	4.30	0.00	62.50	CBP and CCDB
20-49	11.63	4.04	0.00	44.44	CBP and CCDB
50-99	9.08	4.50	0.00	53.19	CBP and CCDB
100-249	8.97	5.67	0.00	42.02	CBP and CCDB
250-499	5.60	6.02	0.00	38.52	CBP and CCDB
500-999	3.91	6.12	0.00	68.92	CBP and CCDB
1,000 +	2.63	5.31	0.00	48.16	CBP and CCDB
C. Size Distribution Index					
Uniform Index	24.08	8.26	12.75	75.19	CBP and CCDB
High Uniform Index	26.92	6.78	17.28	66.96	CBP and CCDB
Low Uniform Index	21.90	5.96	12.67	68.84	CBP and CCDB
Atkinson Index (.5)	15.87	44.18	0.29	35.02	CBP and CCDB
Atkinson Index (2)	77.03	32.99	5.15	98.88	CBP and CCDB

*CBP is the County Business Patterns and CCDB is the County and City Data Book.

Table 2.2. Descriptive Statistics

Variable Code	Variable Description	Mean	Std. Dev.	Source
GREMP	Employment Growth 1990-2000	17.54	16.49	CCDB
GRINC	Per Capita Income Growth 1990-2000	16.09	9.79	CCDB
GRPOP	Population Growth 1990-2000	9.65	13.29	CCDB
AG EMP	Percent Employment Agriculture	12.60	10.61	CCDB
MAN EMP	Percent Employment in Manufacturing	14.61	10.65	CCDB
SERVICE EMP	Percent Employment in Services	20.31	6.86	CCDB
TRADE EMP	Percent Employment in Trade	18.65	4.70	CCDB
EST DENSITY	Establishment Density	5.57	67.65	CBP
POP 25-44	Percent Population 25-44	29.26	3.40	CCDB
COLL PLUS	Percent Population College Degree +	13.37	6.40	CCDB
UNEMP	Unemployment Rate	6.18	2.93	CCDB
PROP TAX	Property Tax Per Capita	544.38	403.87	CCDB
TOTAL GOVT EXP	Total Government Expenditure Per Capita	5299.87	2327.06	CCDB
POP DENSITY	Population Density (per square mile)	202.56	1429.71	CCDB
URBAN	Percent Urban	26.33	44.04	ERS
POP WHITE	Percent Population White	87.56	15.33	CCDB
POP 65+	Percent Population Over 65	14.98	4.32	CCDB
POP 5-17	Percent Population 5-17	19.79	2.67	CCDB
BIRTHS	Births per 1,000 Residents	14.58	3.00	CCDB
NAT AMENITY	Natural Amenity Index	0.04	2.29	ERS
HOMEOWN COST	Median Monthly Homeownership Cost	236.03	96.30	CCDB
POVERTY	Percent Population in Poverty	16.31	7.76	CCDB
FIRE POL EXP	Per Capita Expenditure Fire and Police	103.55	68.98	CCDB
EDUC EXP	Per Capita Expenditure Education	945.16	321.34	CCDB
LN EMP	Log Employment	9.37	1.42	CCDB
LN INC	Log Per Capita Income	9.60	0.21	CCDB

*CBP is the County Business Patterns: U.S. Census, CCDB is the County and City Data Book: U.S. Census and ERS is the Economic Research Service: USDA

Table 2.3. Regression Results: Employment Share

VARIABLES	(1) Employment Growth	(2) Employment Growth	(3) Income Growth	(4) Income Growth
SHARE 1-4	0.2177**	0.623***	0.0079	-0.010
	(0.108)	(0.108)	(0.063)	(0.066)
SHARE 5-9	0.127	0.380***	-0.1211**	0.020
	(0.095)	(0.090)	(0.056)	(0.054)
SHARE 10-19	0.125	0.231***	0.1201**	0.208***
	(0.084)	(0.082)	(0.049)	(0.050)
SHARE 20-49	0.472***	0.293***	0.1578***	0.125***
	(0.077)	(0.075)	(0.045)	(0.045)
SHARE 50-99	0.324***	0.248***	0.1650***	0.105***
	(0.067)	(0.063)	(0.039)	(0.038)
SHARE 100-249	0.302***	0.225***	0.2003***	0.133***
	(0.062)	(0.058)	(0.037)	(0.035)
SHARE 250-499	0.136**	0.157***	0.1107***	0.079**
	(0.059)	(0.054)	(0.034)	(0.033)
SHARE 500-999	0.148***	0.139***	0.0413	0.027
	(0.056)	(0.052)	(0.033)	(0.031)
AG EMP		0.082*		0.010
		(0.049)		(0.030)
MAN EMP		0.098**		0.104***
		(0.042)		(0.026)
SERVICE EMP		0.267***		0.141***
		(0.050)		(0.031)
TRADE EMP		0.396***		-0.001
		(0.074)		(0.045)
EST DENSITY		0.009		0.022***
		(0.007)		(0.004)
POP 25-44		1.148***		1.043***
		(0.167)		(0.102)
COLL PLUS		0.612***		0.488***
		(0.078)		(0.047)
UNEMP		0.130		0.009
		(0.126)		(0.077)
PROP TAX		-0.007***		-0.001
		(0.001)		(0.001)
TOTAL GOVT EXP		-0.001***		-0.001***
		(0.000)		(0.000)

Table 2.3. (cont'd)

	(1) Employment Growth	(2) Employment Growth	(3) Income Growth	(4) Income Growth
POP DENSITY		-0.001*		-0.001***
		(0.000)		(0.000)
URBAN		0.039***		0.020***
		(0.008)		(0.005)
POP WHITE		0.108***		0.066***
		(0.026)		(0.016)
POP 65+		-0.129		0.532***
		(0.148)		(0.091)
POP 5-17		0.499***		0.282***
		(0.167)		(0.102)
LN EMP 1990		-0.007***		0.002
		(0.002)		(0.001)
LN INC 1990		-0.677***		-1.459***
		(0.153)		(0.094)
Constant	1.0256	-58.769***	10.091***	-27.214***
	(2.721)	(8.960)	(1.597)	(5.465)
Rho	.7041***	.738***	.514***	.605***
	(0.040)	(0.045)	(0.043)	(0.047)
Observations	3,046	3,046	3,046	3,046

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.4. Regression Results for Gini Coefficient Indices: Employment Growth

VARIABLES	(1) Employment Growth	(2) Employment Growth	(3) Employment Growth
UNIFORM INDEX	0.337**		
	(0.151)		
UNIFORM INDEX SQ	-0.003		
	(0.002)		
HIGH UNIFORM INDEX		0.544**	
		(0.267)	
HIGH UNIFORM INDEX SQ		-0.006	
		(0.004)	
LOW UNIFORM INDEX			-0.608***
			(0.179)
LOW UNIFORM INDEX SQ			0.009***
			(0.003)
Constant	-50.575***	-54.015***	-34.601***
	(8.982)	(9.603)	(9.126)
Rho	.633***	.633***	.625***
	(0.045)	(0.044)	(0.045)
Full Set of Covariate	Yes	Yes	Yes
Observations	3,046	3,046	3,046

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.5. Regression Results for Gini Coefficient Indices: Income Growth

VARIABLES	(1) Income Growth	(2) Income Growth	(3) Income Growth
UNIFORM INDEX	0.480***		
	(0.090)		
UNIFORM INDEX SQ	-0.009***		
	(0.001)		
HIGH UNIFORM INDEX		0.742***	
		(0.160)	
HIGH UNIFORM INDEX SQ		-0.01***	
		(0.002)	
LOW UNIFORM INDEX			0.119
			(0.107)
LOW UNIFORM INDEX SQ			-0.006***
			(0.002)
Constant	-26.009***	-31.06***	-18.957***
	(5.392)	(5.783)	(5.472)
Rho	.708***	.607***	.608***
	(0.047)	(0.047)	(0.047)
Full Set of Covariate	Yes	Yes	Yes
Observations	3,046	3,046	3,046

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.6. Regression Results: Atkinson Index

	(1)	(2)	(3)	(4)
	Atkinson .5		Atkinson 2	
	Income Growth	Employment Growth	Income Growth	Employment Growth
ATKINSON	-0.415	0.681***	0.352***	0.784***
	(0.631)	(0.103)	(0.062)	(0.103)
AG EMP	0.007	0.074	-0.013	0.164***
	(0.030)	(0.050)	(0.028)	(0.046)
MAN EMP	0.066***	0.008	0.094***	-0.013
	(0.023)	(0.038)	(0.023)	(0.037)
SERVICE EMP	0.134***	0.258***	0.146***	0.247***
	(0.031)	(0.051)	(0.030)	(0.051)
TRADE EMP	0.030	0.521***	0.051	0.446***
	(0.045)	(0.074)	(0.044)	(0.072)
EST DENSITY	0.023***	0.010	0.021***	0.008
	(0.004)	(0.007)	(0.004)	(0.007)
POP 25-44	1.049***	1.204***	1.025***	1.221***
	(0.102)	(0.167)	(0.101)	(0.167)
COLL PLUS	0.4820***	0.639***	0.503***	0.679***
	(0.048)	(0.078)	(0.048)	(0.078)
UNEMP	0.0236	0.231*	0.030	0.210*
	(0.077)	(0.126)	(0.077)	(0.125)
PROP TAX	-0.001**	-0.007***	-0.001**	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
TOTAL GOVT EXP	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
POP DENSITY	-0.001***	-0.001**	-0.001***	-0.008*
	(0.000)	(0.000)	(0.000)	(0.000)
URBAN	0.019***	0.0401***	0.026***	0.049***
	(0.005)	(0.008)	(0.005)	(0.008)
POP WHITE	0.068***	0.120***	0.063***	0.108***
	(0.016)	(0.026)	(0.016)	(0.026)
POP 65+	0.555***	-0.029	0.524***	-0.019
	(0.090)	(0.148)	(0.090)	(0.148)
POP 5-17	0.328***	0.536***	0.288***	0.492***
	(0.102)	(0.167)	(0.102)	(0.166)
LN EMP 1990	0.001	-0.008***	0.002*	-0.006***
	(0.001)	(0.002)	(0.001)	(0.002)
LN INC 1990	-1.4770***	-0.678***	-1.416***	-0.701***
	(0.094)	(0.155)	(0.094)	(0.153)

Table 2.6. (cont'd)

	(1)	(2)	(3)	(4)
	Atkinson .5		Atkinson 2	
	Income Growth	Employment Growth	Income Growth	Employment Growth
Constant	-21.1746***	-49.045***	-24.09***	-50.235***
	(5.309)	(8.713)	(5.288)	(8.699)
Rho	1.095***	1.037***	1.099***	1.020***
	(0.046)	(0.045)	(0.046)	(0.045)
Observations	3,046	3,046	3,046	3,046

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.7. Instrumental Variables Regression Results: Employment Growth

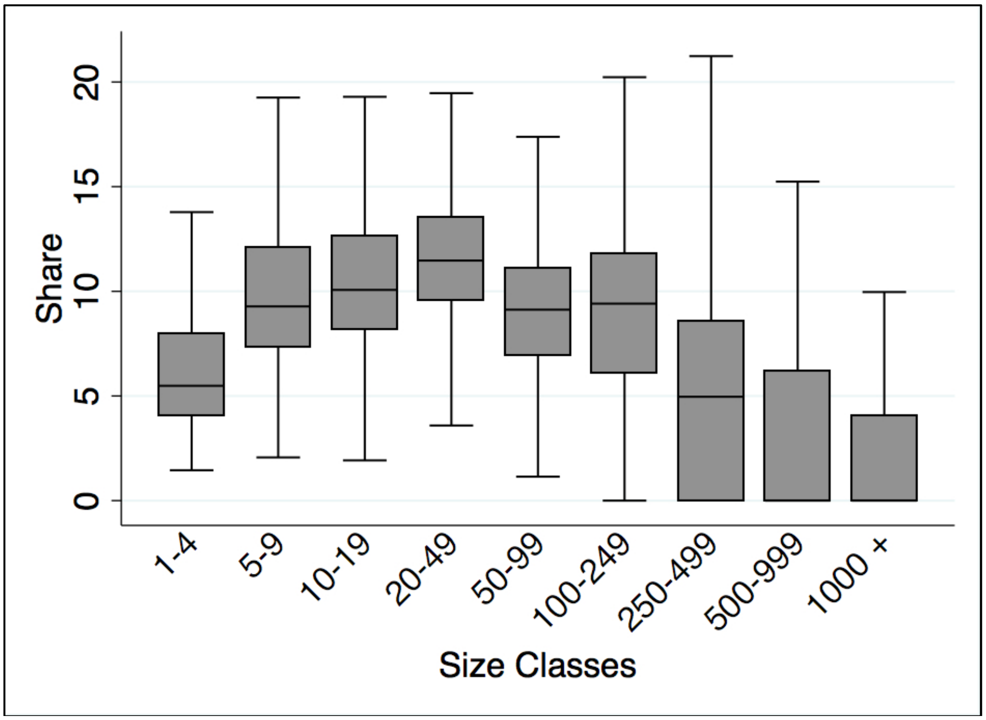
VARIABLES	(1) Population Growth: First Stage	(2) Employment Growth: 2SLS
SHARE 1-4	0.279***	0.379***
	(0.083)	(0.084)
SHARE 5-9	0.087	0.283***
	(0.070)	(0.070)
SHARE 10-19	0.069	0.129**
	(0.063)	(0.064)
SHARE 20-49	0.142**	0.166***
	(0.058)	(0.058)
SHARE 50-99	0.085*	0.157***
	(0.049)	(0.050)
SHARE 100-249	0.164***	0.128***
	(0.045)	(0.046)
SHARE 250-499	0.074*	0.114***
	(0.042)	(0.042)
SHARE 500-999	0.023	0.099**
	(0.041)	(0.041)
AG EMP	0.213***	0.071**
	(0.035)	(0.035)
MAN EMP	0.146***	0.035
	(0.030)	(0.031)
SERVICE EMP	0.112***	0.163***
	(0.039)	(0.040)
TRADE EMP	0.093*	0.405***
	(0.056)	(0.056)
EST DENSITY	0.019***	0.007
	(0.006)	(0.005)
POP 25-44	0.384***	0.563***
	(0.124)	(0.128)
COLL PLUS	0.075	0.485***
	(0.055)	(0.056)
UNEMP	-0.027	0.034
	(0.088)	(0.086)
PROP TAX	-0.005***	-0.004***
	(0.001)	(0.001)
TOTAL GOVT EXP	-0.001***	-0.001***
	0.000	0.000
POP DENSITY	-0.001***	(0.001)
	0.000	0.000
URBAN	0.016***	0.022***
	(0.006)	(0.006)

Table 2.7. (cont'd)

	(1) Population Growth: First Stage	(2) Employment Growth: 2SLS
POP WHITE	-0.002	0.093***
	(0.019)	(0.016)
POP 65+	-0.409***	0.244**
	(0.110)	(0.110)
POP 5-17	-0.392***	0.715***
	(0.116)	(0.115)
LN EMP 1990	-0.010***	-0.003*
	(0.001)	(0.001)
LN INC 1990	-0.761***	-0.614***
	(0.119)	(0.108)
BIRTHS	0.623***	
	(0.098)	
NAT AMENITY	1.791***	
	(0.101)	
HOMEOWN COST	0.033***	
	(0.005)	
POVERTY	-0.318***	
	(0.050)	
FIRE POL EXP	0.007*	
	(0.004)	
EDUC EXP	0.001	
	(0.001)	
GR POP		0.705***
		(0.042)
Constant	1.182	10.092***
	(7.266)	(1.597)
Rho		.524***
		(0.043)
Observations	3,046	3,046

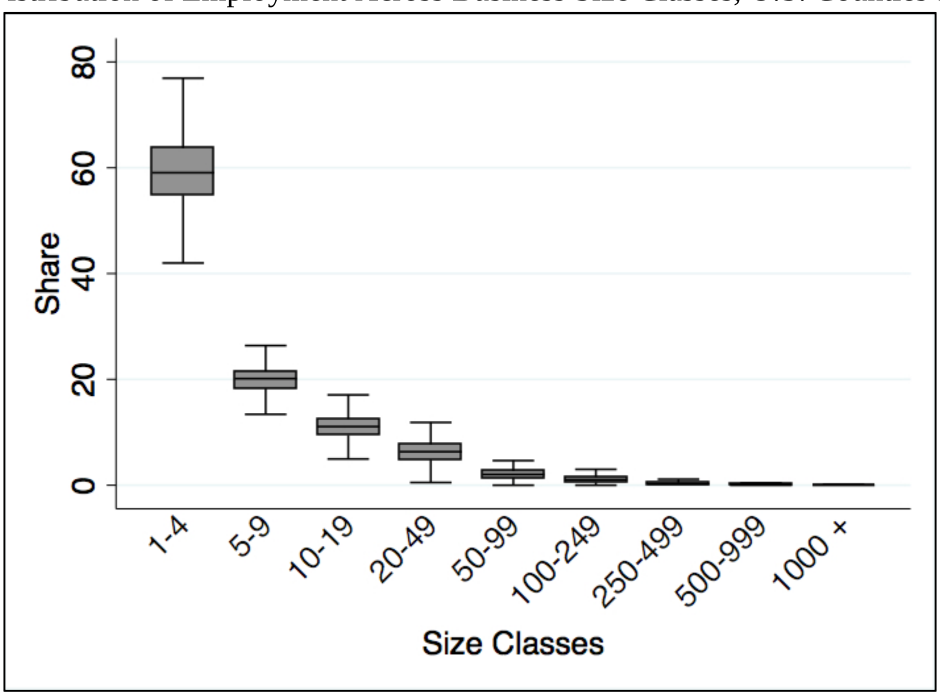
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 2.1. Distribution of Establishments by Employee Size Categories, U.S. Counties 1990



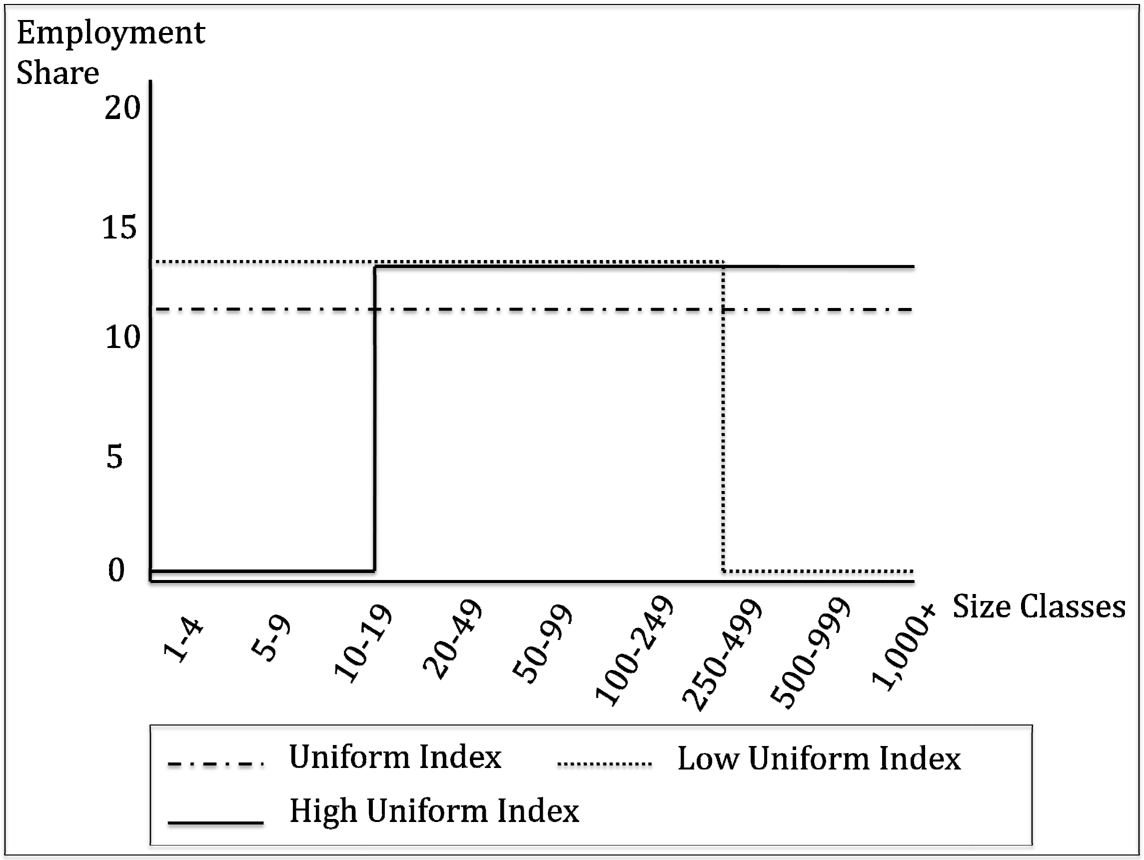
* lower adjacent value, 25th percentile, median, 75th percentile, upper adjacent value

Figure 2.2. Distribution of Employment Across Business Size Classes, U.S. Counties 1990



* lower adjacent value, 25th percentile, median, 75th percentile, upper adjacent value

Figure 2.3. Distribution of Indices



APPENDIX

APPENDIX

Table 2.A1. Reduced Form Regression Results for Indices: Employment Growth

VARIABLES	(1) Employment Growth	(2) Employment Growth	(3) Employment Growth
INDEX	0.337**		
	(0.151)		
INDEX SQ	-0.003		
	(0.002)		
HIGH INDEX		0.544**	
		(0.267)	
HIGH INDEX SQ		-0.006	
		(0.004)	
LOW INDEX			-0.608***
			(0.179)
LOW INDEX SQ			0.009***
			(0.003)
AG EMP	0.150***	0.157***	0.201***
	(0.049)	(0.049)	(0.048)
MAN EMP	-0.052	-0.044	-0.042
	(0.037)	(0.038)	(0.038)
SERVICE EMP	0.237***	0.233***	0.222***
	(0.051)	(0.051)	(0.051)
TRADE EMP	0.478***	0.479***	0.369***
	(0.075)	(0.076)	(0.075)
EST DENSITY	0.011*	0.012*	0.011
	(0.007)	(0.007)	(0.007)
POP 25-44	1.237***	1.237***	1.246***
	(0.168)	(0.168)	(0.168)
COLL PLUS	0.633***	0.631***	0.634***
	(0.078)	(0.078)	(0.078)
UNEMP	0.220*	0.209*	0.170
	(0.127)	(0.126)	(0.126)
PROP TAX	-0.006***	-0.006***	-0.005***
	(0.001)	(0.001)	(0.001)
TOTAL GOVT EXP	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)

Table 2.A1. (cont'd)

	(1) Employment Growth	(2) Employment Growth	(3) Employment Growth
POP DENSITY	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)
URBAN	0.037***	0.037***	0.036***
	(0.008)	(0.008)	(0.008)
POP WHITE	0.122***	0.120***	0.115***
	(0.027)	(0.027)	(0.026)
POP 65+	0.009	-0.011	0.031
	(0.150)	(0.150)	(0.149)
POP 5-17	0.551***	0.543***	0.569***
	(0.168)	(0.168)	(0.168)
LN EMP 1990	-0.008***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)
LN INC 1990	-0.778***	-0.781***	-0.791***
	(0.154)	(0.154)	(0.154)
Constant	-50.575***	-54.015***	-34.601***
	(8.982)	(9.603)	(9.126)
rho	.633***	.633***	.625***
	(0.045)	(0.044)	(0.045)
Observations	3,046	3,046	3,046

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.A2. Reduced Form Regression Results for Indices: Income Growth

VARIABLES	(1) Income Growth	(2) Income Growth	(3) Income Growth
INDEX	0.480***		
	(0.090)		
INDEX SQ	-0.009***		
	(0.001)		
HIGH INDEX		0.741***	
		(0.160)	
HIGH INDEX SQ		-0.013***	
		(0.002)	
LOW INDEX			0.119
			(0.107)
LOW INDEX SQ			-0.006***
			(0.002)
AG EMP	0.045	0.035	0.055*
	(0.030)	(0.030)	(0.029)
MAN EMP	0.073***	0.079***	0.091***
	(0.022)	(0.023)	(0.023)
SERVICE EMP	0.139***	0.140***	0.140***
	(0.030)	(0.030)	(0.030)
TRADE EMP	0.017	0.025	-0.027
	(0.045)	(0.045)	(0.045)
EST DENSITY	0.023***	0.023***	0.023***
	(0.004)	(0.004)	(0.004)
POP 25-44	1.050***	1.051***	1.048***
	(0.101)	(0.101)	(0.101)
COLL PLUS	0.490***	0.488***	0.490***
	(0.047)	(0.047)	(0.047)
UNEMP	0.037	0.025	0.012
	(0.076)	(0.077)	(0.076)
PROP TAX	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
TOTAL GOVT EXP	-0.001***	-0.001***	-0.004***
	(0.000)	(0.000)	(0.000)
POP DENSITY	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
URBAN	0.022***	0.021***	0.021***
	(0.005)	(0.005)	(0.005)
POP WHITE	0.067***	0.066***	0.065***
	(0.016)	(0.016)	(0.016)

Table 2.A2. (cont'd)

	(1) Income Growth	(2) Income Growth	(3) Income Growth
POP 65+	0.494***	0.507***	0.489***
	(0.090)	(0.091)	(0.090)
POP 5-17	0.270***	0.290***	0.271***
	(0.102)	(0.102)	(0.101)
LN EMP 1990	0.001	0.002	0.002
	(0.001)	(0.001)	(0.001)
LN INC 1990	-1.495***	-1.488***	-1.502***
	(0.093)	(0.093)	(0.093)
Constant	-26.009***	-31.067***	-18.957***
	(5.392)	(5.783)	(5.472)
rho	.7075***	.607***	.608***
	(0.047)	(0.047)	(0.047)
Observations	3,046	3,046	3,046

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

REREFENCES

REREFENCES

Allanach, C., and S. Loveridge. 1998. “An assessment of maximum-training business visitation programs.” *Economic Development Quarterly*. 12(2), 125-136.

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

Anselin, L., and D. A. Griffith. 1988. "Do Spatial Effects Really Matter in Regression-Analysis." *Papers of the Regional Science Association*, 65, 11-34.

Aquilina, M., R. Klump and C. Pietrobelli. 2006. "Factor Substitution, Average Firm Size and Economic Growth." *Small Business Economics*, 26(3), 203-214.

Aronsson, T., J. Lundberg and M. Wikstrom. 2001. "Regional Income Growth and Net Migration in Sweden, 1970-1995." *Regional Studies*, 35(9), 823-830.

Atkinson, A. B. 1970. “Measurement of Inequality” *Journal of Economic Theory*, 2(3), 244-263.

Barro, R. J., and X. Sala-i-martin. 1991. "Convergence across States and Regions." *Brookings Papers on Economic Activity*, (1), 107-182.

Barro, R. J., and X. Sala-i-martin. 1992. "Convergence." *Journal of Political Economy*, 100(2), 223-251.

Blair, J. P., and R. Premus. 1987. “Major factors in industrial location: a review.” *Economic Development Quarterly*, 1(1), 72-85.

Birch, D.L. 1979. *The Job Generation Process: Final Report to Economic Development Administration*, MIT Program on Neighborhood and Regional Change, Cambridge, MA.

Birch, D.L. 1981. “Who Creates Jobs?” *Public Interest* 65(3), 3-14.

Birch, D.L. 1987. *Job Creation in America: How Our Smallest Companies Put the Most People to Work.* New York: Free Press.

Boarnet, M. G. 1994. "The Monocentric Model and Employment Location." *Journal of Urban Economics*, 36(1), 79-97.

Boarnet, M. G., S. Chalermpong and E. Geho. 2005. "Specification Issues in Models of Population and Employment Growth." *Papers in Regional Science*, 84(1), 21-46.

Bradshaw, T. and E. Blakely. 1999. “What are third-wave economic development strategies? From incentives to industrial policy.” *Economic Development Quarterly* 13 (3), 229–244.

Carlino, G. R. 1982. “Manufacturing agglomeration economies as returns to scale: a production function approach. “ *Papers of the Regional Science Association*, 50(1), 95-108.

Carlino, G. A. and E. S. Mills. 1987. "The Determinants of County Growth." *Journal of Regional Science*, 27(1), 39-54.

Decancq, K. and A. Decoster. 2009. “The Evolution of World Inequality in Well-being.” *World Development*, 37(1), 11-25.

Deller, S. C. T. H. Tsai D. W. Marcouiller, and D. B. K. English. 2001. "The Role of Amenities and Quality of Life in Rural Economic Growth." *American Journal of Agricultural Economics*, 83(2), 352-365.

Deller, S. and J. C. McConnon. 2009. "Microenterprises and Economic Growth: A Panel Study of the US States 1977-1997." *Applied Economics Letters*, 16(13), 1307-1312.

Devereux, M. P., R. Griffith and H. Simpson. 2007. “Firm location decisions, regional grants, and agglomeration externalities.” *Journal of Public Economics*. 91(3-4), 413-435.

Drabenstott, M. 2006. “Rethinking federal policy for regional economic development.” *Federal Reserve Bank of Kansas City Economic Review*. QI, 115-142.

Drukker, D. M., P. Egger, and I.R. Prucha. 2010. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. Technical report, Department of Economics, University of Maryland

Duffy, N. E. 1994. "The Determinants of State Manufacturing Growth-Rates - a 2-Digit-Level Analysis." *Journal of Regional Science*, 34(2),137-162.

Duffy-Deno, K. T. 1998. "The Effect of Federal Wilderness on County Growth in the Intermountain Western United States." *Journal of Regional Science*, 38(1), 109-136.

Edmiston, K. D. 2004. "The Net Effects of Large Plant Locations and Expansions on County Employment." *Journal of Regional Science*, 44(2), 289-319.

Fotopoulos, G. (Forthcoming). “Nonlinearities in regional economic growth and convergence: the role of entrepreneurship in the European Union.” *Annals of Regional Science*.

Glaeser, E. L.; J. A. Scheinkman and A. Shleifer. 1995. "Economic-Growth in a Cross-Section of Cities." *Journal of Monetary Economics*, 36(1), 117-143.

Graves, P. E. and P. D. Linneman. 1979. "Household Migration - Theoretical and Empirical Results." *Journal of Urban Economics*, 6(3), 383-404.

Graves, P. E. and P. R. Mueser. 1993. "The Role of Equilibrium and Disequilibrium in Modeling Regional Growth and Decline - a Critical Reassessment." *Journal of Regional Science*, 33(1), 69-84.

Hammond, G. W. and M. S. Tosun. 2011. "The Impact of Local Decentralization on Economic Growth: Evidence from Us Counties." *Journal of Regional Science*, 51(1), 47-64.

Hansen, E. R. 1990. “Agglomeration economies and industrial decentralization: the wage-productivity trade-offs," *Journal of Urban Economics*, 28(2), 140-159.

Hausman J. 1983. Specification and estimation of simultaneous equations models. In: Griliches Z, Intriligator M (eds) *Handbook of econometrics*. North Holland, Amsterdam

Henry, M. S.; B. Schmitt; K. Kristensen; D. L. Barkley and S. M. Bao. 1999. "Extending Carlino-Mills Models to Examine Urban Size and Growth Impacts on Proximate Rural Areas." *Growth and Change*, 30(4), 526-548.

Higgins, M. J.; D. Levy and A. T. Young. 2006. "Growth and Convergence across the United States: Evidence from County-Level Data." *Review of Economics and Statistics*, 88(4), 671-681.

Higgins, M. J.; A. T. Young and D. Levy. 2009. "Federal, State, and Local Governments: Evaluating Their Separate Roles in Us Growth." *Public Choice*, 139(3-4), 493-507.

Hodge, T. R. 2011. “Inventory of State Job Creation and Job Retention Incentives,” 2010. *Policy Brief*. East Lansing: North Central Regional Center for Rural Development, Michigan State University.

Krugman, P. (1991). *Geography and Trade.* Leuven University Press, Leuven, Belgium.

LeSage, J. and R.K. Pace. 2009. *Introduction to Spatial Econometrics.* New York: CRC Press.

Loveridge, S. 1996. "On the Continuing Popularity of Industrial Recruitment." *Economic Development Quarterly*, 10(2), 151-158.

Loveridge, S. and D. Nizalov. 2007. "Operationalizing the Entrepreneurial Pipeline Theory: An Empirical Assessment of the Optimal Size Distribution of Local Firms." *Economic Development Quarterly*, 21(3), 244-262.

Mills, E. S. and R. Price. 1984. "Metropolitan Suburbanization and Central City Problems." *Journal of Urban Economics*, 15(1), 1-17.

Morse, G. (ed.). (1990). *The Retention and Expansion of Existing Businesses.* Ames, IA: Iowa State University Press.

Mueser, P. R. and P. E. Graves. 1995. "Examining the Role of Economic-Opportunity and Amenities in Explaining Population Redistribution." *Journal of Urban Economics*, 37(2), 176-200.

Olfert, M. R. and M.D. Partridge. 2010. “Best Practices in Twenty-First-Century Rural Development and Policy.” *Growth and Change*, 41(2), 147–164.

Parker, S.C. (2004). *The economics of self-employment and entrepreneurship*. Cambridge University Press, Cambridge, MA.

Rickman, D. S. 2010. “Modern Macroeconomics and Regional Economic Modeling.” *Journal of Regional Science*, 50(1), 23-41.

Robbins, D. K., L. J. Pantuosco; D. F. Parker and B. K. Fuller. 2000. "An Empirical Assessment of the Contribution of Small Business Employment to Us State Economic Performance." *Small Business Economics*, 15(4), 293-302.

Rupasingha, A. and J. B. Chilton. 2009. "Religious Adherence and County Economic Growth in the US." *Journal of Economic Behavior & Organization*, 72(1), 438-450.

Schumpeter, J.A. 1942. *Capitalism, Socialism and Democracy.* Harper & Row, New York.

Schumpeter, J.A. 1961. *The Theory of Economic Development.* Harvard University Press, Cambridge, MA.

Shaffer, R., S. Deller and D. Marcouiller. 2006. "Rethinking Community Economic Development." *Economic Development Quarterly*, 20(1), 59-74.

Shaffer, S. 2006. "Establishment Size and Local Employment Growth." *Small Business Economics*, 26(5), 439-454.

Steinnes, D. N. and W. D. Fisher. 1974. "Econometric Model of Intraurban Location." *Journal of Regional Science*, 14(1), 65-80.

Von Bargen, P., D. Freedman and E.R. Pages. 2003. "The Rise of the Entrepreneurial Society." *Economic Development Quarterly*, 17(4), 315-324.

Walzer, N. (ed.). 2007. *Entrepreneurship and Local Economic Development*. Lexington Books, Lanham, MD.

CHAPTER 3: Growing Fuel and Jobs? The Effect of Ethanol Plants on Local Employment

3.1 Introduction

Ethanol production has risen dramatically in the United States since the early 2000s (figure 3.1). Several key drivers have encouraged promotion of the biofuel industry. Charles et al. (2007) argue that the development of biofuels has come primarily from policy considerations for rural development, energy independence, climate change, and its potential as a renewable resource. Extensive government intervention at both the state (Cotti and Skidmore, 2007) and federal level[10] (Low and Isserman, 2009) has been used as an impetus to increasing ethanol production. Consequently, ethanol production in the U.S. is due primarily to social and political interests and not market forces (Cascone, 2007). Ethanol policies and subsequent production have a wide array of effects. The academic literature has begun to quantify the effect of ethanol on consequences ranging from local implications on residential home values (Hodge, 2011) to global issues such as climate change (Hahn and Cecot, 2009). In particular, policy makers (Obama,[11] 2010; Yacobucci, 2007), political pundits (Buchanan,[12] 1999), and ethanol producers

[10] The original motivation to encourage ethanol was the Clean Air Act of 1963. The purpose was to use ethanol as an oxygenate in gasoline to reduce carbon monoxide emissions. Other major federal policies include the Energy Policy Act of 2005 mandating the use of 7.5 billion gallons by 2012, and the Energy Independence and Security Act of 2007, which increased the mandate to 36 billion gallons of renewable fuels by 2022.

[11] President Obama (2010) stated: "So there shouldn't be any doubt that renewable, homegrown fuels are a key part of our strategy for a clean-energy future - a future of new industries, new jobs in towns like Macon, MO and new independence."
Accessed at <http://www.whitehouse.gov/the-press-office/remarks-president-barack-obama-poet-biorefining-macon-missouri>

[12] Conservative political commentator Pat Buchanan (1999) stated " . . . just as I support the independence of the family farm, I support a policy of U.S. energy independence that includes a strong stand for ethanol. This industry creates 40,000 jobs, ads $12 billion in net farm income each year . . . ''

(Poet,[13] 2012) alike often tout the local development benefits that an ethanol plant has on a community. Several states even provide incentives to encourage locally owned facilities, because of the perceived local economic benefit (Hueth and Walker, 2008). However, the validity of local economic development claims have gone relatively unexplored in the academic literature.

In this paper, I use county-level data from 12 Corn Belt states to examine the effect of an ethanol plant on local employment. I use the timing of ethanol plant construction and production dates to identify the local economic effect of an ethanol plant locating in a community. Industry groups collect the location and production capacity of many ethanol plants (e.g. Ethanol Producers Magazine, 2012). However, the date (month and year) of construction and production is not systematically collected by anyone. I use information collected from multiple sources to construct a unique county-level panel data set of ethanol plants in 12 Corn Belt states. Using a difference-in-difference identification strategy, I estimate the average effect of an ethanol plant on employment levels. Furthermore, knowing the beginning dates of construction and production allows me to estimate the dynamic effect of an ethanol plant on the local economy. The dynamic model identifies the entire response function of a local economy to an ethanol plant in a way that previous input-output models are unable to.

The results show a positive and statistically significant employment multiplier, with the average employment impact of approximately 125-200 jobs per plant during the production phase. Conversely, the analysis shows little evidence of a positive economic effect resulting from

[13] "A POET ethanol plant has a considerable positive effect on a community. It delivers a big impact during the construction phase and then keeps giving in the form of jobs, taxes, increased demand for corn and of course the products it delivers, ethanol, distillers grains and other environmentally-friendly co-products. Better yet, a POET ethanol plant can drive long-term economic development throughout its community and region." Accessed at <http://www.poet.com/inspiration/plants.asp>

the construction of an ethanol plant. The dynamic estimates suggest that the employment multiplier grows over the first several years of ethanol production, and yields a long-run employment impact of approximately 275 jobs per plant.

Ethanol plants provide an interesting case study as a mechanism for economic development. In particular, examining the local economic development effects associated with an ethanol plant provides an opportunity to shed light on the more general topic of rural economic development. Irwin et al. (2010) argue that rural economies should no longer be considered solely farm or agriculturally dominated areas. Furthermore, the authors note that manufacturing has become increasingly important in rural areas. In a broader context, studying ethanol production facilities is useful, because they use a homogenous production technology that offers local economic benefits similar to those of any rural manufacturing plant of similar size and high wages (Low and Isserman, 2009).

The majority of local and regional economic impact studies on ethanol facilities use regional input-output models. The input-output models have been well vetted over the years. However, several studies draw attention to its potential misuse in measuring the effect of biofuel production (Low and Isserman, 2009; Leistritz and Hodur, 2008; Swenson 2006). Blanco and Isenhouer (2010) provide the lone exception of a study that uses historic data. The authors use an econometric model and data from 2005 – 2006 to estimate the effect of ethanol production on local employment and wages. However, as argued in section 3.2, the author's estimation strategy does not make a convincing case for estimating a causal relationship or provide intuitive results that are useful for understanding the magnitude of the effect. In this study, I use historic data to estimate econometrically the employment multiplier for ethanol plants over the course of the ethanol boom from 1990 – 2011 (as depicted in figure 3.1). Using a difference-in-difference strategy the econometric models are estimated such that they can be interpreted intuitively by policy makers and in the context of the previous input-output results.

The rest of the paper is organized as follows. The next section gives a review of the literature on the consequences of ethanol policies and production. I then present the data and empirical methods. Subsequent sections describe the results and concluding remarks.

3.2 Literature review: impact of ethanol production

In this section, I present a brief overview on the range of effects of ethanol production described in the literature. I pay special attention to the local economic development effects of biofuel production.

Several studies have used cost-benefit techniques in an attempt to quantify the aggregate effect of ethanol production. Hahn and Cecot (2009) focus on the benefits of ethanol in reducing greenhouse gases and energy security. The cost side of the equation includes production and distribution, along with government support programs and pollution. They find that increasing ethanol production to 10 billion gallons yields a negative cost-benefit calculation of $3 billion annually. Furthermore, Du et al. (2009) examine the role of ethanol support programs on producer and consumer welfare. The welfare analysis shows a total social cost of approximately $0.89 billion.

Alternatively, another line of inquiry examines the implications of ethanol production on the environment. Ethanol production and consumption has both environmental costs and benefits. The United Nations Environment Program (2009) provides a review of studies using life-cycle analysis for biofuels. In particular, using life-cycle analysis, Liska et al. (2009) find that corn ethanol has 48 – 59% lower greenhouse gas emissions than gasoline. However, Timilsina and Shrestha (2011) argue that biofuels only reduce greenhouse gas emissions absent of related land use changes. While greenhouse gases are a significant environmental outcome, it is also important to consider pollution from agricultural production practices. Langpap and Wu (2011) develop an integrated economic and physical model. The authors show that land use and crop mix changes due to ethanol production will have a large effect on agricultural pollution. Similarly, Lankoski and Ollikainen (2011) find that most agricultural production technologies for biofuels have negative net environmental consequences.

The effect of biofuel production on local and global commodity prices has been the subject of a hotly contested debate among academics and policy makers (Runge and Senauer, 2007). While this debate is far from settled, several studies have begun to shed light on the issues empirically. Research has suggested that ethanol plants increase local grain (McNew and Griffith, 2005; Ugarte et al., 2007) and natural gas (Whistance and Thompson, 2010) prices in the U.S. Zhang et al. (2010) use time series data to examine the short-run and long-run relationships between world commodity and fuel prices. The authors do not find any direct relationship between fuel and agricultural prices in the long-run. However, in the short-run the results suggest that sugar, the common world input for ethanol production used in Brazil, influences other commodity prices. Finally, several studies suggest that U.S. ethanol production influences international food costs (Tokgoz et al. 2008); however, the extent of the effect may depend on the staple food grain in a developing country (Elobeid and Hart, 2007).

Economic development in the agricultural sector and in rural areas has been considered a key driving force behind ethanol policies (Charles et al., 2007). Two studies of note have used economic impact techniques to predict the aggregate effect for the agricultural sector in the United States. The Environmental Protection Agency (EPA 2007) used the Forest and Agricultural Sector Optimization Model to estimate agricultural income. The EPA considers the implications of the Renewable Fuel Standard (7.5 billion gallon per year in 2012) and the Energy Information Administration's projection (9.9 billion gallon per year in 2012). Based on the United States meeting the Renewable Fuel Standard, the model suggests agricultural income will increase by $2.65 billion, a 5% increase. Similarly, agricultural income would rise by $5.41 billion, a 10% increase, under the Energy Information Administration projection. Ugarte et al. (2007) consider a longer time horizon using a dynamic agricultural sector model with an economy-wide input-output model. The authors estimate that a steady increase in ethanol production between the years 2007 – 2030 will lead to a cumulative gain in net farm income of $210 billion. Neither the EPA (2007) nor Ugarte et al. (2007) estimate the aggregate effect of ethanol production on employment.

The majority of research that predicts the effect of ethanol production on local jobs and incomes use economic input-output models. Models of this type have a long history in applied regional economics. They use linkages between sectors of the economy to construct economic multipliers. In many applications these multiplier-based models estimate the direct, indirect, and induced effects of the project of interest. For ethanol production, the direct employment effect comprises the number of employees working directly at the plant. The indirect effect results from purchases of goods and services needed to operate the facility, which supports local jobs. The induced effect results from jobs that are created as employees from the direct and indirect effects spend their earnings. The sum of these three effects then measures the total effects of an ethanol plant on a local economy. Parcell and Westhoff (2006) summarize the work on local and regional economic impact studies prior to 2005. Table 3.1 shows a reproduction of the previous results discussed in Parcell and Westhoff (2006). The variation in the total employment impact ranges from 104 – 1,806 for plants producing 50 million gallons per year (MGY) and under.

While Parcell and Westoff estimate the total employment impact of 264 for a 60 MGY ethanol plant, based on projections from the previous literature.

There are several shortcomings to using the studies listed in table 3.1. First, several of the reports were conducted by consulting firms, who fail to report a detailed methodology of the analysis. Second, the studies provide information on the state, but not the community used in the analysis. Finally, there is no evidence that any of the analysis went through the scrutiny of a peer-review process.

Both Low and Isserman (2009) and Swenson (2006) provide critiques of the previous research, and alternative measures of the local impacts of an ethanol plant. Low and Isserman (2009) argue that special consideration must be given when estimating the local economic impacts of the ethanol industry. The authors suggest using a two-stage process. The first stage involves carefully thinking about the local environment and institutional context in which the ethanol plant is located. The second step involves modifying the input-output table to correspond with the institutional context. Results from these recent studies are shown in table 3.2. Depending the ethanol plant size and community characteristics, the total employment effect ranges from 99 – 250 jobs.

Blanco and Isenhouer (2010) contribute to this line of inquiry by using county-level data to estimate econometrically the effect of ethanol production on employment and wages. Their empirical analysis uses counties in 12 states for the years 2005 and 2006. The authors use a modified version of the empirical methodology proposed by Hanson (2001). They include measures of state wages, state income, and national employment, along with time and state fixed effects as control variables to control for unobserved heterogeneity in a pooled cross-sectional model. Thus, their empirical model uses variation in ethanol production between counties in

2005 and 2006 to estimate the marginal effect of increasing ethanol production[14] on the per-capita employment and wage per job. The results show that ethanol production has a small positive statistically significant effect on per capita employment and wages. However, the analysis is not without its limitations. First, the authors do no credibly deal with the possible endogeniety of ethanol plant locations decisions on their labor market outcome variables. The authors only use two years of data at the start of the ethanol production boom shown in figure 3.1, and do not consider the timing of when plants started production. Finally, the estimated parameters are difficult to interpret in the context of the previous input-output models. For example, when using the full sample of data, increasing ethanol production by one standard deviation (0.025 billion gallons per year) leads to an average per-capita employment increase of 3.83%. This result is difficult to interpret or use for cost-benefit purposes, where policymakers often desire a more intuitive measure of the economic impact of a project.

From the review of the literature it is clear that ethanol production has wide-ranging implications. To quantify fully the costs and benefits of an ethanol plant or broader policies, it is important to consider many different factors. One important factor is the effect of an ethanol plant on local development outcomes. However, there is disagreement among input-output and econometric techniques on the magnitude of the effect. The input-output models are based on several restrictive assumptions and have a high level of uncertainty surrounding them, while the econometric literature is sparse and provides few studies using historic data.

---

[14] The authors also look at ethanol capacity level in each county noting that production and capacity levels are highly correlated.

3.3 Empirical methods and data

3.3.1 Empirical strategy

The objective of this study is to quantify the local economic impact of an ethanol locating in a county. A positive association between a county with an ethanol plant and elevated employment levels would hardly be convincing evidence of a positive economic multiplier. Therefore, to estimate the effect of an ethanol plant locating in a county on the total employment level, I use a "difference-in-differences" strategy. The empirical model uses variation in both the timing of a county acquiring an ethanol plant and the fact that many counties never obtain a plant. Thus, one can think of counties with an ethanol plant as the treatment group, and those without a plant as the control group. The first specification of interest is the following model:

$$EMP_{cst} = \beta_1 Plant\ Construction_{cst} + \beta_2 Plant\ \text{Pr}oduction_{cst} + \alpha_c + \gamma_t + tZ_c + \varepsilon_{cst} \qquad (3.1)$$

where EMP_{cst} is the employment level in county *c*, state *s*, at time *t*. The variables of interest are $Plant\ Construction_{cst}$ and $Plant\ \text{Pr}oduction_{cst}$. $Plant\ Construction_{cst}$ is a dummy variable, set equal to one when county *c* has an ethanol plant under construction at time *t*. Similarly, $Plant\ \text{Pr}oduction_{cst}$ is a dummy variable equal to one when county *c* has a facility producing ethanol. The estimated parameters β_1 and β_2 can then be interpreted as the average change in employment attributable to an ethanol plant in either the construction or production phase. Using panel data allows me to control for several different types of unobserved heterogeneity that could potentially confound the estimated effect of an ethanol plant on the employment level. α_c and γ_t are county and time fixed effects, respectively. The county fixed effect α_c controls for observable and unobservable differences across counties that are constant over time, while γ_t

controls for common shocks that affect employment in all counties, but vary over time. ε_{cst} is a random disturbance term.

However, factors that influence local employment may vary within a state or county over time, potentially confounding estimates of the effect of an ethanol plant on employment. The time-varying local factors may partially be controlled for by the time fixed effect γ_t. Nonetheless, to account for potentially important local-level time-variant heterogeneity, the specification includes both W_{st} and tZ_c. W_{st} is a state-month interaction fixed effect, which is a fully flexible specification to allow different time patterns for each state. In the empirical estimates, Z_c is interacted with a linear trend *t* for each county, such that county trends are allowed to differ linearly from state trends. tZ_c represents time-varying county characteristics affecting the local labor market, such as demographics and policy variables. This represents a flexible way to control for local heterogeneity in the labor market over time. Furthermore, tZ_c helps to identify the effect of interest by mitigating preexisting trends between the treatment and control group.

I also use the following specification to estimate the dynamic effect of ethanol production on county-level employment:

$$EMP_{cst} = \sum_{i=0}^{n} \beta_{-i} Plant\ \Pr oduction_{cst} + \alpha_c + \gamma_t + tZ_c + \varepsilon_{cst} \tag{3.2}$$

Here the variable of interest is $Plant\ \Pr oduction_{cst-i}$, which represents the number of ethanol plants *i* periods after ethanol production started in county *c*. Using the number of ethanol plants is a more flexible and accurate specification than a series of dummy variables for each time period (see Wooldridge, 2002 p. 314). This is due to the fact that several counties in the sample have multiple ethanol plants between 1990 and 2011. The additional ethanol facilities are not accounted for with dummy variables that take the value of one for those counties that have an ethanol plant *i* periods after the initial production start date. The dynamic model in equation 3.2 allows for the differentiation between long-run and short-run economic multiplier effects. Equation 3.2 also uses the same fixed effects strategy as equation 1 to mitigate the possibility of preexisting trends and omitted variable bias.

Additionally, Angrist and Pischke (2009) argue that using a dynamic model with numerous time periods lends itself to a test of causality in the spirit of Granger (1969). Granger's key insight suggests that conditioning on fixed effects, lags of the ethanol plant treatment should predict local employment, while leads should not. In equation 3.2, $Plant\ \mathrm{Pr}oduction_{cst-i}$ constitute lags for the post treatment effect of an ethanol plant on local employment. Adding leads to the start of ethanol production, i.e. $Plant\ \mathrm{Pr}oduction_{cst-i}$ to equation 3.2, and estimating β_{+i}, constitutes a robustness test on the causal influence of an ethanol plant on employment. The dynamic model outlined in equation 3.2 focuses on the effect of the production phase of an ethanol plant. By adding leads to the dynamic model I will be able to also tease out the construction effect. It is important to bear in mind that the average construction period before ethanol production begins is approximately 10 months for the full sample. While it is possible that there are anticipatory impacts prior to construction phase (i.e. β_{+i} for *i* greater than 10 months), it is likely that they are small.

Difference-in-differences research designs are always set up as implicit treatment-control comparisons (Angrist and Pischke, 2009). Meyer (1995) argues that one of the main goals of a research design is finding treatment and control groups that are comparable. In its simplest form, the key identifying assumption in models that include at least time and cross-section fixed effects is that employment trends are the same for the treatment and control groups in the absence of the treatment. The treatment then induces a deviation or shift from the common trend. Studies often use all non-treated groups as the control group. For example, in the archetypal model using a panel of states over several years, all non-treated states are often used as the control group. However, counties are often more heterogeneous than states, such that it is important to consider the validity of the control group. The range of fixed effects included in models 1 and 2 are intended to control for several forms of unobserved heterogeneity in local labor markets. Nonetheless, differences between the treatment and control groups may remain.

To account for potential differences between the treatment and control groups, I also segment the full sample of data. Low and Isserman (2009) provide useful insights into the factors that influence ethanol plant location decisions. The authors argue that ethanol plants tend to locate in rural or mixed rural areas that are in close proximity to inputs, transportation infrastructure, and output markets for byproducts of production. The local location decision is inevitably complex. However, at the county level, the most important location factor is proximity to an adequate corn supply. Corn is the primary input to production and is likely highly correlated with relevant transportation infrastructure (e.g. railroad access) and byproduct users (e.g. distillers grains for cattle consumption). Furthermore, Swenson and Eathington (2006) argue that it is only cost effective to transport corn 50 miles or less to an ethanol plant. Figure 3.2 shows the average historic corn production between 1980-1990 for counties with and without ethanol plants in the sample. Since investing in an ethanol production facility is a long-term decision, it is useful to consider historic rather than contemporaneous corn production. The historic corn yield is also likely a better indicator of the corn supply than the yield in the years directly leading up to construction and production. Figure 3.2 shows that ethanol plants tend to locate in counties with historically high corn yield. Furthermore, it suggests that counties with small quantities of corn production provide little overlapping support as a control group for the treated ethanol plant counties. To account for this, I consider a subset of the full data set. Specifically, I also estimate equations 1 and 2 for non-urban counties with historic corn production above 2.5 million bushels. I obtained historic corn production data from the National Agricultural Statistics Service, and use the 1993 Rural-urban Continuum Code, also known as the Beale Code, to classify urban and non-urban counties.

To estimate the models in equations (3.1) and (3.2), I use unweighted ordinary least squares (OLS) regressions. To control for serial correlation, I correct the standard errors by clustering by county following Arellano (1987).

3.3.2 Data and variables

The employment data in the analysis to follow come from the Local Area Unemployment Statistics (LAUS) of the Bureau of Labor Statistics. The employment data are complemented with detailed information on the timing of construction and production for ethanol plants in 12 states from 1990 – 2011. The full data set consists of information from 1,054 counties over 264 months, resulting in a panel containing 278,256 observations.

The LAUS collects monthly estimates of total employment at various levels of aggregation, including county and county equivalents.[15] The LAUS uses a broad definition of employment. Employed persons are defined as those who worked as paid employees, worked in their own businesses, or worked 15+ hours in a family enterprise during the reference week. It also includes individuals who are employed, but temporarily absent from their job (e.g. vacation, personal matters, etc.). The employment measures include agricultural, nonagricultural, and self-

[15] For more information on the LAUS estimation methodology see http://www.bls.gov/lau/laumthd.htm

employed workers. Both the LAUS and most input-output models (Stevens and Lahr, 1988) do not distinguish between full-time and part-time employment. However, each employed person is counted only once in the LAUS, even if he or she holds multiple jobs.

Industry groups like the Renewable Fuels Association and trade publications such as Ethanol Producers Magazine provide information on the location of ethanol plants. However, to the best of my knowledge, ethanol facility construction and production dates are not systematically collected anywhere. To overcome the data limitations, I use detailed information collected for ethanol plants in 12 States from 1990-2011. The plant list from Ethanol Producers Magazine[16] was used as a starting point for gathering data on all plants and locations in the United States. This initial information was used to conduct an extensive search to identify construction and production start dates for all dedicated ethanol plants. Plants above 10 MGY were chosen due to the availability of information, and because ethanol production is often the primary business activity for plants of 10 MGY or larger.[17] The search included looking through trade publications, local newspapers, and contacting ethanol production facilities directly. Table 3.A1 of the Appendix shows a list of the construction and production start dates for all 142 ethanol plants in the sample.

Figure 3.3 highlights counties in the sample with at least one ethanol plant between 1990 and 2011. According to the plant listing from Ethanol Producers Magazine, the 142 ethanol plants in the sample account for over 80% of the total dedicated ethanol facilities in the U.S.

---

[16] The plant Ethanol Producers Magazine plant list can be accessed at http://www.ethanolproducer.com/plants/listplants/USA/

[17] Using a sample of ethanol plants 10 MGY or greater eliminates research facilities and business operations where ethanol production is a side venture. For example, several beverage manufacturers and recycling centers have begun operating small ethanol plants by using their waste materials.

Iowa has the most ethanol plants of any state in the sample with 34, while North Dakota possesses the fewest with only 3 ethanol facilities. The majority of the ethanol plants are located in Iowa, Nebraska, eastern South Dakota, and southern Minnesota. These areas are noted for their productive agricultural land and are considered the heart of the Corn Belt (Hart, 1986). With the exception of Ohio, all of the states within the sample had some form of state-level ethanol support program over the study time period[18].

Table 3.3 displays descriptive statistics for counties both with and without an ethanol plant. In the full sample, ethanol plants took on average 9.8 months to construct, and produced ethanol for approximately 52 months. Non-ethanol plant counties had an average of 31,307 jobs compared to 18,997 jobs for ethanol plant counties. As noted in the preceding discussion of the empirical strategy, it is important to consider the trends or growth paths of the treatment and control groups. Ethanol plant counties in the full sample experienced lower employment growth rates than counties without an ethanol plant over several pertinent time periods. In the sample time period (1990-2011), non-ethanol plant counties had over a 6% higher employment growth rate. Similarly, in the period leading up to the sample (1980-1990), non-ethanol plant counties' employment growth was over 4% more than the growth in ethanol plant counties. This suggests that these groups have different employment growth paths. Ethanol plant counties were more likely to be non-urban, and had almost two-and-one-half times the amount of historic corn production.

It is possible that the full sample does not provide suitable overlapping support or 'observational equivalence' between the treatment and control groups. Table 3.3 also shows summary statistics for non-urban counties with historic average corn production greater than 2.5

---

[18] See Cotti and Skidmore (2007) for a list of tax credits and subsidies for ethanol by state.

million bushels per year. The summary statistics in table 3.3 show that average employment level and employment growth are similar for ethanol and non-ethanol plant counties. In particular, employment growth in the decade of the 1980s is 1.92% per year for ethanol plant counties and 2.4% per year for non-ethanol plant counties. This suggests that there is not a significant difference in the growth paths of the two groups. Meanwhile, ethanol plant counties still have an edge in historic corn production, but a smaller difference when compared to the full sample. The non-urban, high corn counties described in table 3.3 compare favorably to the counties proposed by Low and Isserman (2009).

Figure 3.4 shows the production start dates for plants in the sample. The growth of ethanol plants in the 12 Corn Belt states shares a similar exponential trend with the growth of ethanol production in the U.S. depicted in figure 3.1. Ethanol was still an infant industry in the 1990s. The ethanol industry experienced a boom in the number of plants in the mid 2000s, peaking in 2007. The peak of new plants going online coincided with what has come to be known as "The Great Recession" in the U.S. The recession took place between December 2007 and June 2009 (NBER, 2012). To test the robustness of the results I also examine only the pre-recession time period (1990 – 2007).

Ethanol plants can be viewed as homogenous manufacturing facilities. Nonetheless, they are built in different sizes and vary in the capacity of ethanol that they are able to produce. While the capacity of an ethanol plant and its actual production can be different, anecdotal evidence suggests that plants produce at approximately their full capacity. Figure 3.5 shows the distribution of ethanol capacity in millions of gallons per year for the full sample of data. The distribution of capacity is bimodal. It is likely that the clustering of ethanol plants around the capacity size of 60 MGY is due to the Small Ethanol Producer Tax Credit.[19] The SEPTC provides a $0.10 per gallon blender tax incentive for the first 15 million gallons of ethanol sold from a plant with not more than 60 million gallons of capacity. Two large ethanol facilities owned by the same global agribusiness firm reportedly produce 275 and 300 MGY respectively. The 275 MGY plant is included in the non-urban, high corn subsample, while the larger 300 MGY plant is not. The results that follow are robust with respect to excluding these relatively large facilities. While ethanol plants vary in their production capacity, due to the uniformity of available technology, it is unlikely that they differ substantially in construction and operations.

3.4 Results

Table 3.4 presents estimates of equation 1 using the full sample of data, while table 3.5 shows a parallel set of results for the non-urban, high corn subset of counties. The results in tables 3.4 and 3.5 can be interpreted as the average effect of an ethanol plant on employment during construction and production. Columns 1, 3, and 5 examine both the construction and production phase, while the columns 2, 4, and 6 only consider the effect of an ethanol plant in the production phase. Each column corresponds to a different specification. The latter columns use techniques that control for increasing dimensions of unobserved heterogeneity. The first two columns present the basic model, which includes time and county fixed effects. The third and fourth columns add a state-month interaction fixed effect. Finally, the fifth and sixth columns show estimates for models that add a county-specific linear trend to the aforementioned models with time, county, and state-month fixed effects. The models estimated in the last two columns control for the highest degree of unobserved heterogeneity. The county specific time trends

[19] More information on the SEPTC can be found with the Department of Energy online at http://www.afdc.energy.gov/afdc/laws/law/US/352

attempt to prevent the estimated impact of ethanol plants from capturing differences in preexisting employment trends between the treatment and control groups.

The results of the full sample in table 3.4 show striking differences among the three specifications. The basic specification with county and time fixed effects suggests that ethanol plant construction and production is associated with a decrease in the average employment level. When adding in state-month interaction fixed effects (columns 3 and 4), the negative relationship between ethanol plants and employment remains. However, the estimate for the average production effect becomes positive and statistically significant when flexibly allowing state trends to vary over time. The final specification adds in a linear trend for each county. The county-specific linear trend flexibly allows for different trends between the treatment and control groups. This specification yields positive and statistically significant results for the production phase in both models. The construction effect is positive and small, but statistically insignificant.

Table 3.5 displays my preferred specification using the non-urban, high corn subset of counties. As argued in the previous section, segmenting the data into non-urban agriculturally productive counties provides a higher probability of common support between the treatment and control counties. The results in table 3.5 are robust across different specifications. For all specifications, I find that an ethanol plant in the production phase is associated with an increase in the average employment level for the period from 1990-2011. In particular, I find that, on average, an ethanol plant in the production phase generates between 124 and 185 total jobs. This result is smaller in magnitude than the full sample of counties that control for a variety of forms of heterogeneity and county specific linear trends (Table 3.4 columns 3.5 and 3.6).[20]

Table 3.6 presents estimates of β_{-i} for the dynamic specification[21] described in equation 2, with the county employment level as the dependent variable. Examining the dynamics allows me to trace out the full employment adjustment path. In presenting the employment dynamics I focus on the non-urban, high corn counties. The coefficients in the table show the effect of an ethanol facility on local employment for every 2 quarters (half a year) since ethanol production started. The final coefficient (5+ years) shows the long-run employment effect of an ethanol plant. As before, each column corresponds to a different specification, where the latter columns include fixed effects that control for increasing dimensions of unobserved heterogeneity.

The results in table 3.6 are shown graphically in figure 3.6, and are robust across each of the three specifications. They reveal that the employment multiplier for an ethanol plant in a non-rural, high-corn county unfolds over the course of several years. All three specifications suggest that the employment multiplier associated with an ethanol plant rises slowly over the first several years of production. This implies that it takes several years for the local economy to

[20] I check the robustness of the results in several ways. Tables A2 and A3 in the appendix show similar results for the average effect from construction and production of an ethanol plant over the pre-recession time period from 1990 - 2007. Table A4 and A5 show pre-recession results for the dynamic model specifications. They suggest that the Great Recession of 2008 isn't driving the magnitudes of the estimated employment benefits for ethanol plant counties. Furthermore, table A6 in the appendix show that the magnitude of the average ethanol plant effect on employment is robust to including counties adjacent to ethanol plants as explanatory variables.

[21] Wolfers (2006) argues that the difference-in-difference estimation strategy may yield misleading results in average effect using a strategy similar to equation 1. This could occur if cross-section specific trends also pick up the effect of a policy, and not just the preexisting trends. Thus, Wolfers argues it is instructive to trace out the dynamic effect of a policy or project to allow the cross-section trend (in this case, the county-specific trend) to identify preexisting trends.

adjust to the increased economic activity associated with the ethanol facility. The employment effect then spikes after 4 to 5 years before settling to the long-run effect. The long-run effect varies depending on the specification, and is higher than the average effect found in table 3.5. This trajectory follows anecdotal evidence provided by several ethanol facility managers, where firms associated with the input and output of ethanol production began to cluster near the ethanol plant.

I check the robustness of the results by analyzing the timeframe before an ethanol plant started production in a county. A causal interpretation of the previous findings would be weakened if employment levels were changing in counties that ethanol plants located in (compared to non-ethanol plant counties) before construction and production began. In order to examine this issue, I follow the discussion in the empirical strategy and add leads prior to construction to the dynamic specification. The leads added to specification 2 are coded such that an ethanol plant would start production in 2 quarters, 4 quarters, and so on up to 3 years. Figure 3.7 graphically shows the results for the modified specification displayed in table 3.7.

The average ethanol plant took less than one year to construct. Thus, the first year before production constitutes the construction phase, and leads prior to 1 year comprise the causality test of interest. Industry specialists from outside the region build most ethanol plants. Therefore, it is likely that the economic multiplier effect for the construction phase is small and short lived. The employment multiplier for construction differs by specification. The results displayed in columns 1 and 2 suggest that the construction phase is not statistically different from zero. However, column 3, which includes a county-specific time trend shows that the employment response during construction is positive (approximately 60 to 70 jobs) and statistically significant at the 10% level. The results in table 3.7 show weak evidence at best of a positive construction multiplier. Now turning to the time period before ethanol construction. The coefficients for the variables leading construction in the first and third specifications are both individually and jointly statistically indistinguishable from zero. Specification 2 suggests that it is possible that ethanol plant counties were on a negative trend prior to the start of construction. Overall, the coefficients to the left of the vertical line in figure 3.6 (before production) do not provide strong evidence of a preexisting trend in ethanol plant counties.

3.5 Conclusion

On the surface, one might expect that a firm choosing to locate in a region would spur local economic growth. The direct effect (people employed directly by the firm) suggests that this would be the case. Traditional economic impact techniques (e.g. input-output models) account for the supply linkages and induced spending of a firm's location decision. However, they are unable to account for other positive or potentially negative economic forces. Negative effects can occur because of increasing costs, competition for labor, and congestion for public services and infrastructure (Edmiston, 2004). While it is possible that some of these negative spillovers could be present for an ethanol plant locating in a county, nevertheless, it is likely that positive spillovers are the more influential economic force for an ethanol plant locating in rural communities. The positive agglomeration forces increase the attractiveness of a location to other firms.

The objective of this research was to examine the effect of ethanol production on local employment levels. Few studies examine this question with historical data, largely due to the data constraints related to local ethanol production. I use a data set containing the timing of ethanol plant construction and production to estimate the net employment effect, using a difference-in-differences framework. Further, I add leads and lags to the start of ethanol production to examine the dynamic response of an ethanol plant on a local economy. When using non-urban, high corn counties, the results suggest a positive and statistically significant employment multiplier, with the overall average local employment impact of approximately 125-200 jobs. This average local employment effect during the production phase is consistent with the more recent input-output modeling experiments (e.g. Low and Isserman 2007). Conversely, the analysis shows little evidence of a positive economic effect resulting from the construction of an ethanol plant. The dynamic estimates suggest that the local employment multiplier grows over the first several years of ethanol production, and yields a long-run employment impact of approximately 275 jobs per plant.

The impetus for ethanol production has come from state subsidies and tax breaks and federal policies such as consumption mandates, tax credits, and import tariffs. The federal policies lead to higher ethanol and gasoline prices for all consumers. Thus, it is challenging to determine the total financial cost from ethanol policies. Nonetheless, the local employment benefits from these policies are regionally concentrated. It is important for policy makers to take into account the overall financial costs of ethanol policies, as well as the local benefits. These results are useful to policy makers, who often tout the benefits to communities where ethanol is produced. My analysis shows that, on average, an ethanol plant does create positive employment spillovers to non-urban counties. However, policy makers should be cautious about justifying the significant policy support on the basis of local employment benefits alone. Finally, this research could be helpful to policymakers in examining cellulosic ethanol. Cellulosic ethanol is often viewed as the next generation of biofuels. It shares many similarities to current ethanol production techniques, and would likely share a similar local employment response.

Table 3.1. Summary of Previous Studies Investigating Total Local and Regional Economic Effects of Ethanol Production from Parcell and Westhoff (2006)

Source	Year	Size (MGY)*	Location	Direct Jobs	Indirect Jobs	Total Jobs
Swenson	2005	41	IA	32	135	167
BBI International Consulting	2004	20	ID	28	363	339
Petersan	2003	24	NE	31	73	104
BBI International Consulting	2003	10	HI	22	154	176
Urbanchuk and Kappell	2002	40	n/a	41	694	732
Resource Systems Group	2000	50	NY	53	1,753	1,806
Parcell and Westhoff**	2006	60	-	54	210	264

*Production capacity in millions of gallons per year (MGY)
** Based on projections from previous studies listed in Table 1.

Table 3.2. Summary of Previous Economic Impact Results

Source	Size (MGY)*	Location (County)	Location Type	Direct Jobs	Indirect Jobs	Induce d Jobs	Total Jobs
Low and Isserman (2009)	60	Coles, IL	Mixed Rural	35	83	34	152
Low and Isserman (2009)	60	Harlan, NE	Rural	35	50	15	99
Low and Isserman (2009)	100	Hamilton, IL	Rural	39	97	17	153
Low and Isserman (2009)	100	Kankakee, IL	Mixed Rural	39	152	59	250
Swenson (2006)	50	3 County Region, IA	-	35	75	23	133
Swenson (2008)	100	3 County Region, IA	-	46	95	29	170

*Production capacity in millions of gallons per year (MGY)

Table 3.3. Descriptive Statistics

	Full Sample		Non-Urban, High Corn Sample	
	Ethanol Plant Counties	Non-ethanol Plant Counties	Ethanol Plant Counties	Non-ethanol Plant Counties
Average Construction Time (in months)	9.79	-	10.31	-
Average Production Time (in months)	51.69	-	54.70	-
Average Employment	18,997	31,307	10,517	8,860
Employment Growth (1990-2011)	-3.88%	2.45%	-6.85%	-3.76%
Employment Growth (1990-2005)	12.38%	17.22%	11.00%	10.84%
Employment Growth (1980-1990)	3.77%	8.30%	1.92%	2.40%
% Non-Urban	80.92%	71.94%	-	-
Historic Average Corn Production*	13,103,960	5,449,718	14,633,200	9,289,700
Number of Counties	131	923	88	280

* The historic corn production calculated by averaging the number of bushels produced in the years 1980, 1985, and 1990

Table 3.4. Average Impact of Ethanol Plants on Employment Levels 1990-2011: Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Construction	-713.1***		-711.2***		26.77	
	(251.6)		(72.72)		(33.28)	
Production	-473.1	-384.8	-470.6***	-382.5***	621.8***	611.5***
	(339.8)	(313.4)	(93.87)	(85.53)	(65.61)	(57.68)
Observations	278,256	278,256	278,256	278,256	278,256	278,256
R-squared	0.056	0.055	0.011	0.011	0.059	0.059
Number of Counties	1,054	1,054	1,054	1,054	1,054	1,054
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
State-Month Fixed Effect	No	No	Yes	Yes	Yes	Yes
County Trend, Linear	No	No	No	No	Yes	Yes
Cluster	County	County	County	County	County	County

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3.5. Average Impact of Ethanol Plants on Employment Levels 1990-2011: Non-Urban, High Corn Counties

	(1)	(2)	(3)	(4)	(5)	(6)
Construction	-35.25		-34.89		-63.17***	
	(93.94)		(23.55)		(17.79)	
Production	189.9	195.1*	190.4***	195.7***	124.7***	150.0***
	(127.6)	(116.6)	(40.07)	(37.67)	(19.59)	(16.05)
Observations	97,152	97,152	97,152	97,152	97,152	97,152
Number of Counties	368	368	368	368	368	368
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
State-Month Fixed Effect	No	No	Yes	Yes	Yes	Yes
County Trend, Linear	No	No	No	No	Yes	Yes
Cluster	County	County	County	County	County	County

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3.6. Dynamic Impact of Ethanol Plants on Local Employment Levels 1990-2011: Non-Urban, High Corn Counties

	(1)	(2)	(3)
First half year since production started	41.44	39.93	103.1***
	(78.78)	(25.20)	(19.07)
1 year of production	55.23	57.30**	123.4***
	(86.64)	(27.15)	(20.61)
1.5 years of production	141.3	139.2***	204.1***
	(93.99)	(29.82)	(22.49)
2 years of production	144.1	147.4***	220.5***
	(94.14)	(32.00)	(22.16)
2.5 years of production	193.8*	190.6***	273.1***
	(98.68)	(35.36)	(26.30)
3 years of production	211.4**	215.5***	300.0***
	(105.6)	(40.39)	(27.31)
3.5 years of production	231.3*	229.0***	315.3***
	(119.5)	(49.45)	(28.54)
4 years of production	306.4**	311.0***	381.5***
	(128.8)	(58.37)	(31.35)
4.5 years of production	370.7***	367.3***	418.9***
	(139.3)	(62.33)	(33.22)
5 years of production	302.3**	311.1***	422.1***
	(141.5)	(67.03)	(30.13)
5+ years of production	271.1*	271.6***	332.6***
	(161.2)	(61.20)	(33.95)
Observations	75,072	75,072	75,072
Number of Counties	368	368	368
Time Fixed Effect	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes
State-Month Fixed Effect	No	Yes	Yes
County Trend, Linear	No	No	Yes
Cluster	County	County	County

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3.7. Dynamic Impact of Ethanol Plants on Local Employment Levels 1990-2011 Including Leads: Non-Urban, High Corn Counties

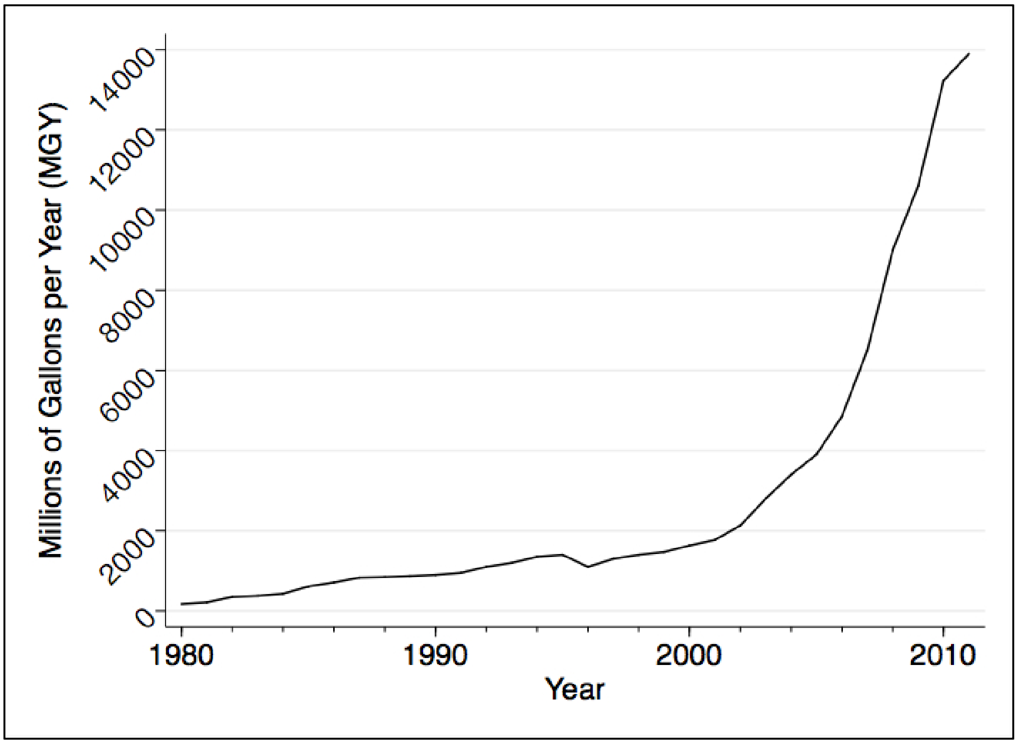
	(1)	(2)	(3)
3 years before production	-55.09	-56.68**	17.64
	(68.47)	(27.79)	(18.99)
2.5 years before production	-93.14	-90.71***	-10.58
	(66.74)	(26.32)	(19.09)
2 years before production	-81.30	-83.09***	2.599
	(73.17)	(29.14)	(24.96)
1.5 years before production	-92.24	-89.37***	-0.0412
	(70.92)	(25.57)	(24.25)
1 year before production	-19.72	-22.75	71.56**
	(75.99)	(26.22)	(29.62)
First half year before production starts	-33.30	-29.08	59.79*
	(76.03)	(26.66)	(30.64)
First half year since production started	26.03	24.65	107.2***
	(91.15)	(33.23)	(32.83)
1 year of production	48.03	51.61	125.0***
	(99.30)	(39.46)	(34.52)
1.5 years of production	132.4	127.7***	203.2***
	(112.4)	(38.55)	(42.75)
2 years of production	153.7	163.3***	241.9***
	(111.7)	(43.57)	(47.84)
2.5 years of production	149.9	141.9***	238.3***
	(124.9)	(41.04)	(54.61)
3 years of production	173.1	182.1***	306.8***
	(127.9)	(45.49)	(57.11)
3.5 years of production	209.5	202.3***	316.7***
	(156.6)	(53.33)	(62.72)
4 years of production	262.6*	272.4***	379.3***
	(156.4)	(51.74)	(65.13)
4.5 years of production	337.3*	331.1***	399.4***
	(177.5)	(52.83)	(67.94)
5 years of production	323.5*	334.9***	393.7***
	(172.7)	(50.74)	(67.49)
5+ years of production	264.9	265.2***	343.5***
	(194.6)	(64.18)	(69.70)

Table 3.7. (cont'd)

	(1)	(2)	(3)
Observations	61,824	61,824	61,824
Number of Counties	368	368	368
Time Fixed Effect	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes
State-Month Fixed Effect	No	Yes	Yes
County Trend, Linear	No	No	Yes
Cluster	County	County	County

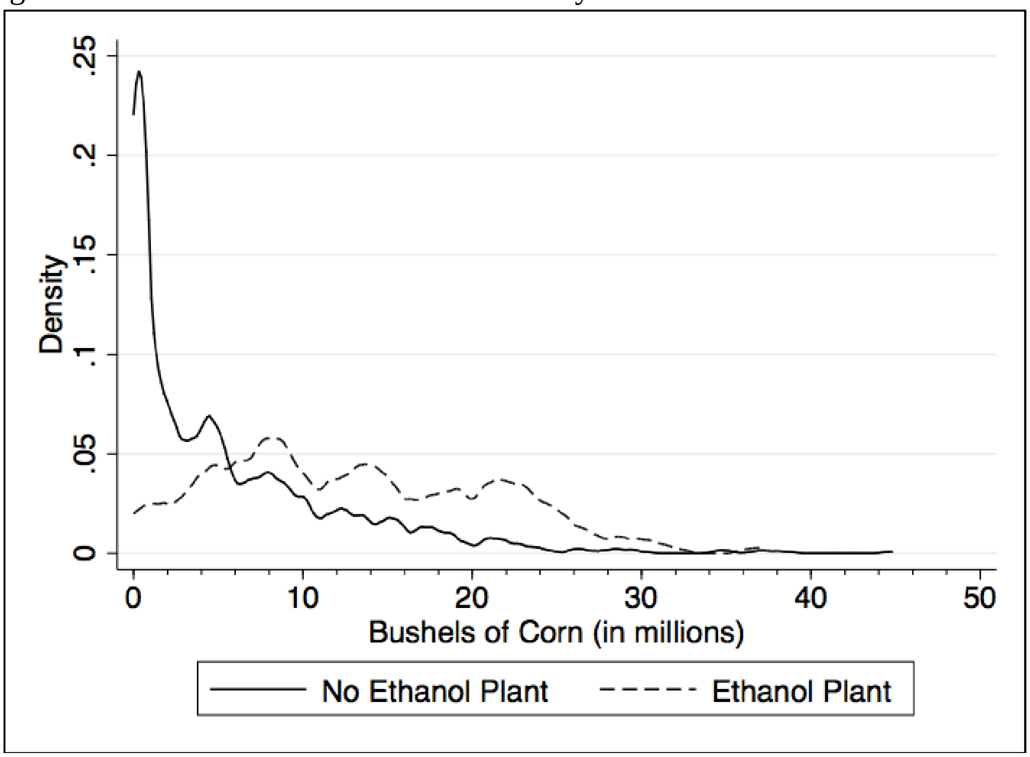
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Figure 3.1. Fuel Ethanol Production in the United States, 1980 – 2011



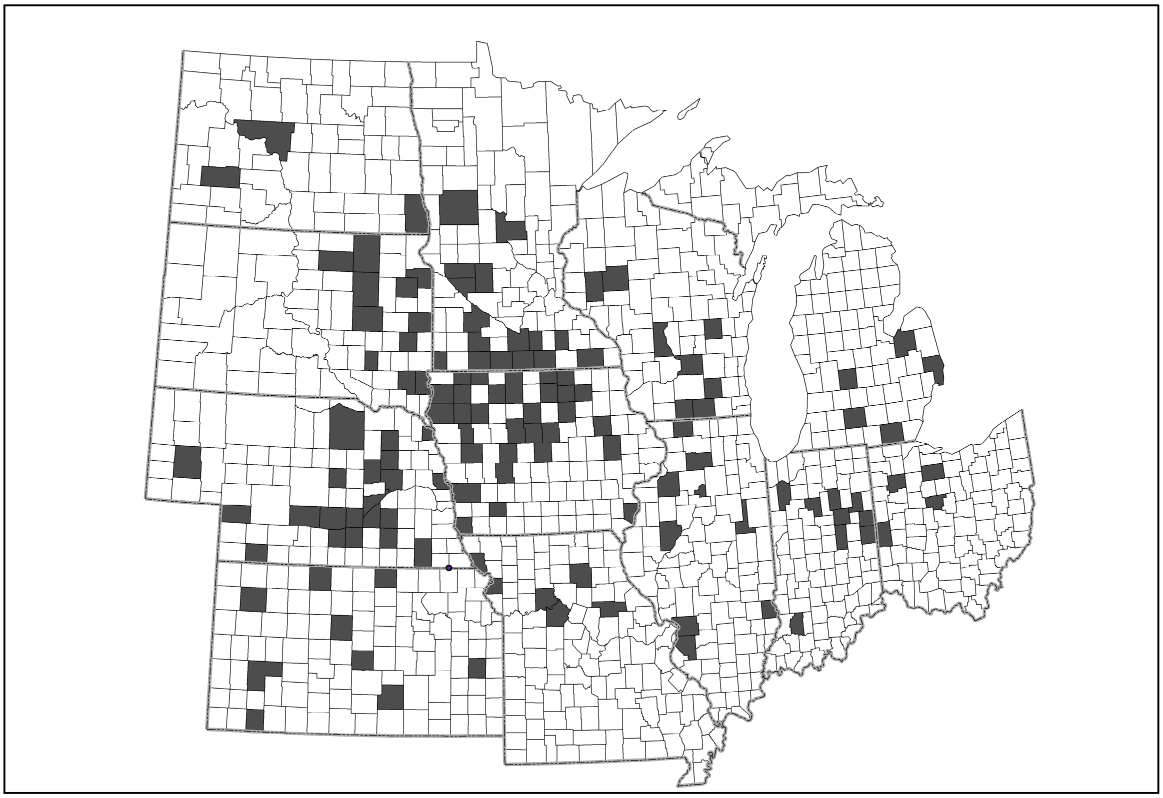
Source: Industry Statistics from the Renewable Fuels Association (2012)

Figure 3.2. The Distribution of Historic County-Level Corn Production 1980-1990



Source: Ethanol plant data from the Renewable Fuels Association (2012) and corn production data from the USDA National Agricultural Statistics Service (2012)

Figure 3.3. Full Sample: Ethanol Plants 1990 - 2011



* Note counties with at least 1 ethanol plant between 1990 - 2011 are shaded

Figure 3.4. Ethanol Production Start Dates 1990 – 2011

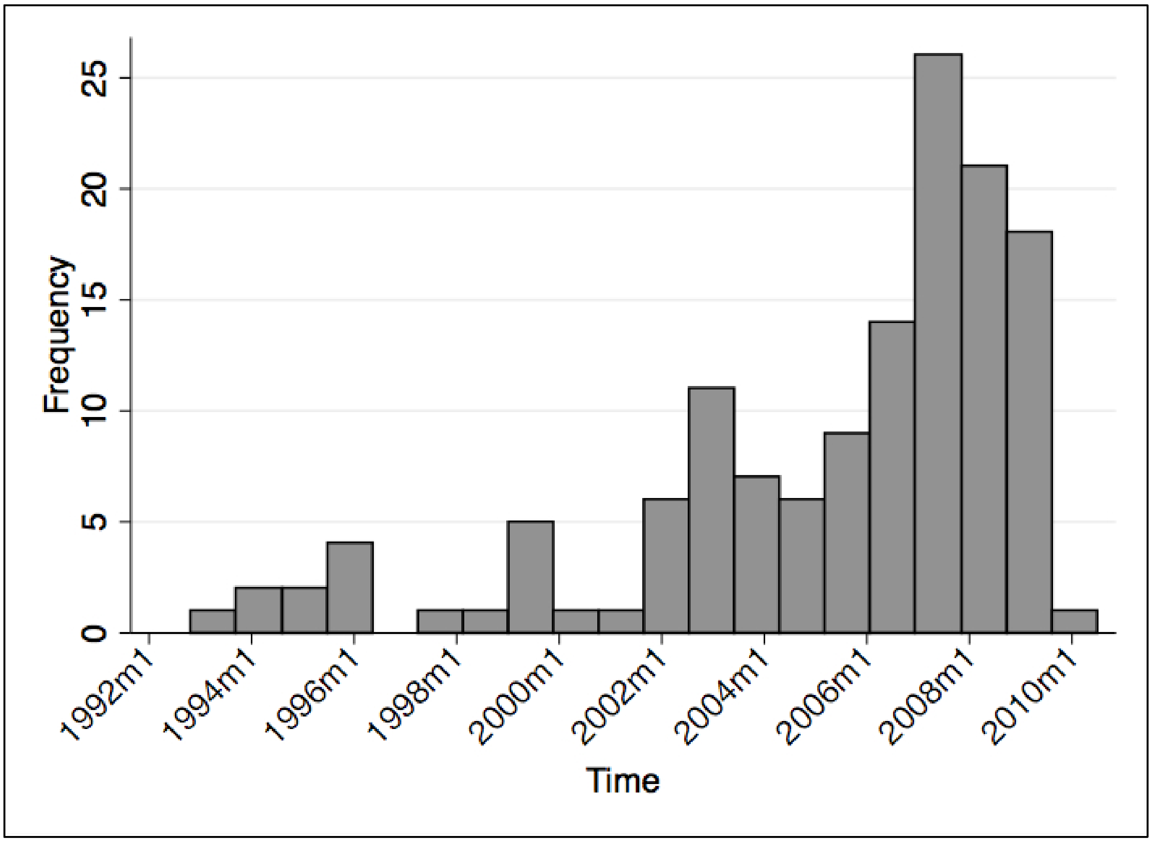


Figure 3.5. Distribution of Ethanol Plant Capacity

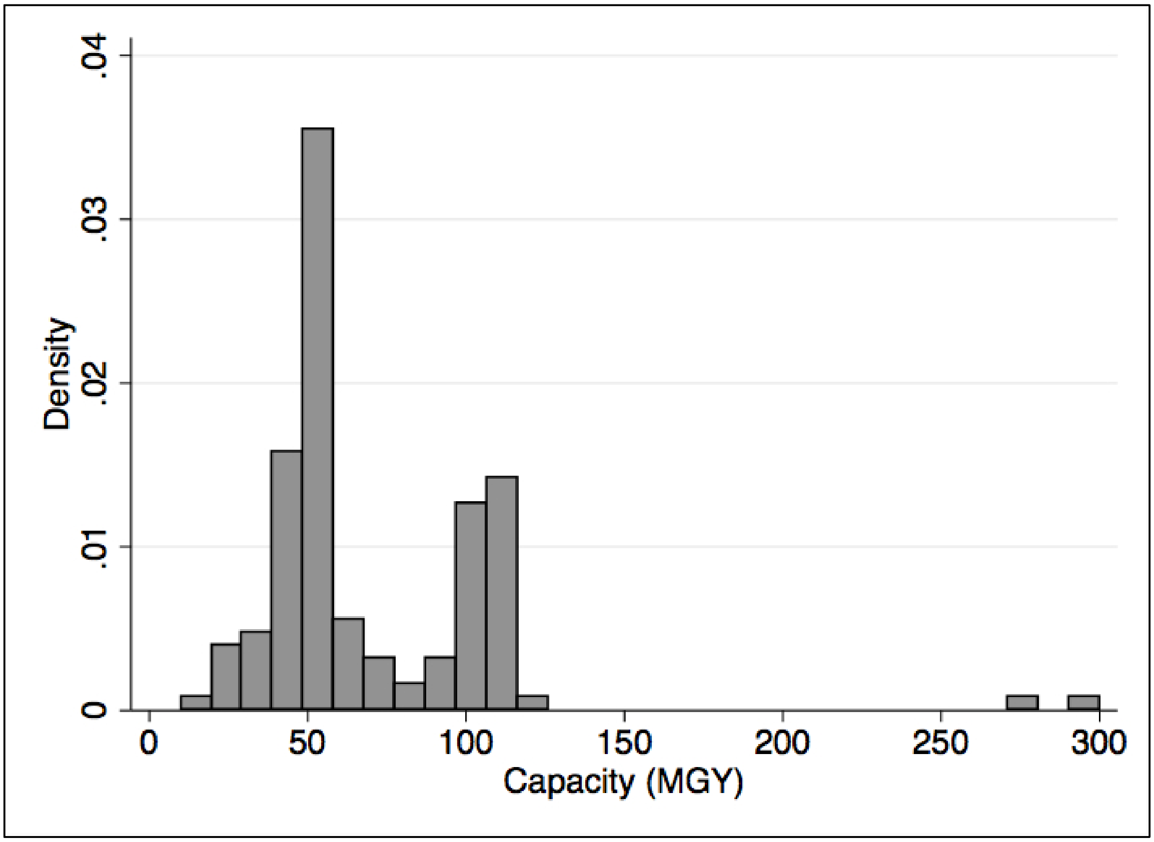


Figure 3.6. Response of employment to an ethanol plant

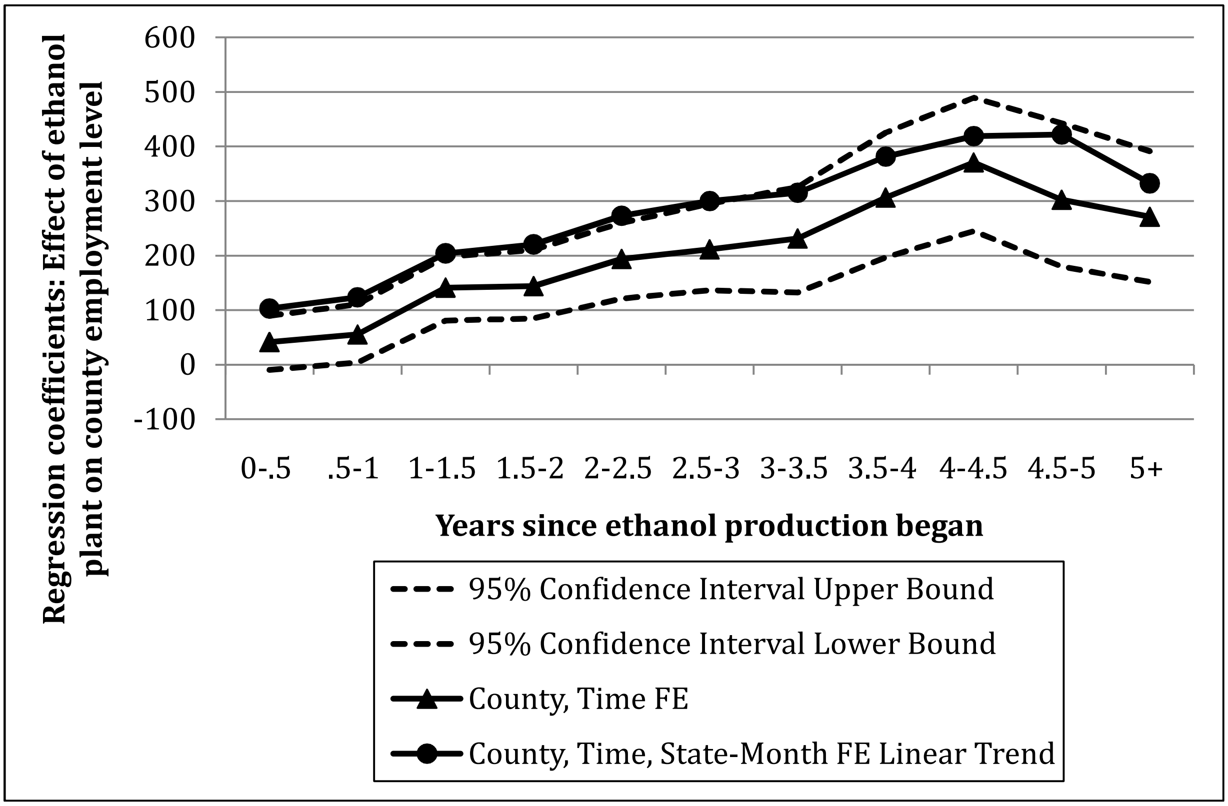
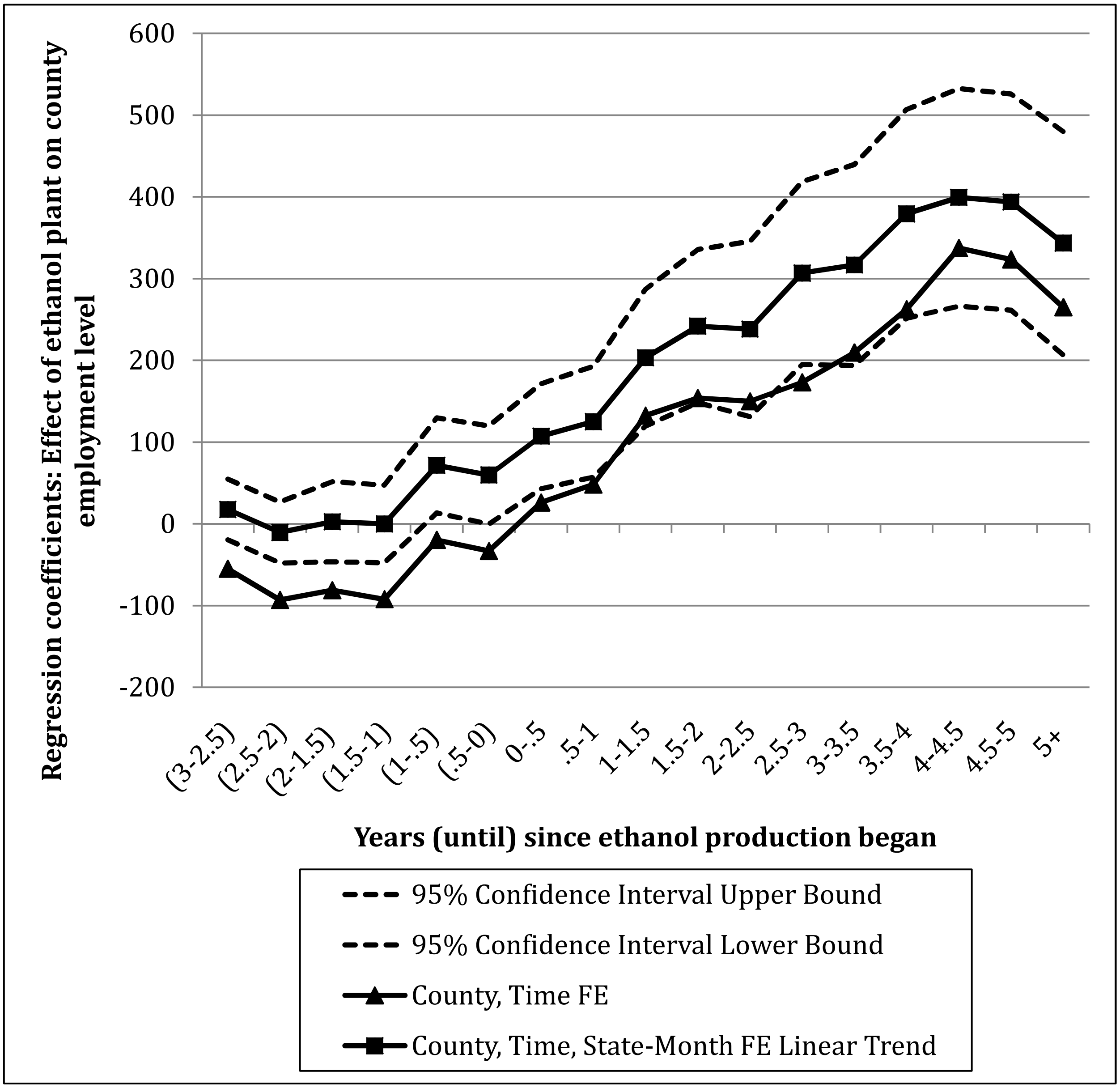


Figure 3.7. Response of employment to an ethanol plant, including leads



APPENDIX

APPENDIX

Table 3.A1. Ethanol Plants

Plant Name	Location (County, State)	Capacity*	Production Start Date	Construction Start Date
Ag Processing Inc.	Adams, NE	52	Nov-99	Jun-98
Poet Biorefining-Corning	Adams, IA	60	May-11	Apr-10
East Kansas Agri-Energy LLC	Anderson, KS	35	Jun-09	Sep-08
Poet Biorefining-Laddonia	Audrain, MO	50	Sep-10	Nov-09
Heartland Grain Fuels LP	Beadle, SD	30	Nov-03	Jun-02
Poet Biorefining-Lake Crystal	Blue Earth, MN	56	May-09	Aug-08
Valero Renewable Fuels LLC	Boone, NE	100	Oct-11	Mar-10
Valero Renewable Fuels LLC	Brookings, SD	120	Dec-07	Jan-07
Heartland Grain Fuels LP	Brown, SD	48	Dec-97	Mar-96
Poet Biorefining-Groton	Brown, SD	50	May-07	Apr-06
Hawkeye Energy Holdings LLC	Buchanan, IA	115	Jun-10	Jan-09
LifeLine Foods LLC	Buchanan, MO	40	Jul-11	Mar-10
Valero Renewable Fuels LLC	Buena Vista, IA	100	Nov-10	Jun-09
Abengoa Bioenergy of Nebraska LLC	Buffalo, NE	88	Jul-11	Dec-09
Hawkeye Energy Holdings LLC	Butler, IA	115	Oct-12	Aug-11
The Andersons Albion Ethanol LLC	Calhoun, MI	55	Aug-10	Sep-09
Show Me Ethanol LLC	Carroll, MO	55	May-12	Mar-11
The Andersons Clymers Ethanol LLC	Cass, IN	110	May-11	Feb-10
Golden Grain Energy LLC	Cerro Gordo, IA	80	Dec-08	Oct-07
Little Sioux Corn Processors LP	Cherokee, IA	92	Apr-07	Nov-05
Homeland Energy Solutions LLC	Chickasaw, IA	100	Apr-13	Jul-11
Ace Ethanol LLC	Chippewa, WI	42	Jun-06	Jun-05
Granite Falls Energy LLC	Chippewa, MN	50	Nov-09	Aug-08
Glacial Lakes Energy LLC	Codington, SD	100	Dec-04	Sep-05
Didion Ethanol LLC	Columbia, WI	50	Mar-12	Oct-10
United Wisconsin Grain Producers LL	Columbia, WI	55	Apr-09	Oct-07
Poet Biorefining-Bingham Lake	Cottonwood,MN	30	Jul-01	Mar-00

Table 3.A1. (cont'd)

Plant Name	Location (County, State)	Capacity*	Production Start Date	Construction Start Date
Amaizing Energy LLC	Crawford, IA	55	Sep-09	Aug-08
Lincolnland Agri-Energy LLC	Crawford, IL	45	Jul-08	Mar-07
Siouxland Ethanol LLC	Dakota, NE	50	May-11	Nov-09
The Andersons Marathon Ethanol LLC	Darke, OH	110	Mar-12	Sep-10
Grain Processing Corp.	Daviess, IN	20	Mar-03	Mar-02
Poet Biorefining-Mitchell	Davison, SD	60	Dec-10	Oct-09
Cornhusker Energy Lexington LLC	Dawson, NE	40	Dec-09	Mar-07
Big River Resources West Burlington	Des Moines, IA	92	Apr-08	Nov-06
Green Plains-Superior	Dickinson, IA	55	Jul-12	Aug-10
Al-Corn Clean Fuel	Dodge, MN	36	May-00	Jun-99
Western Wisconsin Energy LLC	Dunn, WI	40	Sep-10	May-08
Glacial Lakes Energy LLC	Edmunds, SD	100	Jun-12	Sep-10
Corn Plus LLLP	Faribault, MN	44	Nov-98	Mar-97
Advanced BioEnergy LLC	Fillmore, NE	100	Oct-11	Dec-09
Poet Biorefining-Preston	Fillmore, MN	46	Aug-02	Jun-01
Bonanza BioEnergy LLC	Finney, KS	55	Oct-11	May-10
Valero Renewable Fuels LLC	Floyd, IA	110	Apr-11	Mar-10
One Earth Energy LLC	Ford, IL	100	Jun-13	Oct-11
Poet Biorefining-Glenville East	Freeborn, MN	45	Sep-03	Mar-02
Green Plains-Shenandoah	Fremont, IA	55	Jun-11	Nov-09
Riverland Biofuels LLC	Fulton, IL	38	Jun-11	Oct-10
E Energy Adams LLC	Gage, NE	50	Nov-11	Jun-10
Central Indiana Ethanol LLC	Grant, IN	40	Mar-11	Oct-09
Poet Biorefining-Big Stone	Grant, SD	75	Jun-06	Mar-05
Badger State Ethanol LLC	Green, WI	55	Oct-06	Jul-05
Louis Dreyfus Commodities	Greene, IA	100	May-13	Mar-12
Hawkeye Energy Holdings LLC	Guthrie, IA	115	Oct-12	May-11
Poet Biorefining-Coon Rapids	Guthrie, IA	54	Aug-06	Jul-05
Biofuel Energy Corp.	Hall, NE	115	Mar-11	May-10
Nebraska Energy LLC	Hamilton, NE	50	Oct-99	Mar-98
Poet Biorefining-Jewell	Hamilton, IA	60	Mar-10	Apr-09
Hawkeye Energy Holdings LLC	Hardin, IA	100	Nov-08	Nov-07

Table 3.A1. (cont'd)

Plant Name	Location (County, State)	Capacity*	Production Start Date	Construction Start Date
Big River Resources Galva LLC	Henry, IL	100	May-13	Jan-13
Patriot Renewable Fuels LLC	Henry, IL	100	Aug-12	Feb-11
Trenton Agri Products LLC	Hitchcock, NE	40	Mar-08	Jul-07
Golden Triangle Energy Co-op Inc.	Holt, MO	20	Feb-05	Mar-04
NEDAK Ethanol LLC	Holt, NE	44	Dec-12	Jun-10
Platinum Ethanol LLC	Ida, IA	110	Sep-12	Nov-10
Quad County Corn Processors	Ida, IA	30	Feb-06	Mar-05
Carbon Green BioEnergy Woodbury LLC	Ionia, MI	50	Sep-10	May-09
Heron Lake BioEnergy LLC	Jackson, MN	50	Sep-11	Nov-09
Iroquois Bio-Energy Company LLC	Jasper, IN	40	Jan-11	Sep-09
Poet Biorefining-Portland	Jay, IN	60	Sep-11	Aug-10
Valero Energy Corp.	Jefferson, WI	110	Jul-11	Sep-10
Castle Rock Renewable Fuels LLC	Juneau, WI	50	Feb-12	Oct-10
Bushmills Ethanol LLC	Kandiyohi, MN	49	Dec-09	Oct-08
KAAPA Ethanol LLC	Kearney, NE	40	Nov-07	Jul-06
Global Ethanol/Midwest Grain Process	Kossuth, IA	100	Nov-06	Jun-05
Dakota Ethanol LLC	Lake, SD	50	Sep-05	Apr-04
Illinois River Energy LLC	Lee, IL	100	Nov-10	Nov-08
Global Ethanol/Midwest Grain Process	Lenawee, MI	57	Mar-11	Aug-09
Poet Biorefining-Hudson	Lincoln, SD	55	May-08	May-07
Archer Daniels Midland Co.	Linn, IA	275	Oct-12	Jun-10
Penford Products Corp.	Linn, IA	37	Sep-12	Nov-10
Poet Biorefining-Macon	Macon, MO	36	May-04	May-03
Abengoa Bioenergy of Illinois LLC	Madison, IL	88	Jul-12	Dec-11
Elkhorn Valley Ethanol LLC	Madison, NE	40	Sep-11	Sep-09
Poet Biorefining-Alexandria	Madison, IN	65	Apr-12	Feb-11
Poet Biorefining-Marion	Marion, OH	68	Oct-12	May-11
Biofuel Energy Corp.	Martin, MN	115	Jan-12	Oct-10
Valero Renewable Fuels LLC	Martin, MN	110	Jun-13	Nov-10
Blue Flint Ethanol LLC	McLean, ND	50	Feb-11	Nov-09
Green Plains-Central City	Merrick, NE	100	Jul-08	Jul-07
Absolute Energy LLC	Mitchell, IA	100	Feb-12	Aug-10

Table 3.A1. (cont'd)

Plant Name	Location (County, State)	Capacity*	Production Start Date	Construction Start Date
Bridgeport Ethanol LLC	Morrill, NE	50	Oct-12	Sep-11
Central Minnesota Ethanol Co-op	Morrison, MN	20.5	Mar-03	Apr-02
Valero Renewable Fuels LLC	O'Brien, IA	110	Aug-12	Nov-10
Poet Biorefining-Ashton	Osceola, IA	55	Mar-08	Jul-06
Otter Tail Ag Enterprises LLC	Otter Tail, MN	57.5	Mar-12	Oct-10
Poet Biorefining-Emmetsburg	Palo Alto, IA	50	Apr-09	Apr-08
Standard Ethanol Madrid LLC	Perkins, NE	44	Jul-11	Dec-09
Prairie Horizon Agri-Energy LLC	Phillips, KS	40	Jul-10	Sep-09
Husker Ag LLC	Pierce, NE	67	Mar-07	Nov-05
Archer Daniels Midland Co.	Platte, NE	100	Nov-96	Nov-95
Archer Daniels Midland Co.	Platte, NE	300	Jul-14	Jan-11
Plymouth Energy LLC	Plymouth, IA	50	Feb-13	Oct-10
Southwest Iowa Renewable Energy LLC	Pottawattamie, IA	110	Feb-13	Feb-11
Marquis Energy LLC	Putnam, IL	100	May-11	Sep-10
Poet Biorefining-Leipsic	Putnam, OH	60	Jan-12	Oct-10
Cardinal Ethanol LLC	Randolph, IN	100	Oct-12	Oct-10
Highwater Ethanol LLC	Redwood, MN	55	Jul-13	Aug-12
Nesika Energy LLC	Republic, KS	10	Oct-11	Oct-10
Kansas Ethanol LLC	Rice, KS	55	May-12	Jan-11
Hankinson Renewable Energy LLC	Richland, ND	110	10/2/12	Aug-10
Agri-Energy LLC	Rock, MN	21	Feb-03	May-01
United Ethanol LLC	Rock, WI	55	Jan-11	Sep-09
White Energy Russell LLC	Russell, KS	50	Oct-05	Sep-04
Mid-Missouri Energy Inc.	Saline, MO	40	Jan-09	Oct-07
Abengoa Bioenergy of Kansas LLC	Sedgwick, KS	25	Dec-06	Sep-05
Poet Biorefining-Fostoria	Seneca, OH	68	Sep-12	Aug-11
Arkalon Energy LLC	Seward, KS	110	Dec-11	Aug-10
Siouxland Energy & Livestock Co-op	Sioux, IA	55	Dec-05	Apr-04
Redfield Energy LLC	Spink, SD	50	Dec-10	Nov-09
Center Ethanol Co. LLC	St. Clair, IL	50	Feb-12	Oct-10
Marysville Ethanol LLC	St. Clair, MI	50	Oct-11	May-11
Red Trail Energy LLC	Stark, ND	50	Dec-10	Jul-09
Adkins Energy LLC	Stephenson, IL	43	Aug-06	Sep-05

Table 3.A1. (cont'd)

Plant Name	Location (County, State)	Capacity*	Production Start Date	Construction Start Date
Lincolnway Energy LLC	Story, IA	50	May-10	Nov-08
Chippewa Valley Ethanol Co. LLLP	Swift, MN	45	Apr-00	Jun-99
Western Plains Energy LLC	Thomas, KS	45	Jan-08	May-07
NuGen Energy LLC	Turner, SD	110	Feb-12	Mar-11
Poet Biorefining-Chancellor	Turner, SD	45	Mar-07	Apr-06
Poet Biorefining-Chancellor	Turner, SD	100	Mar-12	Nov-11
Poet Biorefining-Caro	Tuscola, MI	50	Nov-06	Jul-05
Green Plains -Ord	Valley, NE	50	Jul-13	Dec-09
Poet Biorefining-North Manchester	Wabash, IN	68	Sep-12	Jul-11
Guardian Energy LLC	Waseca, MN	110	Aug-13	Jun-11
Cargill Inc.	Washington, NE	85	Apr-99	Jul-97
Poet Biorefining-Gowrie	Webster, IA	60	May-10	Apr-09
Valero Renewable Fuels LLC	Webster, IA	110	Oct-09	Jul-08
Green Plains-Bluffton	Wells, IN	110	Sep-12	Nov-11
Utica Energy LLC	Winnebago, WI	52	Apr-07	Jun-06
Poet Biorefining-Hanlontown	Worth, IA	55	Feb-08	Apr-07
Corn LP	Wright, IA	50	Dec-09	Oct-08
Abengoa Bioenergy Corp.	York, NE	55	Dec-97	Mar-96

*Ethanol plant capacity levels were taken from Ethanol Producer Magazine.

Table 3.A2. Average Impact of Ethanol Plants on Employment Levels 1990-2007: Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Construction	-763.1***	-759	147.2			
	(115.5)	(2,425)	(98.89)			
Production	-991.2***	-985	563.4***	-873.2***	-867.6	510.6***
	(94.49)	(1,985)	(93.48)	(92.79)	(1,949)	(85.98)
Observations	227,664	227,664	227,664	227,664	227,664	227,664
Number of Counties	1,054	1,054	1,054	1,054	1,054	1,054
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
State-Month Fixed Effect	No	Yes	Yes	No	Yes	Yes
County Trend, Linear	No	No	Yes	No	No	Yes
Cluster	County	County	County	County	County	County

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.A3. Average Impact of Ethanol Plants on Employment Levels 1990-2007: Non-Urban, High-Corn Counties

	(1)	(2)	(3)	(4)	(5)	(6)
Construction	-61.41***	-60.04	-77.52***			
	(21.52)	(137.5)	(17.34)			
Production	117.5***	118.8	104.9***	128.9***	129.9	134.4***
	(17.53)	(112.1)	(16.54)	(17.07)	(109.1)	(15.17)
Observations	75,072	75,072	75,072	75,072	75,072	75,072
Number of Counties	368	368	368	368	368	368
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
State-Month Fixed Effect	No	Yes	Yes	No	Yes	Yes
County Trend, Linear	No	No	Yes	No	No	Yes
Cluster	County	County	County	County	County	County

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.A4. Dynamic Impact of Ethanol Plants on Local Employment Levels 1990-2007: Non-Urban, High Corn Counties

	(1)	(2)	(3)
First half year since production started	70.80	66.06**	99.94***
	(73.98)	(30.54)	(19.04)
1 year of production	51.65	60.63	95.60***
	(87.37)	(37.58)	(23.96)
1.5 years of production	114.4	105.4***	143.7***
	(97.86)	(32.97)	(28.98)
2 years of production	144.3	153.7***	214.2***
	(95.96)	(38.17)	(39.07)
2.5 years of production	158.7	150.8***	198.9***
	(115.6)	(36.53)	(41.83)
3 years of production	217.1*	227.2***	269.8***
	(128.9)	(47.98)	(47.92)
3.5 years of production	279.7*	272.7***	298.3***
	(155.8)	(49.67)	(49.31)
4 years of production	348.5**	360.6***	342.1***
	(163.6)	(50.73)	(51.92)
4.5 years of production	375.8**	370.1***	326.3***
	(169.2)	(40.28)	(57.29)
5 years of production	349.8**	359.5***	345.2***
	(167.6)	(40.26)	(51.96)
5+ years of production	276.6	276.1***	285.1***
	(200.4)	(54.29)	(50.40)
Observations	57,408	57,408	57,408
Number of Counties	368	368	368
Time Fixed Effect	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes
State-Month Fixed Effect	No	Yes	Yes
County Trend, Linear	No	No	Yes
Cluster	County	County	County

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3.A5. Dynamic Impact of Ethanol Plants on Local Employment Levels 1990-2007 Including Leads: Non-Urban, High Corn Counties

	(1)	(2)	(3)
3 years before production	-55.19	-56.69**	24.09
	(67.91)	(27.91)	(19.71)
2.5 years before production	-92.80	-90.30***	-3.975
	(66.34)	(26.40)	(19.11)
2 years before production	-81.04	-82.83***	9.369
	(72.74)	(29.07)	(25.47)
1.5 years before production	-90.47	-86.51***	0.299
	(72.92)	(26.74)	(24.62)
1 year before production	-18.23	-20.46	64.35**
	(81.12)	(28.68)	(29.68)
First half year before production starts	-18.55	-13.87	65.22**
	(79.88)	(31.02)	(29.04)
First half year since production started	45.61	41.33	128.0***
	(98.59)	(38.09)	(33.53)
1 year of production	24.92	34.34	126.1***
	(109.8)	(44.38)	(38.23)
1.5 years of production	86.45	78.02*	176.8***
	(118.0)	(39.75)	(45.75)
2 years of production	115.3	125.2***	249.4***
	(111.7)	(42.22)	(50.68)
2.5 years of production	128.6	121.2***	236.3***
	(131.4)	(42.36)	(56.32)
3 years of production	186.3	196.9***	309.8***
	(145.4)	(52.62)	(61.25)
3.5 years of production	248.2	241.8***	341.2***
	(170.5)	(55.23)	(64.78)
4 years of production	317.9*	330.6***	386.9***
	(176.8)	(55.41)	(68.53)
4.5 years of production	346.0*	340.8***	373.8***
	(180.4)	(45.31)	(75.62)
5 years of production	319.5*	329.9***	394.5***
	(177.2)	(45.48)	(69.87)
5+ years of production	240.0	240.2***	340.8***
	(212.5)	(64.61)	(71.82)

Table 3.A5. (cont'd)

	(1)	(2)	(3)
Observations	61,824	61,824	61,824
Number of Counties	368	368	368
Time Fixed Effect	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes
State-Month Fixed Effect	No	Yes	Yes
County Trend, Linear	No	No	Yes
Cluster	County	County	County

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3.A6. Average Impact of Ethanol Plants on Own and Adjacent County Employment Levels 1990-2011: Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Construction	23.1	56.2	-49.2			
	(265.5)	(242.5)	(185.3)			
Production	161.6	199.3	213.3	132.2	107.4	117.6
	(454.79)	(585.45)	(681.8)	(397.4)	(149.3)	(83.4)
Adjacent County Construction	1.2	-24.3	-58.2			
	(34.3)	(71.0)	(99.2)			
Adjacent County Production	15.4	6.8	-22.5	62.2	34.3	-55.3
	(99.8)	(82.3)	(150.3)	(100.2)	(190.2)	(223.1)
Observations	75,072	75,072	75,072	75,072	75,072	75,072
Number of Counties	368	368	368	368	368	368
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
State-Month Fixed Effect	No	Yes	Yes	No	Yes	Yes
County Trend, Linear	No	No	Yes	No	No	Yes
Cluster	County	County	County	County	County	County

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

REFERENCES

REFERENCES

Angrist, J. D. and J. S. Pischke. 2009. *Mostly Harmless Econometrics: an Empiricists Companion*, Princeton, NJ: Princeton University Press.

Arellano, M. 1987. "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economic Statistics*, 49, 431-434.

Blanco, L. and M. Isenhouer. 2010. "Powering America: The Impact of Ethanol Production in the Corn Belt States." *Energy Economics*, 32(6), 1228-1234.

Buchanan, P. (1999) "A Family Farm Bill of Rights." Retrieved April 4, 2012, from http://buchanan.org/blog/pjb-a-family-farm-bill-of-rights-327.

Cascone, R. 2007. "Biofuels: What Is Beyond Ethanol and Biodiesel? - Emerging Technologies to Produce Gasoline and Diesel from Renewable Biological Resources Will Reshape the Agricultural, Chemical and Refining Industries." *Hydrocarbon Processing*, 86(9), 95-109.

Charles, M. B., R. Ryan; N. Ryan and R. Oloruntoba. 2007. "Public Policy and Biofuels: The Way Forward?" *Energy Policy*, 35(11), 5737-5746.

Cotti, C. and M. Skidmore. 2010. "The Impact of State Government Subsidies and Tax Credits in an Emerging Industry: Ethanol Production 1980-2007." *Southern Economic Journal*, 76(4), 1076-1093.

Du, X. D., D. J. Hayes and M. L. Mallory. 2009. "A Welfare Analysis of the Us Ethanol Subsidy." *Review of Agricultural Economics*, 31(4), 669-676.

Edmiston, K. D. 2004. "The Net Effects of Large Plant Locations and Expansions on County Employment." *Journal of Regional Science*, 44(2), 289-319.

Elobeid, A and Hart, C. 2007. "Ethanol Expansion in the Food versus Fuel Debate: How Will Developing Countries Fare?" *Journal of Agricultural and Food Industrial Organization*, 5, 1542 – 1585.

Ethanol Producers Magazine 2012. "Ethanol Plants: Plant List." Retrieved March 14, 2012, from http://www.ethanolproducer.com/plants/listplants/USA/

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, 37(3), 424-438.

Hahn, Robert and Caroline Cecot. 2009. "The Benefits and Costs of Ethanol: An Evaluation of the Government's Analysis." *Journal of Regulatory Economics*, 35(3), 275-295.

Hanson, G. H. 2001. "Us-Mexico Integration and Regional Economies: Evidence from Border-City Pairs." *Journal of Urban Economics*, 50(2), 259-87.

Hart, J. F. 1986. "Change in the Corn Belt." *Geographical Review*, 76(1), 51-72.

Hodge, T.R. 2011. “The Effect of Ethanol Plants on Residential Property Values: Evidence from Michigan.” *Journal of Regional Analysis and Policy*, 41(2), 148 – 167.

Hueth, B. and Walker W.D. 2008. “Local Ownership in Biofuels Production: A Strategy for Rural Development?” *Choices*, 23(4), 46-48.

Irwin, E.G., A.M. Isserman, M. Kilkenny, and M.D. Partridge. 2010. “A Century of Research in Rural Development and Regional Issues.” *American Journal of Agricultural Economics*, 92(2), 522-553.

Langpap, C. and J. J. Wu. 2011. "Potential Environmental Impacts of Increased Reliance on Corn-Based Bioenergy." *Environmental & Resource Economics*, 49(2), 147-171.

Lankoski, J. and M. Ollikainen. 2011. "Biofuel Policies and the Environment: Do Climate Benefits Warrant Increased Production from Biofuel Feedstocks?" *Ecological Economics*, 70(4), 676-687.

Leistritz, L. and Hodur, N. “Local and Regional Economic Impacts of Biofuel Development” in M. Khanna (ed.), *Transition to a Bioeconomy: Environmental and Rural Development Impacts*, Proceedings of the Farm Foundation/USDA Conference, St. Louis, Missouri, October 15-16, 2008, Farm Foundation, Oak Brook, IL.

Liska, A. J., H. S. Yang, V. R. Bremer, T. J. Klopfenstein, D. T. Walters, G. E. Erickson and K. G. Cassman. 2009. "Improvements in Life Cycle Energy Efficiency and Greenhouse Gas Emissions of Corn-Ethanol." *Journal of Industrial Ecology*, 13(1), 58-74.

Low, S. A. and A. M. Isserman. 2009. "Ethanol and the Local Economy Industry Trends, Location Factors, Economic Impacts, and Risks." *Economic Development Quarterly*, 23(1), 71-88.

McNew, K. and D. Griffith. 2005. "Measuring the Impact of Ethanol Plants on Local Grain Prices." *Review of Agricultural Economics*, 27(2), 164-180.

Meyer, B. D. 1995. "Natural and Quasi-Experiments in Economics." *Journal of Business & Economic Statistics*, 13(2), 151-161.

National Bureau of Economic Research. 2012. “US Business Cycle Expansions and Contractions.” Accessed at <http://www.nber.org/cycles.html> January, 2012.

Parcell, J.L. and Westhoff P. 2006. “Economic Effects of Biofuel Production on States and Rural Communities.” *Journal of Agriculture and Applied Economics*, 38(2), 377-387.

Renewable Fuels Association. 2012. “Industry Statistics” Retrieved April 4, 2012, from http://www.ethanolrfa.org/pages/statistics

Runge, C.F. and Senauer, B. 2007. “How Biofuels Could Starve the Poor.” *Foreign Affairs*. pp. 41 – 53.

Stevens, B.H. and Lahr, M.L. 1988. “Regional Economic Multipliers: Definition, Measurement and Application.” *Economic Development Quarterly,* 2, 88-96.

Swenson, D. and Eathington, L. 2006. “Determining the Regional Economic Values of Ethanol Production in Iowa Considering Different Levels of Local Investment.” Ames: Iowa State University.

Swenson D. 2006. “Input-Outrageous: The Economic Impacts of Modern Biofuels Production.” Paper presented at the Mid-Continent Regional Science Association and the Biennial IMPLAN National Users Conference, Indianapolis, IN. Retrieved February 2011, from http://www.econ.iastate.edu/outreach/menuCommOut reach .asp?code=3C

Timilsina, G. R. and A. Shrestha. 2011. "How Much Hope Should We Have for Biofuels?" *Energy*, 36(4), 2055-2069.

Tokgoz, S., A. Elobeid, J. Fabiosa, D. J. Hayes, B. A. Babcock, T. H. Yu, F. X. Dong and C. E. Hart. 2008. "Bottlenecks, Drought, and Oil Price Spikes: Impact on U.S. Ethanol and Agricultural Sectors." *Review of Agricultural Economics*, 30(4), 604-622.

Ugarte, Dgdlt, B. C. English and K. Jensen. 2007. "Sixty Billion Gallons by 2030: Economic and Agricultural Impacts of Ethanol and Biodiesel Expansion." *American Journal of Agricultural Economics*, 89(5), 1290-1295.

United Nations Environment Programme. 2009. “Towards Sustainable Production and Use of Resources: Assessing Biofuels.” Online at http://www.unep.fr/scp/rpanel/pdf/assessing_biofuels_full_report.pdf

U.S. Environmental Protection Agency (EPA) April 2007. “Regulatory Impact Analysis: Renewable Fuel Standard Program.” Prepared by: Assessment and Standards Division. EPA420-R-07004. Available at http://www.epa.gov/oms/renewablefuels/420r07004.pdf. Accessed February, 2012.

Whistance, J. and W. Thompson. 2010. "How Does Increased Corn-Ethanol Production Affect Us Natural Gas Prices?" *Energy Policy*, 38(5), 2315-2325.

Wolfers, J. 2006. "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results." *American Economic Review*, 96(5), 1802-1820.

Wooldridge, J.M. 2006. *Introductory Econometrics: A Modern Approach*, 3rd ed. New York: Thompson.

Yacobucci, B. 2007. “Ethanol and Biofuels.” *Congressional Research Service*. The Library of Congress.

Zhang, Z. B., L. Lohr; C. Escalante and M. Wetzstein. 2010. "Food Versus Fuel: What Do Prices Tell Us?" *Energy Policy*, 38(1), 445-451.