FORECASTING COUNTY-LEVEL UNEMPLOYMENT

ACCOUNTING FOR SPATIAL CORRELATION

A Dissertation

Presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

DUSTY SWEET

Dr. Peter Mueser, Dissertation Advisor

DECEMBER 2011

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

FORECASTING COUNTY-LEVEL UNEMPLOYMENT ACCOUNTING FOR SPATIAL CORRELATION

presented by Dusty Sweet,

a candidate for the degree of doctor of philosophy, and hereby certify that, in their opinion, it is worth of acceptance.

_____

Professor Peter Mueser

_____

Professor Joseph Haslag

_____

Professor J. Isaac Miller

_____

Professor Michael Podgursky

_____

Professor Christopher Wikle

DEDICATIONS

I would like to thank my parents for giving me the ability, my grandparents for giving me the drive, my sister Kerri for giving me the humility, and my wife Michelle, whose support and love made this possible.

ACKNOWLEDGEMENTS

I would like to thank first and foremost Dr. Peter Mueser for his patience, his wisdom, and his guidance, as well as all the valuable help from the committee. I would also like to think Dr. Michael McCracken for starting me on the path and Ms. Lynne Riddell for keeping me there.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABSTRACT

This paper analyzes the effect of including a spatial component in models that predict monthly county-level unemployment in Missouri.

The initial analysis seeks to explore the general spatial structure of unemployment using spatial covariance measures, correlation statistics, and Moran's I, which is the most commonly used measure of spatial association. The effects of counties being similar in nature are also considered.

Following these analyses, regression models are estimated that predict monthly unemployment for counties using both spatial and non-spatial, both fitting separate models for each county and model that constrain parameters to be similar across counties. In addition to investigating the importance of spatial ties, models that investigate the extent to which similarities in industry or related characteristics can be used in models.

Finally, the accuracy of out-of-sample forecasts is examined for both spatial and non-spatial models. It is found that while evidence of a spatial component does appear in the results, its ability to contribute to statistical modeling or forecasting accuracy is mixed.

1. INTRODUCTION AND LITERATURE REVIEW

In the last twenty years an increasing number of spatial econometrics papers have appeared throughout the economics literature. Coupled with the emergence of data compiled at the state level and below (e.g, metropolitan statistical areas, counties) it is naturally applied to economic research.

When this new focus in econometrics is applied to specific areas of economics such as unemployment rates, it holds great promise for many areas. Several recent papers have illustrated that unemployment contains a spatial dependence at multiple levels of aggregation. There is recent work on spatial correlation in unemployment at both the state and county levels as well as levels that lie in between these two. Unemployment is a common measure of an economy's overall health. The ability to measure and forecast it more accurately may improve predictions about the future of the economy. On a more local level, if the spatial structure of unemployment can be exploited properly it will allow regional planning authorities to be better prepared for changing economic conditions.

The main goal of this study is to explore the nature of the spatial dependence in unemployment at the county level and to investigate whether spatial models of monthly county unemployment can be developed to improve predictions. Using data for counties in Missouri, this study starts by analyzing the overall spatial structure of unemployment in Missouri using correlation coefficients and covariances. The relationships between county unemployment rates are analyzed in terms of physical distance between counties, as well as metropolitan status, agricultural concentration, manufacturing concentration,

and educational concentration. Moran's I (the most commonly used measure of spatial dependence) is also calculated using alternative weighting structures based on geographic distance. Moran's I is calculated using simple contiguity-based weights, an exponential weighting function, and weights based on the inverse of distance between counties. Statistical significance of Moran's I for these different weighting schemes is also analyzed.

The next analysis undertaken after these exploratory exercises constructs both spatial and nonspatial models predicting monthly county-level unemployment in Missouri using ordinary least squares. These models are estimated using all data available. In all of the models estimated (nonspatial and spatial alike), the coefficients are both restricted to be the same across counties and allowed to vary across counties. There are also situations where there is a mixture of these specifications; some models allow a few coefficients to vary across counties while simultaneously restricting other coefficients to be the same across counties. Goodness of fit measures and related statistics are calculated to determine which structure is most likely to fit the Missouri county-level unemployment. This step allows for both temporal and spatial lags, and there are models with varying number of both temporal and spatial lags to measure the importance of different numbers of lags.

Based on the results of the models, the analysis turns to examining whether there are mediating factors that explain why some counties have more spatial dependence with their neighbors than others. Some of the characteristics that might affect these relationships are the size of the labor force in the county, the types of industry that are present in the county, and whether or not the county has out-of-Missouri counties as

predictors. For example, analysis is done to determine whether unemployment in counties with larger labor forces (relative to the other counties in the state) has a stronger spatial component than unemployment in counties with smaller labor forces.

Following this analysis, both spatial forecasting models and nonspatial forecasting models are built for each county-level unit and compare the forecasting accuracy of the two using out-of-sample forecasts estimated from fixed numbers of observations. This differs from the previous chapter in that this analysis estimates forecasting accuracy from fewer observations than the full sample. The previous chapter neither calculates forecasting accuracy nor uses anything less than the full number of observations in the dataset. The initial step is to estimate spatial and nonspatial models using a consistent sample size and examine the effects of restricting county coefficients to be constant across counties as well as allowing them to vary across counties, in large part corresponding to models examined in the prior chapter's analysis. This not only gives a preliminary sense of which structure is more accurate but also produces baseline estimates for comparison with those produced in later analyses.

The next move is to consider different numbers of temporal lags as predictors as well as different numbers of observations, again based on out-of-sample forecasts as explained in the previous paragraph. The analyses relating to labor force size of the county, industrial makeup of the county, and whether or not the county has out-of-Missouri counties as predictors are repeated as well.

When this study is complete, the overall structure of unemployment at the county level in Missouri will have been examined from several viewpoints. The general measures of correlation and covariance gives an initial sketch. The analyses that follow

use models based on the full sample of data, and the analysis is concluded with forecasting accuracy measures based on coefficients estimated from subsets of the available data. With these results in hand, we will not only have an idea of what, if any, spatial structure exists in county-level unemployment in Missouri but also the extent to which forecasting accuracy can be improved using measures based on this spatial structure.

The remainder of the present chapter is organized in the following fashion. The first sections introduce the general idea of spatial correlation and its structure and discuss some theoretical work both in and out of Economics. From there, some of the recent applied work using spatial models is discussed. Once the general theoretical and applied models have been discussed, the methodology that I have used to analyze spatial correlation in county-level unemployment in Missouri is laid out.

1.1 Spatial Correlation

Cliff and Ord (1973) define spatial correlation as the case where values or characteristics in one area affect or are affected by the same particular characteristic in one or more of its neighbors. A simple example is where a variable in one region is related to the same variable in a geographically proximate area. This is similar to serial correlation, but one difference distinguishes them. Whittle (1954) notes that in time-series data, there is a natural order present. There is a strong presumption as to the direction of the causation, moving from the past to the present, or from the present to the future. With spatial correlation, the direction of the causation is less clear. The effects of the dependence can move in as many directions as there are spatial units being measured. Given the proper conditions, there are some similarities between a time series model and

a simple spatial model. A basic first-order autoregressive model in time is denoted by the following:

$$\text{Equation 1.1 } Y_t = \rho Y_{t-1} + \varepsilon_t$$

where Y_t and Y_{t-1} are the values of Y at times t and t-1, respectively. For a given location, ρ is the correlation coefficient, and ε_t is the residual term.

A one-directional spatial model would be analogous to the AR(1) model. However this would imply spatial dependence in only one direction. While this is not impossible, it is not very likely. A more realistic analog to the AR(1) model in a spatial context would be the following:

$$\text{Equation 1.2 } Y_i = \beta Y_{i-1} + \delta Y_{i+1} + \varepsilon_i$$

where Y_{i-1} and Y_{i+1} refer to the spatial neighbors on either side of Y_i. Although the properties of this first-order spatial model do not translate perfectly from time-series to a spatial context, they are similar. This extra direction of the correlation makes the spatial structure more technically complicated.

Consider the following estimation equation:

$$\text{Equation 1.3 } Y_{i,t} = \alpha + \rho Y_{i,t-1} + \beta_1 Y_{i-1,t-1} + \ldots + \beta_k Y_{i-k,t-1} + \varepsilon_{i,t}$$

where Y_{it} is the unemployment rate for particular spatial unit i (e.g., county) at time t, and the subscripts i to i-k denote the rates in different locations. k indicates the number of neighbors that are included in the model. The subscript t-1 indicates a lag in time. The temporal lag is not restricted to t-1; depending on the data and the conditions of the model it may extend to prior periods, with the number of temporal lags based on the situation. For example, $Y_{i-2,t-3}$ would be the unemployment rate in the second neighbor measured three periods prior to time t. α is a constant term that may be constrained to

equal zero in some cases. ρ is the coefficient on an area's own temporal lag, and $\beta_1 \ldots \beta_k$ are coefficients on the temporally lagged values of the bordering spatial units. This illustrates a general space-time model that we use as both a starting point for other models in the literature as well as the base model in our research.

This discussion leads to the problem of which neighbors should be used. Arbia (2006) describes the formation of a neighborhood, which is defined as consisting of those spatial units that are being used as predictors for the unit under study. He considers three ways that neighborhoods are usually formed. The first is a critical cut-off neighborhood, where only spatial units that are within a particular distance are considered part of the neighborhood. A similar method is the nearest neighbor definition, which simply identifies the closest unit. The third discussed by Arbia is the contiguity-based neighborhood, in which those units sharing a border with the area under study to be considered part of the neighborhood. The main analysis in this paper will use the contiguity-based approach for determining the neighborhoods, although we will consider distance-based approaches as well.

Another potential problem is correlation between the predictors. The premise of spatial correlation is that neighbors' values of a variable are dependent upon each other in some way. If the predictors are correlated with the area of interest, the predictors themselves are expected to be related to each other (since they would be neighbors of one another). Semple and Green (1984) describe clustering as a possible solution to this problem. The idea behind clustering is that if areas are similar in enough ways, they can be fused together to form a single unit, reducing the complexity of the model. Their rationale for using clustering is primarily to decrease the dimensions of the equation.

However, another result is that several independent variables are compressed into one, which eliminates the possibility of multicollinearity, and the associated difficulties in interpreting coefficients. As an example, if county i and county j are predictors for county k and the two (i and j) are closely correlated, estimates will be unstable, and in the extreme case the rank condition may fail. If the unemployment rates for the two counties are combined into a single variable, the problem of collinearity between the predictors is alleviated. Several criteria exist when deciding how to cluster the areas together; the method used to cluster is dependent on the situation.

Once it is determined that spatial correlation is present, the next concern is to determine the nature of the dependence. Anselin and Bera (1998) discuss the difference between the spatial lag and spatial error model. In order to first illustrate the spatial lag model for a single spatial unit, note the following equation:

Equation 1.4 $Y_i=\alpha+\beta \sum_{j=1}^{n} w_{ij}Y_j + \sum_{k=1}^{K} \theta_k X_{ik} +\varepsilon_i$

In this setup, Y_i is the dependent variable and the Y_j's are the spatially lagged values. w_{ij} is the weight given to Y_j, indicating how it is spatially related to i, β_i is the spatial dependence coefficient, and X_{ik} is a vector of exogenous predictors with their corresponding coefficients (θ_k). Similar to the time-series case, the parameter of interest for spatial dependence is the β. Because it is the result of the inclusion of the neighbors which are said to be spatially lagged, this is called the spatial lag model. This is similar to Equation 1.3 with the exception that this setup has exogenous predictors.

An alternative to this specification is to include the spatial component in the residuals of the main model. This particular setup is called the spatial error model. Consider the standard regression equation:

Equation 1.5 $Y_i = \alpha + \sum_{k=1}^{K} \theta_k X_{ik} + \varepsilon_i$

Note that no spatial component is included at this stage. The idea behind the spatial error model is that the residuals follow a process that exhibits spatial correlation. Consider the following error structure for a single spatial unit:

Equation 1.6 $\varepsilon_i = \lambda \sum_{j=1}^{n} w_{ij}\varepsilon_j + u_i$

Here the λ is measuring the degree of spatial correlation that is present in the residuals, w_{ij} identifies the relative spatial proximity of locations i and j, and u_i is an independent error term. Anselin and Bera (1998) also show that these two types can be combined into a single model, which can be estimated provided the X_{ik} contains at least one exogenous variable save the constant. They also note that this can be modeled by separating the error term into an error-in-components structure.

Stetzer (1982) reports results based on the estimation of a model as well as using a model for forecasting based on the differing weighting schemes in constructing w_{ij} using simulated data. Using a series of Monte Carlo estimations, he finds that identifying the proper area is the most important goal, that is weighting the predictors so that only the relevant ones are included is more important than finding the correct functional form of the weights. His conclusion about the goal of the weighting matrix is the same when addressing the issue of forecasting as well as estimation.

Stetzer's analyses can be illustrated in terms of the following equation:

Equation 1.7 $Y_{i,t+1} = \alpha + \rho Y_{i,t} + \beta_1 Y_{i-1,t} + \ldots + \beta_k Y_{i-k,t} + \varepsilon_{i,t}$

This is very similar in structure to Equation 1.3, except the temporal subscripts are changed to indicate that the value of Y_i in time t+1 is now forecasted from all available data in time t.

He finds that in most cases identifying the proper area is critical. The exception in the forecasting case is when this area is large. If the proper area is populated by a large number of neighbors, then it becomes more important to use a distance decay setup. Unfortunately the author is not clear on his definition of large with respect to a general case. What this implies for this paper is that the goal should be to accurately identify which counties should be included in independent variables.

One way to begin determining the level of spatial dependence in the data is by use of a chloropleth map. A chloropleth map is simply a map of the regions that is color-coded based on the values under study. While this is not a formal statistical procedure for measuring spatial dependence between units, it gives a nice visual as to whether or not spatial correlation is present (Cressie, 1993).

A formal statistical measure of overall spatial dependence is Moran's I statistic. Moran's I statistic is a measure that tests for spatial association. Taking x_i as the measure of location i, Moran's I is calculated as follows:

Equation 1.8 $$I=\frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

where w_{ij} is the weighting matrix, $\bar{x}$ is the mean for x_i in the sample, and n is the number of locations. Because this is concerned in measuring the relationship between neighbors, $w_{ii} = 0$; that is, the association between a spatial unit and itself is omitted. As above, the weighting matrix is constructed so that more proximate locations receive larger weights and more distant ones lower weights. It compares the spatial variation to the total variation to determine if a significant level of variation is due to spatial correlation. The values of Moran's I normally range from -1 to 1, where positive (negative) values indicate that areas close to one another are more (less) likely to have similar values.

Another often used method of determining spatial correlation is Geary's C statistic. Geary's C is calculated as follows:

$$\text{Equation 1.9 } C=\frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}(x_i-\bar{x})^2}{2\left(\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}\right)\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Geary's C uses the same variables as Moran's I. Geary's C differs from Moran's I in that Geary's C normally takes values between 0 and 2. Values closer to 0 indicate positive spatial correlation, and those closer to 2 indicate negative spatial correlation. A value of 1 implies that there is no spatial correlation. Keller and Shuie (2007) use Geary's C to measure the level of spatial correlation in their data, in addition to Moran's I. They find that the spatial relationship between Chinese rice prices is strong, whether based on Geary's C or Moran's I statistic.

While several alternatives exist to test the overall measure of spatial correlation, Moran's I is by far the most often used, although other measures are occasionally used. Those who use Geary's C argue that it is more sensitive to local differences as the numerator is built by pairwise comparisons of regions as opposed to differences from the mean (Cressie, 1993). It is overwhelmingly the case in the literature that any time multiple measures are used, the substantive results do not differ significantly.

To this point, the models discussed have dealt with either the temporal aspect or spatial aspect of the variable of interest as in Equation 1.3. What about models accounting for both simultaneously? The literature regarding models in both space and time (or spatiotemporal models) is sparse. While the nature of spatial models is important, the model I work with has the added element of temporal dependence. This brings in the necessity of discussing some of the features of space-time models. Banerjee, Carlen and Gelfand (2004) argue that the use of time-series data within the scope of

spatial statistics makes it necessary from a modeling perspective to first determine whether one is dealing with spatial data moving through time or temporal data moving through space. Banerjee, Carlen and Gelfand (2004) give the example of pollution levels in given areas in multiple time periods as spatial data moving through time. The analysis will examine unemployment rates in various locations and observe how they change as time passes, which is similar to the pollution example. An example of the other process would be home sales. From one year to the next, we observe different locations (houses) that are put on the market for sale. This is where we move through different time periods and observe how the locations change, as the location of which homes are put up for sale is different from one time period to the next. Banerjee, Carlen and Gelfand propose modeling the residuals of a space-time model by breaking them into parts. They show three possibilities in a theoretical context: one with only a temporal component, one with only a spatial component, and a third with both types.

1.2 Applications of Spatial Models in Literature

The area of spatial econometrics is an emerging field of research. As noted above, Keller and Shiue (2007) consider the behavior of rice prices in Chinese prefectures of southwestern China based on a spatial model. Based on Moran's I and Geary's C, measures of spatial association, the authors show that rice prices do exhibit some level of spatial correlation. After confirming that a spatial correlation exists, they estimate separate models for each prefecture in their study. Their results suggest that there are distinct patterns of prices for different regions.

Baller, Anselin, Messner, and Deane (2001) study spatial correlation in an arena that is not traditionally studied by economists. Using county-level data from the United

States, they seek to build a spatial model to illustrate the relationship between counties' homicide rates. They start by using Moran's I to estimate an initial measure of the correlation and find that homicide is positively spatially correlated at the county level with values ranging from 0.36 to 0.42. This means that counties of high (low) homicide rates tend to cluster together. They build a model that includes both a spatial lag and a spatial error component. One model uses the five nearest neighbors and the other includes the ten nearest neighbors, with distance between county centroids as the measure of distance. The spatial lag component is captured by the inclusion of neighbors' homicide rates as independent variables, while the spatial error component is intended to capture any residual correlation that is coming from the neighboring county's homicide rate.

Another recent addition to the literature is by Kapoor, Kelejian, and Prucha (2006). They derive a theoretical procedure by which to estimate a spatial model with feasible generalized least squares (FGLS) using an error-in-components structure. Their model seeks to estimate the relationship between county income levels in Virginia. They take their theoretical results and use them to implement FGLS. After deriving several alternative estimators, they conduct a Monte Carlo experiment that is based on the county income for Virginia and show that the alternative methods do not lead to different results. This paper gives a blueprint for using FGLS to estimate a spatial model.

Zhou and Buongiorno (2006) also construct a space-time model to examine timber prices in the southern United States for twenty-one contiguous regions. The authors use a binary row-standardized contiguity matrix for their spatial weights. Their general method follows the procedure laid out by Pfeifer and Deutsch (1980), which is a three-stage procedure to construct a space-time autoregressive moving average

(STARMA) forecasting model. Zhou and Buongiorno first identify the lags that will be appropriate for their data, estimate the parameters on this model, and finally check the validity of the model they have estimated. Zhou and Buongiorno take the STARMA models that they build and make predictions of future values. They compare these predictions to univariate ARIMA models constructed for the same regions. It is an example of an application of a STARMA model in the existing literature. It also provides further evidence that, for predictive purposes, ignoring a spatial component that is present will lead to less accurate predictions.

1.3 Applications of Spatial Models in Unemployment

Turning to spatial dependence specifically related to unemployment, it seems reasonable that the unemployment in one county would have an effect on its neighbors. The magnitude of the effect would likely vary according to the source county of the shock. But it is also possible that there is correlation between counties that is not the result of direct causal impact. Consider the flood in Missouri in 1993. Unemployment in all counties affected by that event would experience increases in their rates. The increase caused by the flood would not be a result of one county's shock influencing another but of both regions experiencing the same condition. Nonetheless it is reasonable to hypothesize that an unemployment shock in one county would have a measurable impact in its neighboring counties.

Cracolici, Cuffaro, and Nijkamp (2007) analyze the provincial unemployment rates for Italy and find that a large portion of the differences in unemployment rates between regions is due to the differing conditions in the North and South. Given the low mobility of labor (particularly in the higher unemployment South) caused by increased

migration costs and increased costs of housing transactions, we may expect that the labor markets do not properly equalize wages between the regions. If an employment shock hits a region causing higher unemployment, we expect those people who are separated from their jobs to search for a new position first in their home region due to low search costs. If their search ends without finding a new job, we would expect them to expand their search to include neighboring areas, which has become easier with the recent expansion of the role in the internet with respect to job search. If a suitable job is found, they accept the position. Cracolici, Cuffaro, and Nijkamp (2007) describe here a process by which a market shock that increases unemployment in one region at one time can lead to increased unemployment in a neighboring region at some future time period.

Patacchini and Zenou (2007) also use Moran's I to measure spatial correlation in unemployment in their British Travel-To-Work Areas[1] (TTWAs) using a binary weight based on distance between the areas. They report values of Moran's I for five different five-year periods' average unemployment rates and find that the correlation increases as more recent data are used; their dataset ranges from 1985 to 2003. The authors look at several distance cutoffs that increase by increments of 20 kilometers in width, i.e., the distance cutoffs are 0-20 km, 0-40 km, etc., up to 120 km. For a given distance, two areas receive a non-zero weight if they are within the given cutoff criterion and a weight of 0 otherwise. The largest values of Moran's I are found in the smallest distance cutoff of 20 km. As Moran's I is calculated for the longer distance cutoffs, Moran's I decreases monotonically, showing that the spatial association largely reflects the impact of immediate adjacency. They also observe that for the earliest data in the sample (1985),

[1] A TTWA is a region with at least 3,500 people where at least 75% of the population that lives in the area also works in that area.

nearly all values of Moran's I are not statistically significant regardless of the distance cutoff. As they calculate the values for later periods, the values of Moran's I increase and many become statistically significant (even those at long distances), although it is still true that the shorter distances produce a higher value of Moran's I than those at longer distances. This implies that the spatial correlation in unemployment is becoming stronger through time.

Another example of spatial dependence in unemployment is from Bronars and Jansen (1987) who worked with monthly unemployment data at the county level. Collecting data from a section of the U.S. Midwest, they estimated a model by combining their counties to form regions of approximately the same size as counties and also aggregating the data for months to form quarterly observations. They subtracted the unemployment rate for each region at each period from the region's average rate over time, and also differenced out the national unemployment rate. The authors also included quarterly dummy variables to capture seasonality.

Their model used temporally lagged unemployment rates of neighboring counties to predict the unemployment rate of a specific area. The authors considered an equation very similar to Equation 1.3, and which is also similar in important respects to the analysis undertaken here. Their spatial units are not counties, although the divisions of their lattice use county positions as reference points and units are of similar size to counties. Their constant term has some predetermined elements, most notably the national unemployment rate and the average mean over time of Y_i. After calculating the parameters on the lags, they used these to simulate the effects of an unemployment shock in one county on surrounding regions. Their research showed that there is an inverse

relationship between the correlation of unemployment for a given pair of spatial units and distance between them. The authors found that a labor market shock can cause effects up to 25 miles from the county that experienced the shock. Using their data to construct a spatial correlation function, they found that once a region was farther than 25 miles the correlation coefficient was no longer significant. This supports the idea of using only using first- or second- order neighbors when predicting a region's unemployment rate. In other words, the prediction equation for the unemployment rate in a county located on the western border of Missouri would not benefit from including the unemployment rate in St. Louis County as one of its independent variables.

Feasel and Rodini (2002) look at county unemployment rates as well. The focus of their paper is on the role of demographic factors and migration, but they also employ a spatial component in their model of county unemployment in California. They employ Moran's I as a diagnostic check on the presence of spatial dependence between county unemployment rates. The authors do not report any of their numerical results for Moran's I, but do use it to conclude that spatial dependence does exist at the county level for their dataset.

Niebuhr (2003) builds a similar model to that of Feasel and Rodini (2002) but uses European regions instead of the United States regions. His paper is also focused on migration, but does work with spatial dependence between unemployment. His spatial weighting matrix is initially set up as a binary contiguity matrix, so that regions that share borders receive a weight of 1 and all others 0, but he further alters the weights so any region that shares a border with the dependent region will be weighted using a distance decay function ($w_{ij} = e^{-cd}$) where the rate of decay is determined by the distance between

centers of regions (d), a decay parameter (c). The weighting matrix is then row-standardized to 1. A positive value of Moran's I implies that spatial correlation is present in the data. Using various specifications, the calculations show Moran's I to range between 0.6 and 0.8. Their estimates of the decay parameter suggest that the spatial correlation has a half-life of roughly forty kilometers (24.8 miles), meaning that it is reduce to half of its original strength after this distance. This suggests that the "reach" of a county in affecting another county's unemployment rate may be somewhat greater than the 25 miles found by Bronars and Jansen (1987).

Longhi, Nijkamp, and Poot (2006) investigate the German labor market. The authors seek to determine how spatial differences between regions affect local real wages. They specify a wage equation for different regions in Germany that includes a spatial component as well as other control variables. The spatial version of their model for wages is analogous to the spatial lag model. To measure the level of spatial correlation present, the authors also use the Moran's I statistic (estimates range from 0.56 to 0.75) to determine the level of spatial dependence with an inverse of distance ($^1/_d$) as spatial weights. While their model is mostly concerned with the wage variation across regions, they do find that the unemployment rates in neighboring regions affect one another.

Conley and Topa (2002) also investigate the possibility of unemployment being correlated spatially. Their goal is to compare areas with similar unemployment rates that tend to cluster together with social and economic distance measures to see if the two are related. Using Census tract data from the city of Chicago, they build a model that not only tests the importance of physical distance but also that of travel time. The authors

suggest that raw physical distance is not necessarily appropriate, but that it might be more useful to consider travel time. They also include personal characteristics of location population such as race and ethnicity. People tend to locate themselves and exchange information with people of similar characteristics, and if there is a racial or an ethnic component of unemployment we would expect that these factors could be used to create a more meaningful distance measure. Another factor Conley and Topa include is occupational distance to capture the extent to which information that might be transmitted through networking.

The authors construct measures of distance for all of their measured characteristics (physical distance, travel time, race, and occupation). Physical distances and travel time are exactly as they seem. For racial and occupational distance, the authors determine how racially (occupationally) different one census tract is from another and calculate the Euclidian distance between the two tracts based on these metrics.

To construct their measure of spatial correlation, Conley and Topa (2002) specify that X_i is unemployment at location s_i. The theory says that as the distance between s_i and s_j becomes smaller, the values of X_i and X_j should have increasing correlation. Because of this the authors state that the covariance of these two X's is a function of the distance between the two. To get estimates of spatial autocorrelation, the authors set up the function listed below.

$$\text{Equation 1.10} \quad f(\delta)=\sum_{i=1}^{N}\sum_{j=1}^{N} W_N\left[\left|\delta-D_{ij}\right|\right](X_{S_i}-\bar{X})(X_{S_j}-\bar{X})$$

In this setup, W_N is a weighting function, δ is a distance that will determine whether or not an observation is "counted," D_{ij} is the distance between s_i and s_j, and $\bar{X}$ the mean of

the X's. If the absolute value of the difference between δ and D_{ij} is larger than a specified cutoff distance, then W_N is 0 and the observation is not counted. Otherwise, it will receive the same weight as all other neighbors being counted, where the sum of these weights will be one. From this setup they find that the correlation decreases as the distance between spatial units increases. They report autocorrelation functions based on their measures of distance and show that the physical distance correlations flatten to 0 after approximately 6-8 kilometers. While including ethnic and education parameters softens the distance correlation to a degree, they do find physical distance between tracts remains relevant.

Elhorst (2000) discusses the ramifications of constructing a model that has a first-order lag in both space and time. He begins by comparing and contrasting the analytical conditions for a model with only a temporal lag, a model with only a spatial lag, and one that includes both. He continues further with a theoretical display of various features of each model including variance-covariance structure, general forms of likelihood functions for each based on normality, and conditions on temporal stationarity. Using data from Europe on regional unemployment and the labor force participation rate, he estimates the effects of regional unemployment on the labor force participation rate in a model that includes both spatial and temporal lags. He goes through analytically the several ways that both spatial and temporal lags can be included in an econometric model and shows that in some cases several options are available for fitting a particular dataset (e.g., spatial lag model corrected for serial correlation, serial lag model corrected for spatial correlation).

Another paper that also builds a space-time model for unemployment comes from Di Giacinto (2006), which considers regional unemployment in Italy. Like Zhou and Buongiorno (2006), Di Giacinto constructs a model based on the work of Pfeifer and Deutsch (1980). His analysis also concludes that spatial correlation displays fairly rapid distance decay; this means there does not appear to be a lasting spatial effect beyond the first spatial neighbor.

While Genton and de Luna (2004) do not use any of the methods proposed by previously mentioned authors, their model considers both space and time in their evaluation of unemployment. They consider the appropriate lengths of both spatial lags and temporal lags (the length of the spatial lag is defined as how many neighbors are included as predictors) and build separate models for the nine U.S. Census regions including lagged values in both space and time, which is analogous to constructing separate county models within a given state.

1.4 Model and Methodology

The first step in the analysis done here is to acquire a preliminary understanding of the spatial structure of county-level unemployment. In chapter 2, I calculate contemporaneous spatial covariance functions following closely the structure of Equation 1.10 as well as a series of correlation coefficients based on changes in unemployment over the life of the dataset. While these two measures often yield similar results, the covariance function will give us a sense of how unemployment rates are related at any given point in time, whereas the correlation functions allow us to observe how changes in unemployment rates are related over time.

These two methods are applied to various distance and characteristic groups to determine what effects not only distance has on the spatial relationships but also how spatial relationships differ based on various characteristics, including manufacturing concentration, agricultural concentration, educational concentration, and metropolitan status. This allows us to see how a county's unemployment rates affect unemployment rates of similar counties.

While the measures of covariance and correlation are useful in painting an overall picture, the next move is to examine the most commonly used measure of spatial correlation, which is Moran's I. This will allow an initial view of the spatial structure of county-level unemployment in Missouri.

Once the preliminary analysis is complete, chapter 3 turns to fitting an appropriate model. The model I fit has elements of most of the previously-mentioned studies. I use monthly county-level unemployment rates in a model that captures spatial dependence to forecast future values of the unemployment rate. The model I estimate can be written with the general structure:

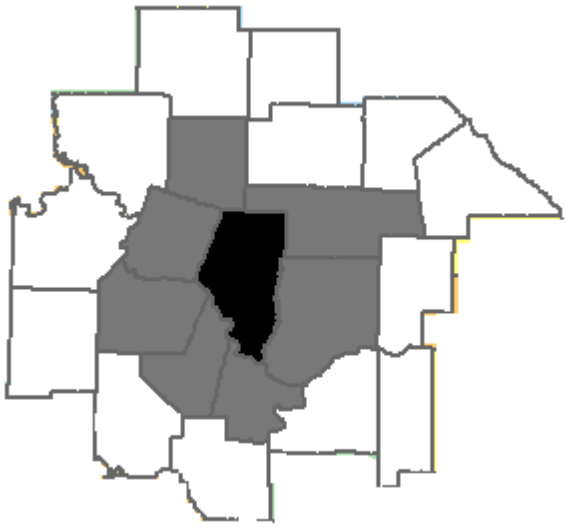
Equation 1.11 $$Y_{i,t}=\alpha_i+\sum_{r=1}^{q}\rho_{i,r}Y_{i,t-r}+\sum_{r=1}^{a}\beta_{i,r}Y_{j(i),t-r}+\sum_{r=1}^{b}\delta_{i,r}Y_{k(i),t-r}+\sum_{r=1}^{v}\theta_{i,r}U_{t-r}+\sum_{m=1}^{11}\eta_{i,m}D_{t,m}+\varepsilon_{i,t}$$

$Y_{i,t}$ is the unemployment rate in county i at time t. $Y_{i,t-r}$ is the lagged value, where r is the length of the temporal lag in months. Several values of q, a, and b, the lag maximum values, will be considered. $Y_{j(i)}$ and $Y_{k(i)}$ are the unemployment rates of the counties immediately adjacent to county i and those adjacent to counties in the set j but not adjacent to county i, respectively. The α_i is a constant term, and $\rho_{i,r}$, $\beta_{i,r}$, $\delta_{i,r}$, θ_i, and $\eta_{i,m}$ are coefficients, all of which are estimated. ε_i is the error term.

The data to be fitted are the monthly unemployment rates from the Bureau of Labor Statistics for January 1990 to November of 2006 for all counties and St. Louis City (a county-equivalent unit) in the state of Missouri. Unemployment for counties from states that border Missouri will be employed to calculate spatial measures. Two other items are included in the model. One is a set of 11 monthly dummies (for 12 months) to capture seasonality indicated by $D_{t,m}$. The original data are not seasonally adjusted, so this is a necessary addition. The second is the unemployment rate of a larger region (U_{t-r}) such as Missouri or the United States. The reason I include this is to capture any correlation between counties that is a result of larger factors than a county-level shock. A national recession would cause an increase in unemployment in all counties. If it is not accounted for in some way, the model may lead to a false positive of spatial correlation (or improper sign on the coefficient, improper magnitude, etc.).

I have estimated a model for each county separately as well as a single model for all counties in the state (and some hybrids of these two). I use Ordinary Least Squares (OLS) to build my forecast models. Figure 1.1 below helps to explain the calculation of the $Y_{j(i),t}$ and $Y_{k(i),t}$.

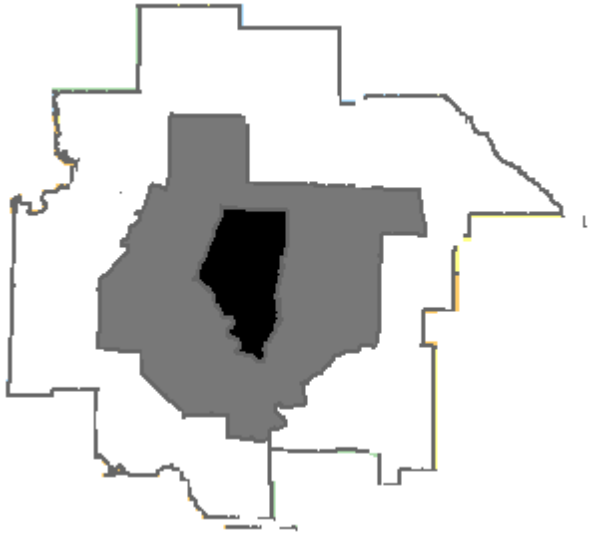
Figure 1.1 Boone County and its neighbors



As an example, consider Y_i for i as Boone county, which is black in the above illustration (the t is not essential to this portion of the discussion). The gray counties are those that share a border with Boone, and the white counties are those that share a border with the gray counties but do not share a border with Boone. $Y_{j(i)}$ is the unemployment rate of the gray area.

The procedure for calculating $Y_{k(i)}$ is identical, except it is the white counties. This would give the unemployment rate for the region that contains all second-order spatial neighbors that are not first-order neighbors. A pictorial representation of the three areas that will be used in this context is Figure 1.2.

Figure 1.2 Boone County and the neighboring supercounties



This method would give the same result as calculating a weighted average of the unemployment rates of the counties involved using labor force as the weights. Because county-level unemployment data are available for all states in the United States, $Y_{j(i),t}$ and $Y_{k(i),t}$ can be calculated for all Missouri counties.

It is important to make a distinction here between the model I fit and other models in the literature. Some models call for the parameter estimates to be the same across units. In contrast, my concern will focus on constructing a model that predicts unemployment. Since in some cases I am constructing a model for each county separately, there is no restriction that the parameters be the same across counties. For example, the coefficient on the state unemployment rate may be different for each county, so that a shock to the state unemployment rate has a different effect on different counties. I also have estimated a single model for all counties in the state to examine the effects (if any) of this restriction.

Why stop at including second-order neighbors? Bronars and Jansen (1987) find that unemployment shocks tend to travel approximately 25 miles. Outside of this radius the shock loses its strength. Our results[2] are consistent with this conclusion and thus nothing further than a second-order spatial neighbor is considered. Although no two counties in Missouri are the same size, most counties are approximately 30-40 miles between edges. We consider various weighting schemes that assign contiguity-based weights as well as some that give decreasing weights based on increasing distance. It turns out to be the case that a county this far away does not have any statistically meaningful impact on the county in question.

The choice to combine the neighbors into a single unit is based on the idea of clustering proposed by Semple and Green (1984). By combining the counties surrounding Boone into a single spatial unit, the dimension of the first-order neighbor is reduced from seven to one. This allows more degrees of freedom when estimating the model and removes the need to estimate parameters for individual adjacent counties, whose unemployment rates may display substantial intercorrelations. This aggregation approach is viable for unemployment rates. The number of people unemployed and in the labor force and total labor force from all counties is readily available. This approach does not correct for a possible correlation between the first-order and second-order spatial neighbors (both of which will be predictors), but two independent variables being correlated is generally easier to work with than having a large number of correlated predictors.

[2] While we include second-order spatial neighbors in some steps of our analyses, we ultimately find that their inclusion gives no benefit and thus they are consistently omitted.

As noted above, using aggregate unemployment is equivalent to weighting the neighboring counties based on labor force. While there is no weighting matrix explicitly specified, weights are implicit in the setup of the model. Clearly those counties which are more heavily populated will have a greater impact on a particular region's unemployment rate.

Giacomini and Granger (2004) present a theoretical model showing that when spatial correlation is present, building a forecasting model that accounts for this spatial structure leads to more efficient forecasts. Bronars and Jansen do similar work but apply it to empirical data. They take county level data, but alter the arrangement of the counties. Genton and de Luna (2004) use the nine census regions of the United States to show how effectively a model built for spatial correlation can capture such dependence. What I do is based on these three previous studies with a slight addition. Instead of just identifying and/or capturing the spatial correlation using the data, I go one step further and construct forecasts to show predictive power. In addition, by combining first-order and second-order spatial neighbors into a single unit for each, I reduce the dimensionality thereby simplifying the model construction.

The underlying theme of both Bronars and Jansen (1987) and Genton and de Luna (2004) is that spatial dependence is present in unemployment, and ignoring it can be costly. Giacomini and Granger (2004) show theoretically that not accounting for this dependence when forecasting is a mistake.

Parameter estimates on the predictors indicate the effect of an incremental change in unemployment in one of the neighboring regions on the county in question. Since the predictors will be temporally lagged, I have estimates of what a change in unemployment

one month in neighboring counties has on unemployment in the next month (assuming a one-period lag).

I also check to see if there is any pattern in the forecasting accuracy with respect to county characteristics. It is plausible that the inclusion of an agricultural county's unemployment rate regardless of geographical proximity will improve forecasts for another agricultural county's unemployment rate. This property is checked for agricultural concentration, manufacturing concentration, educational concentration, and metropolitan status by including similar counties as predictors.

It would also be interesting to see if counties that are located in the interior of the state have a stronger spatial component than those on the border. Perhaps there is more spatial interaction occurring where the exchange of labor allows the economic actors to remain in the state rather than crossing the border to another state. People who work in one state and live in another sometimes face a more complicated tax situation. This could make such people less likely to search across state lines. The hesitation to cross state lines could also be for cultural reasons. Before and during the Civil War, Missouri was a state that allowed slavery while Kansas was a "free state." This may have caused such differences between the characteristics of the people who lived in these states to prevent them from crossing state lines. This is tested by separating the non-Missouri counties as separate predictors. Specifically, in the case of a first-order spatial lag, the variable $Y_{j(i),t}$ includes any adjacent counties outside Missouri. By pulling out the non-Missouri counties from this variable and specifying two first-order spatial lags ($Y_{MOj(i),t}$ and $Y_{Nonj(i),t}$), I am able to compare the effects of those counties located outside of Missouri to those in Missouri based on the coefficients on these variables.

While having these parameter estimates can be useful, they are not the ultimate goal of my research. In chapter 4, I use these results to build forecasting models for each county in Missouri. To gauge how effective the forecasts are in predicting future unemployment, I use a subset of my data to build the models so that I have actual unemployment figures with which to test the accuracy of the out-of-sample forecasts. I estimate my model based on a subset of the months and use those results to predict unemployment in the subsequent month.

I also build non-spatial models for each county for comparison purposes and conduct the same procedure with these non-spatial models as I did for the spatial models, obtaining the necessary parameter estimates to construct forecasts. The non-spatial models provide a benchmark to which I can compare the accuracy of the spatial models. This allows me to see under which circumstances including a spatial lag as a predictor helps predictive accuracy. While the statistical significance of predictive accuracy is not formally tested, the patterns observed give us the informal ability to infer that the difference is very likely to be significant.

Another consideration is the number of prior periods to be used to estimate the model. If there are structural breaks in the processes underlying the model, it may be more appropriate to use only a recent sample to build the forecasting model. Instead of using the full dataset, I use data for 60 months to estimate parameters and use those estimates to forecast the 61st month's unemployment. If there are changes in the structure of unemployment that are shorter (for example) than this time period of 60 months, then a 60-month model will not accurately capture it. It is also possible that using less than the

full sample affects the number of predictors that contribute to prediction. For these reasons, we also consider forecasts made from models of both 40 and 100 months.

The three seminal papers I use as the basis for my research primarily show either theoretically or empirically that spatial models are more appropriate than non-spatial models when there is a spatial process present in the data. Bronars and Jansen (1987) and Giacomini and Granger (2004) build a theoretical model and test it with a data, while Genton and de Luna (2004) use national data to illustrate the presence of spatial correlation. Only Giacomini and Granger (2004) address the possible gains of using a spatial model in forecasting. I explore the nature of spatial correlation in Missouri at the county level using actual data both contemporaneously and over time. I also measure the effects that last month's unemployment has on this month's unemployment and what factors affect this relationship. I also address the gains to forecasting accuracy. In the process, I learn properties of county-level data as it applies to the state of Missouri that can be used for predicting other variables related to unemployment or future planning by government.

2. STRUCTURE OF MISSOURI UNEMPLOYMENT

Before we start building spatial forecasting models for county-level unemployment, both county-level unemployment and the geographic structure of unemployment needs to be examined. It is expected that counties that are more proximate to each other will have more closely correlated unemployment rates. Exactly how the distances should be coded is an initial question. Some spatial measures are based on contiguity, positing that counties sharing a border will display similarities in pattern. Alternatively, distance may play a role among counties that are not contiguous. With that in mind, both distance and contiguity are examined.

In the initial estimates, two different measures are calculated to capture the spatial relationship in unemployment. The first measure captures covariance between unemployment rates based on the spatial relationship between a pair of counties at a given point in time. The second measure examines how the unemployment rate varies over time, considering which unemployment rates tend to move together. These capture conceptually separate patterns, although in practice they generally correspond. The two measures produce similar results in our case.

While it is expected that the results will show that distance matters in understanding patterns of unemployment, it is possible that other factors unrelated to distance may be of equal or greater importance. For example, if two counties both have a large manufacturing base, then any shock that hits the manufacturing sector would affect the employment in both counties regardless of their geographic proximity. Because of this, it is important to check for unemployment patterns for counties that have similar

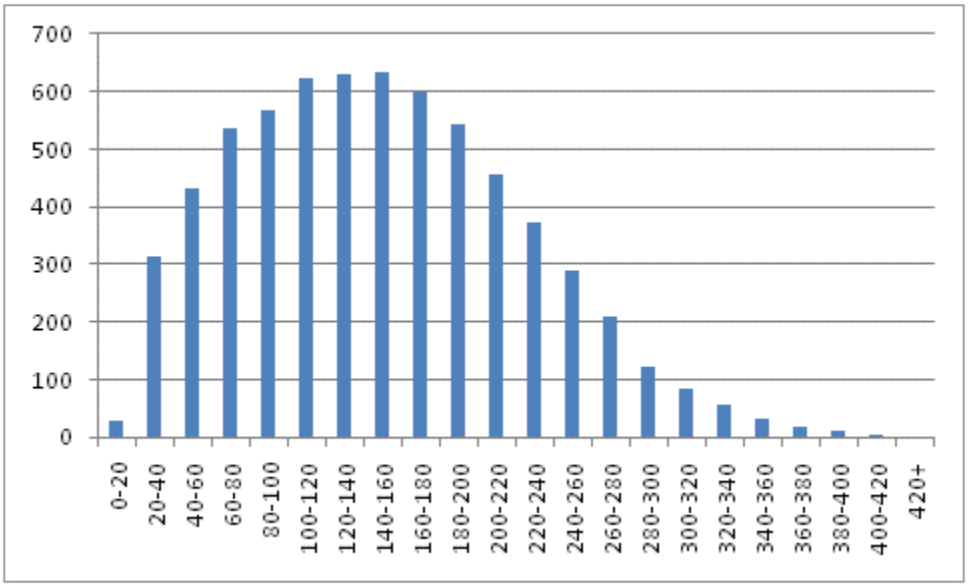
industrial profiles. Similarities in patterns of unemployment across counties based on manufacturing concentration, agricultural concentration, and educational concentration are examined. Similarly, metropolitan status may be an important factor for patterns in county-level unemployment as some shocks will affect metro or rural areas differently. The results show that while these factors identify counties that share some similarities in unemployment patterns, geography is of much greater importance.

After this preliminary analysis is complete, the most widely used measure of spatial correlation, Moran's I, will be used to study alternative measures of geographic proximity. In the calculation of Moran's I, weights specify the importance that one county's unemployment rate has for other counties. In general, weights are greater for pairs that are more proximate. It is common to incorporate contiguity in the weights by assigning a common weight for pairs of contiguous counties and a zero weight for counties that do not share a border. Various weighting structures are tested based on geographic distance as well as contiguity. The focus will be finding the best fitting weighting structure for county-level unemployment in Missouri.

2.1 Counties in Missouri

The first step in the process will be to examine the general spatial structure of county unemployment in Missouri. To lead off the analysis, I have produced a distribution of mileages between geographic county centroids measured in straight-line distance. Most counties in Missouri are roughly similar in size when measured in square miles. Figure 2.1 represents this distribution of distances between all pairs of 115 counties (6,555 distinct pairs).

Figure 2.1: Distribution of mileages between all county centers in Missouri

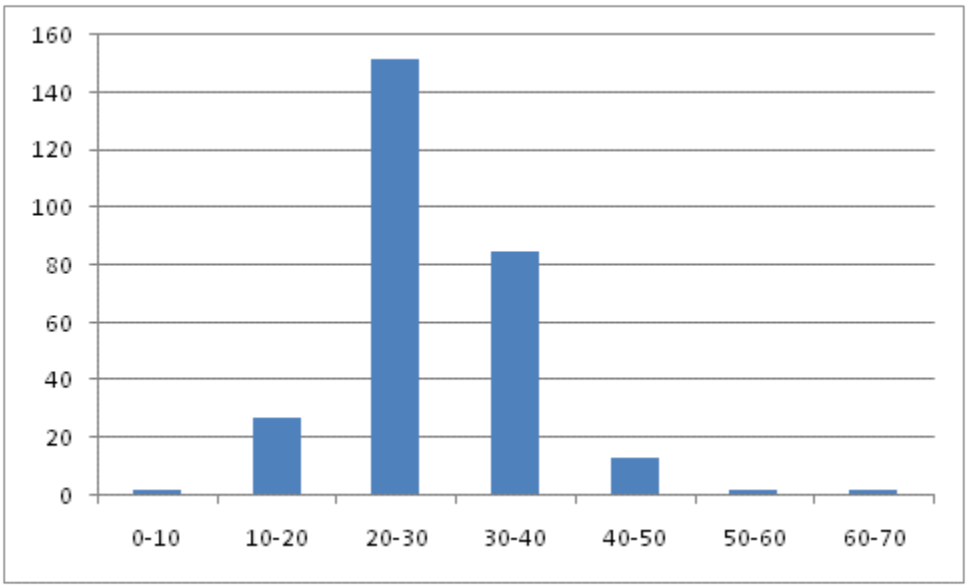


As the graph shows, the most common distance range between counties is 160-180 miles, and no county in the state is farther than 440 miles from another county. If we look at the distance between those pairs of counties that are contiguous, we can see that the average distance between adjacent counties is 38.29 miles. Figure 2.2 presents a distribution of the mileage between counties that are adjacent to each other.

Table 2.1: Descriptive Statistics of mileage between county centers for contiguous counties in Missouri

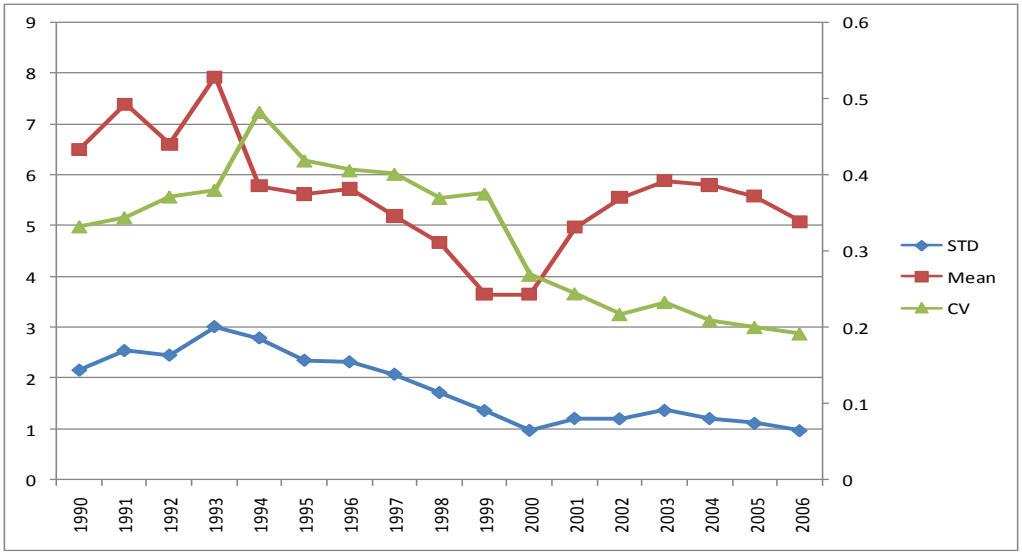
Measure	In Miles
Avg	38.29
Max	68.94
Min	7.19
Stdev	7.37
Pairs	280

Figure 2.2: Distribution of mileages between county centers for contiguous counties in Missouri



It is also helpful to get a sense of the overall picture of county unemployment in Missouri. Figure 2.3 below depicts some descriptive statistics for county unemployment (n=115).

Figure 2.3: Mean, standard deviation, and coefficient of variation of county-level annual average unemployment rates, 1990-2006



As seen on the graph above, the mean and standard deviation are both declining initially. There is a general downward trend during the 1990s with the exception of the spike in 1993. However, the upward shift in the mean since 2000 is not accompanied by a similar shift in the standard deviation. The coefficient of variation takes a dip indicating that the spread of unemployment relative to the mean at the county level tightened at the turn of the century.

2.2 Spatial Correlation of Unemployment at the County Level Based on Distance Measures

The next focus is on the degree of spatial correlation between county-level unemployment rates based on alternative distance measures. The first measure considered will be based on contiguity.

Before any results are presented, it is important to define the way that these correlations were measured. They were calculated using Equation 2.1 which is an adaptation of a method used by Conley and Topa (2002). While this equation is very similar to Equation 1.10, I use a weighting structure based on categories of distance.

Equation 2.1 $f\left(C_k(i,j)\right)=\sum_{i=1}^{N}\sum_{j=1}^{N} W_N[C_k(i,j)]\left(X_i-\bar{X}\right)(X_j-\bar{X})$

$W_N[C_k(i,j)]$ is a weighting function that is nonzero where the relationship between counties i and j is characterized by category k (i.e., they are "matched"), i.e.,

Equation 2.2 $W_N[C_k(i,j)]=0$ if i and j are not matched and $W_N[C_k(i,j)]=\frac{1}{N(C_k)}$ otherwise

where $N(C_k)$ is the number of counties in class C_k

In Equation 2.2, C_k is a function which identifies whether the relationship between counties i and j corresponds with category k. In our first analysis, k=1 will identify counties that are contiguous, whereas k=2 will correspond with counties that are

not contiguous. Specifically, for Category 1 (C_1) which is the contiguous case, two counties (i and j) are matched if they share any length of border with one another. There are 280 pairs of counties that share a border out of the 6,555 possible distinct pairs of counties, so $N(C_1) = 280$.

Assuming two counties are matched together (and thus receive the non-zero weight), their respective unemployment values (X_i and X_j) will be subtracted from the mean of all counties' unemployment ($\overline{X}$) and those two differences will be multiplied together. Once this is done for all non-zero weighted counties, the products are summed to arrive at a measure of the covariance between all pairs of counties that are within a particular category. This is done for a single point in time. This covariance measure is calculated for each year: the annual average of monthly unemployment in each county in Missouri has been used to apply Equations 2.1 and 2.2.

2.3 Contiguous Counties

The first two categories identify whether a pair of counties are contiguous. Figure 2.4 graphs these measures for the data for each year in the sample (1990-2006).

Figure 2.4: Spatial covariance of annual average of monthly unemployment between contiguous and non-contiguous counties from 1990-2006

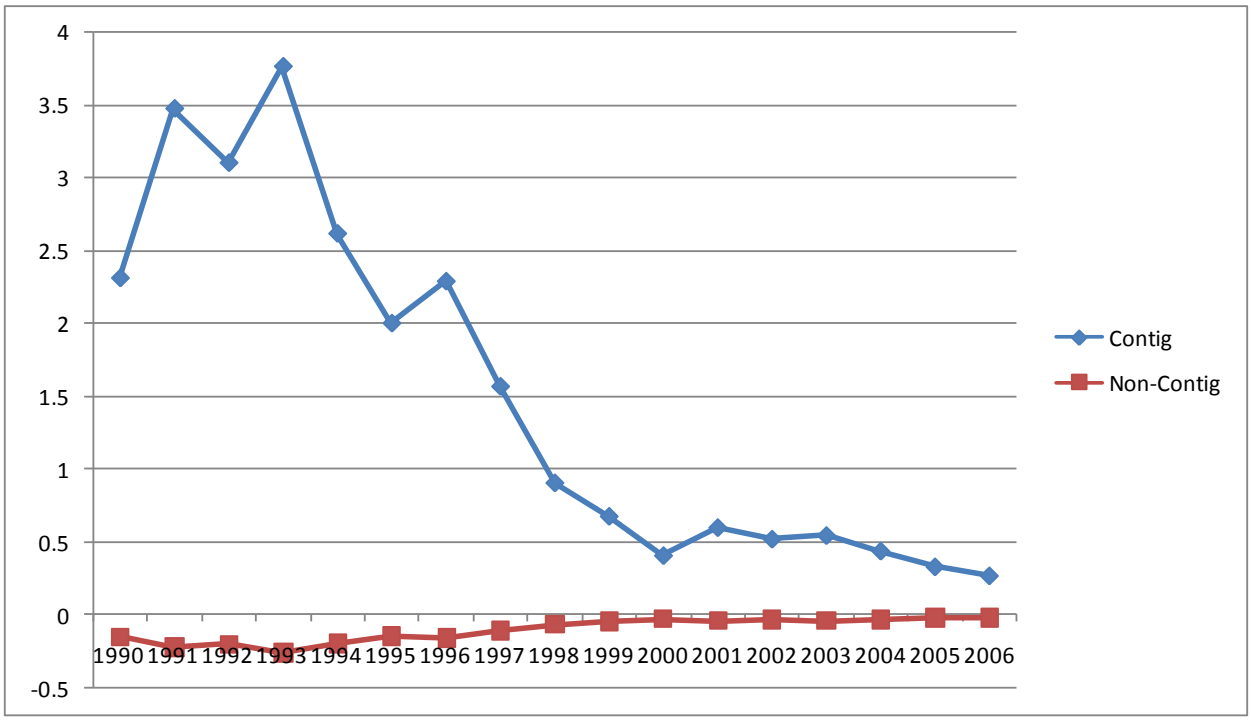


Figure 2.4 presents results based on Equations 2.1 and 2.2.

As the graph shows, the contiguous counties exhibit a high degree of spatial covariance throughout the sample while the non-contiguous counties exhibit a very weak and negative relationship. This result is to be expected; counties that are closer to one another are likely to have similar unemployment rates. One point worth noting is how both measures appear to be converging towards zero at later points in the sample. This is explained by the trend depicted in Figure 3 where the variance and standard deviation of annual county unemployment are also decreasing. Both have a similar path over time to the two lines shown above.

2.4 Mileage Bands

The next category measures are based on mileage between counties. For the purposes of this analysis, I have selected mileage bands of 50 miles. Figure 2.5 shows the categories based on mileage bands.

Figure 2.5: Spatial covariance of annual average of monthly unemployment between counties based on mileage between centers from 1990-2006

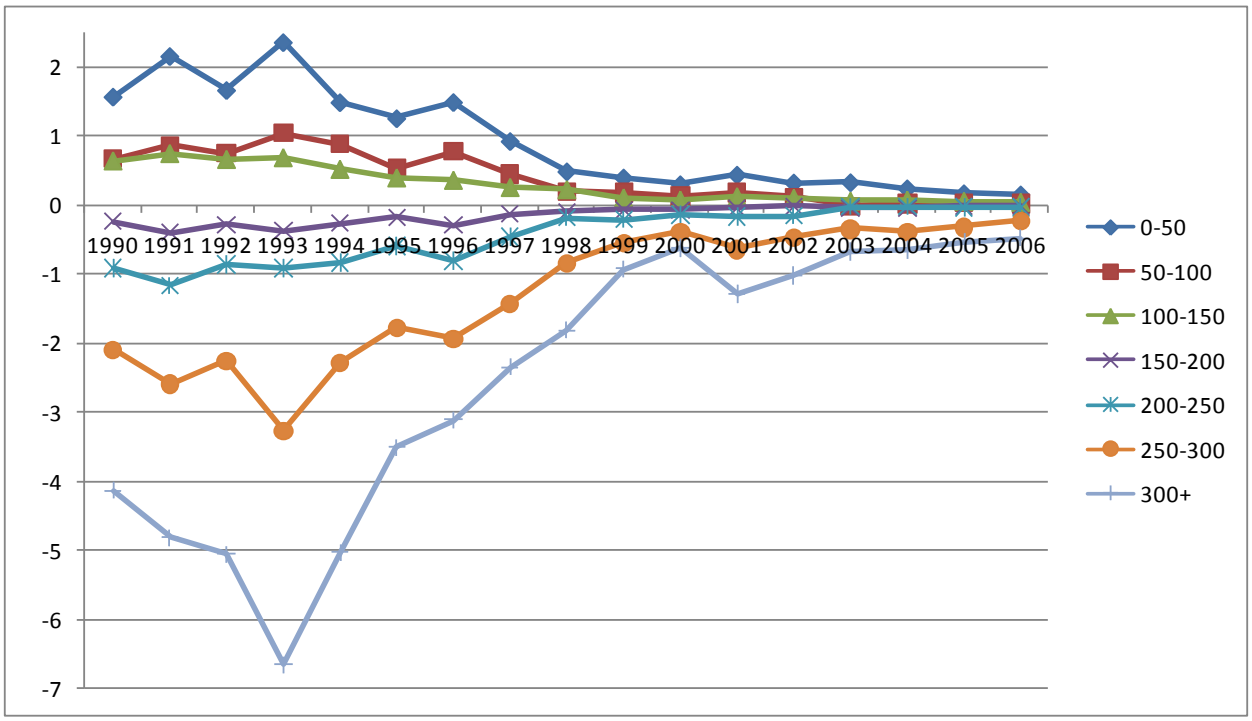


Figure 2.5 presents results based on Equations 2.1 and 2.2.

Consistent with the results based on the contiguity measures, those counties that are geographically proximate exhibit a positive relationship between their unemployment rates; this implies that if one county's unemployment is higher than the average, so is the unemployment rate of its neighbors. As the distance between counties increases, the degree of the spatial relationship tends to soften. These covariance figures also tend to converge towards zero in the later periods, as in the previous graphs. This decline in

covariance is larger than the decline in the variance of unemployment rates by county over the same period.

One interesting result shown here is that as we increase the distance beyond 250 miles, the relationship turns strongly negative. This relationship is expected due to the equation used to construct these measures. For each time period in the sample, the weighted sum of all covariances across all distance categories must equal zero. Due to the fact that the closer counties are showing a positive relationship we expect to see a negative relationship present in the more distant counties.

2.5 Correlations Over Time Based on Contiguity and Mileage

All of the previous measures in this section considered spatial dependence at a single point in time (year). As this research is concerned with prediction, another factor to consider is the correlation between county unemployment rates over time. In the following analyses, we calculate the standard correlation coefficient for a particular pair of counties, where each month in the period of our study is a unit in the analysis.

The previous measure captures the unemployment rates for two counties at one single point in time. The correlation measure that is presented below looks at the entire time series of data and measures how closely the unemployment rate of the pair of counties move through time. Consider a county that is always 1% above the mean and another that is always 1% below. This pair of counties would exhibit a negative covariance at all points in time, based on the measure for a single point in time reported above. In contrast, the correlation of the unemployment rates for this same pair of counties over time would be positive, as they would be moving together in similar magnitudes in the same direction through the course of time. The first measure looks at

deviation from the mean in any single time period, whereas the second measures how closely the two counties' rates move together.

Table 2.2: Descriptive statistics of correlations of unemployment rates over time for contiguous and non-contiguous counties

CORRELATIONS	N of Pairs	Mean	Max	Min	Stdev
Contiguous	280	0.67415	0.98050	-0.06354	0.17570
Non-Contiguous	6275	0.55787	0.96123	-0.24465	0.20260

Figure 2.6: Distribution of correlation coefficients of unemployment rates over time for contiguous counties

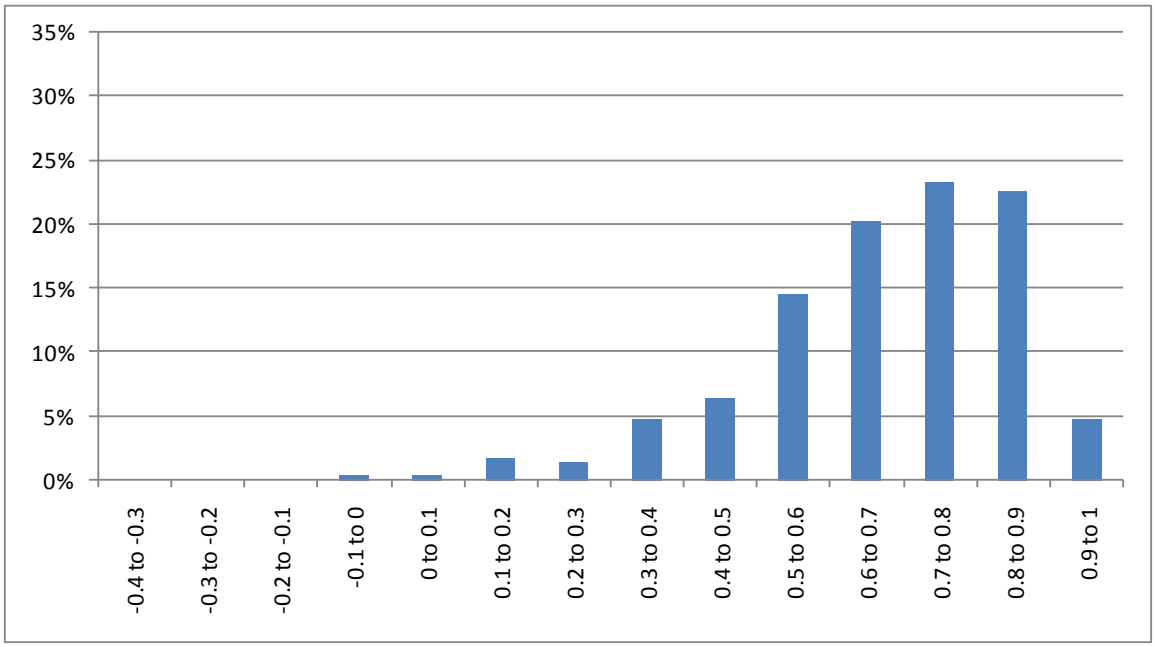


Figure 2.7: Distribution of correlation coefficients of unemployment rates over time for non-contiguous counties

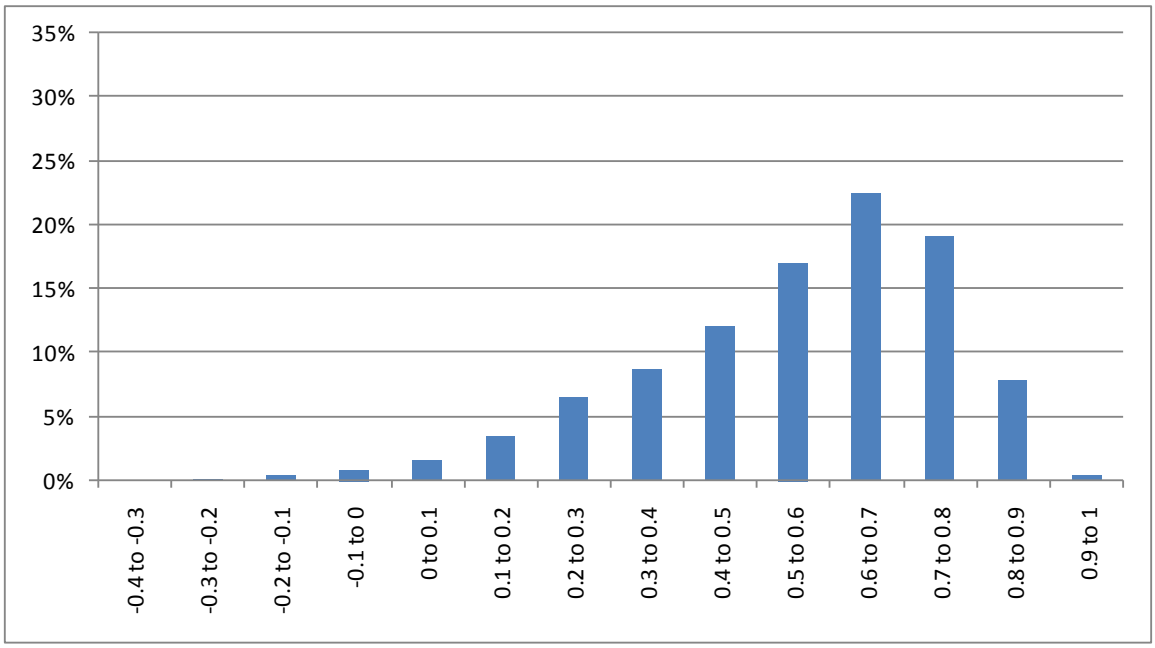


Table 2.2 shows descriptive statistics of the correlation coefficients calculated between contiguous and non-contiguous counties. The subsequent graphs show the distributions of these correlation coefficients. While both categories of counties exhibit a high degree of correlation, the contiguous counties have a higher mean than the non-contiguous counties. We can also see that the non-contiguous counties have a substantially smaller percentage of the higher correlations than those of the contiguous counties. While there is no temporal lag present in the correlation calculations, it does appear from these figures that one contiguous county would be more effective at predicting an adjacent county than it would be in predicting the unemployment in a non-adjacent one.

When mileages between counties are considered instead of simply contiguity, we can see a similar result in that as the distances increase the correlations decrease, which is shown by the results in Table 2.3. The following figures illustrate this effect.

Table 2.3: Descriptive statistics of correlations for unemployment rates over time for counties that are separated by various mileage bands

CORRELATIONS	N of Pairs	Mean	Max	Min	Stdev
0-50 Miles	544	0.64347	0.98050	-0.14915	0.18413
50-100 Miles	1334	0.58952	0.93063	-0.24465	0.19239
100-150 Miles	1591	0.57582	0.92320	-0.17082	0.19054
150-200 Miles	1435	0.55147	0.95031	-0.13176	0.19755
200-250 Miles	990	0.52802	0.96123	-0.11693	0.21629
250-300 Miles	458	0.48170	0.87981	-0.16727	0.22410
300+ Miles	203	0.50121	0.87257	-0.14482	0.21391

Figure 2.8: Distribution of correlation coefficients of unemployment rates over time for counties 0-50 miles apart

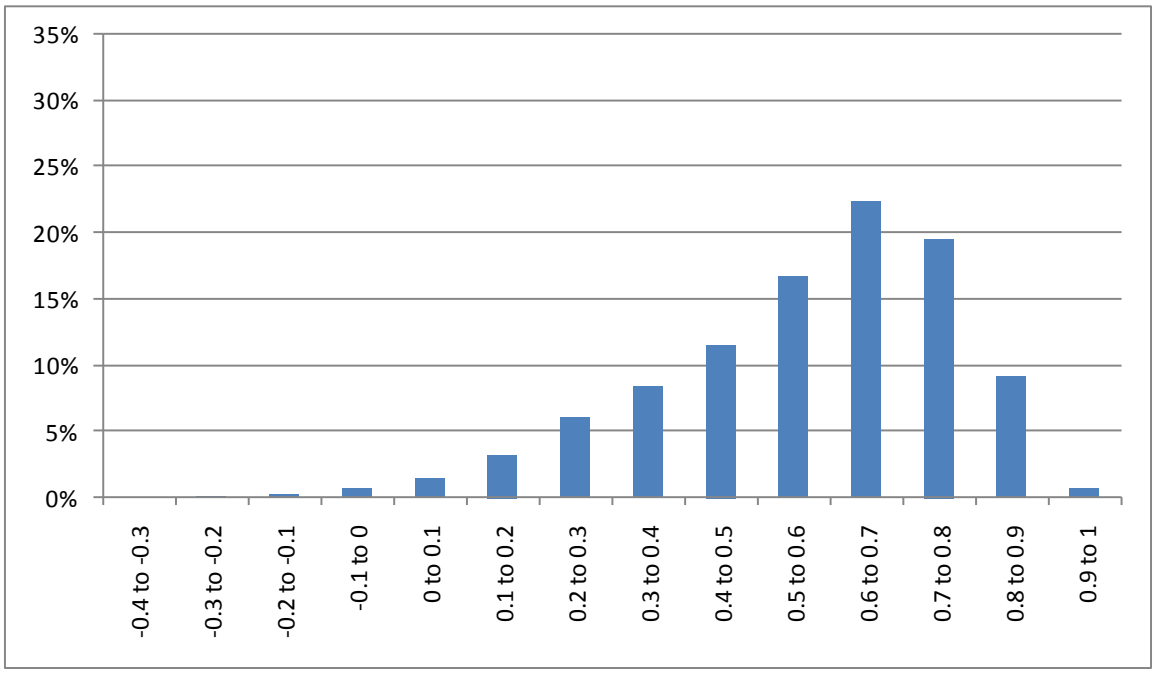


Figure 2.9: Distribution of correlation coefficients of unemployment rates over time for counties 50-100 miles apart

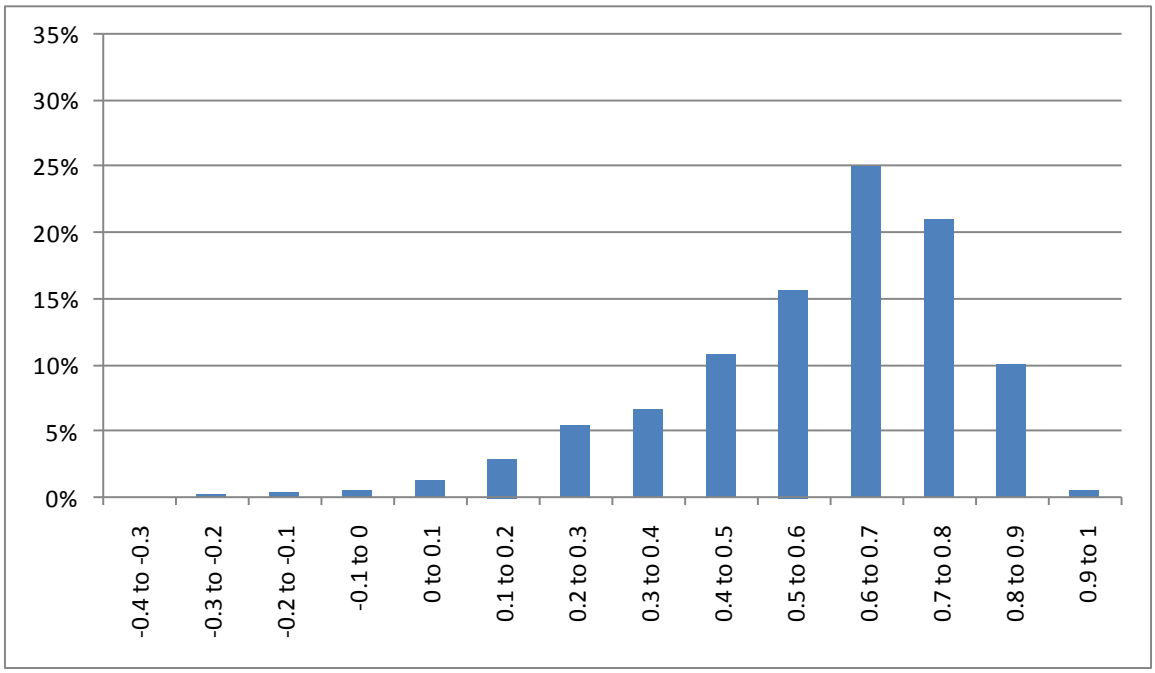


Figure 2.10: Distribution of correlation coefficients of unemployment rates over time for counties 100-150 miles apart

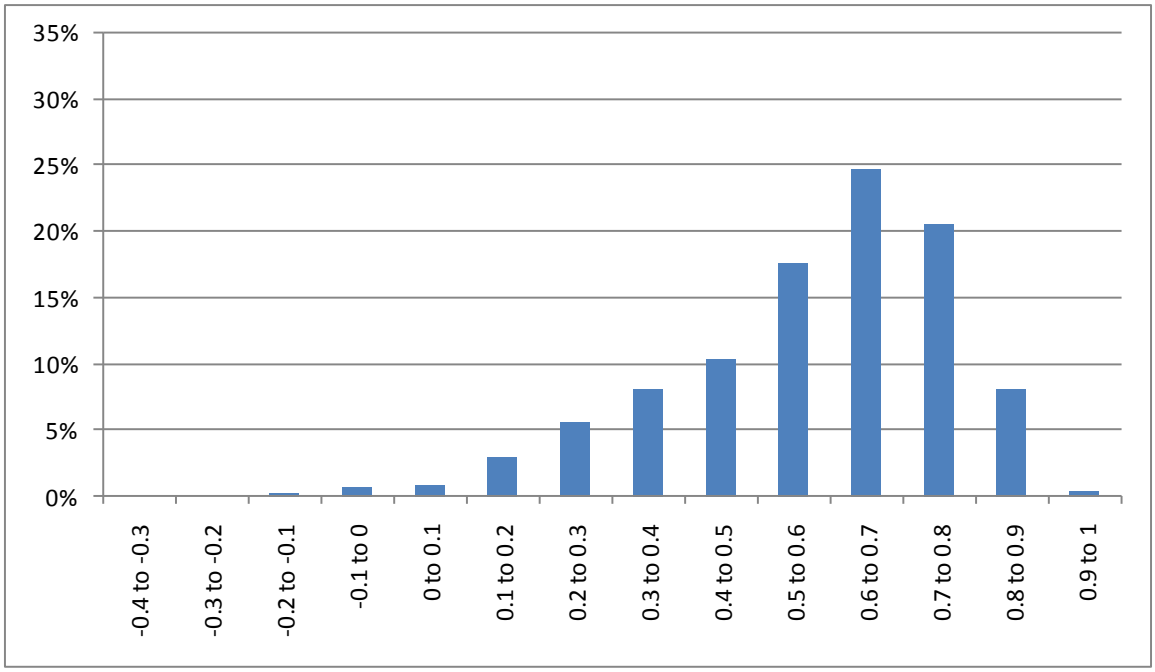


Figure 2.11: Distribution of correlation coefficients of unemployment rates over time for counties 150-200 miles apart

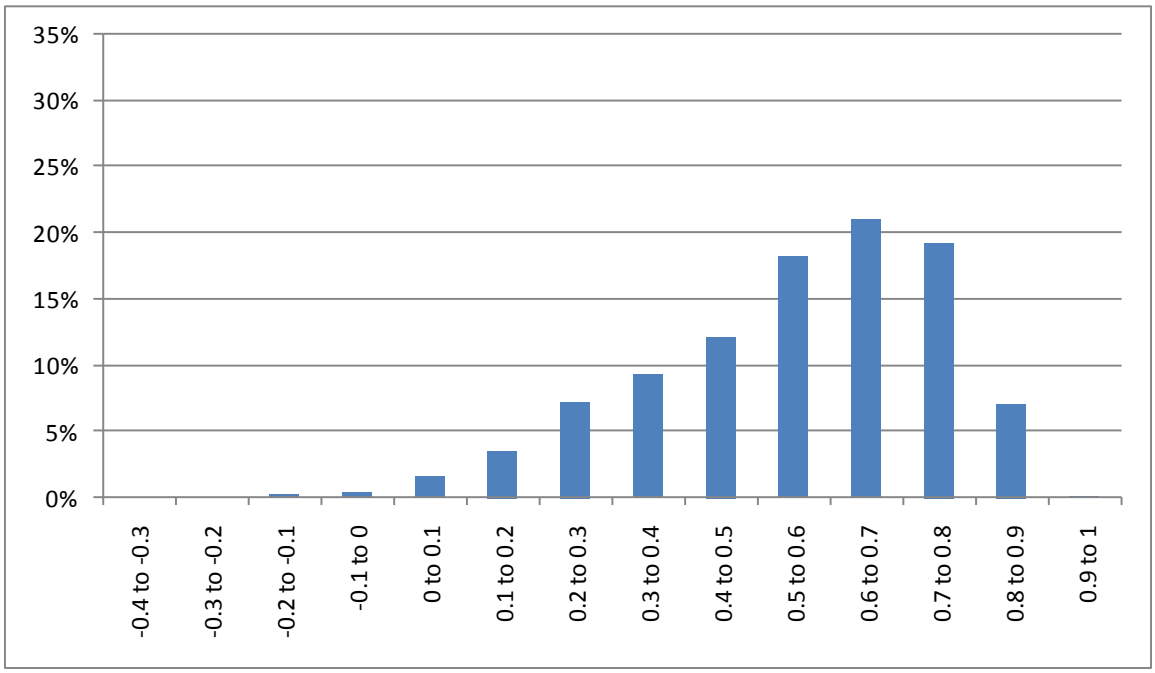


Figure 2.12: Distribution of correlation coefficients of unemployment rates over time for counties 200-250 miles apart

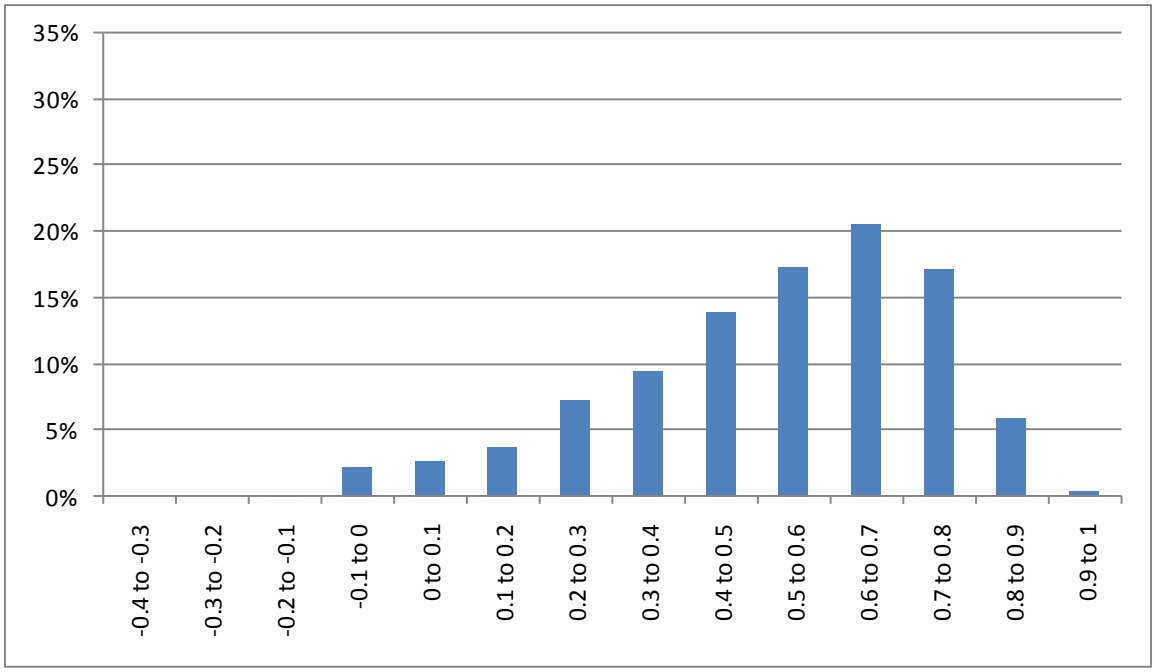


Figure 2.13: Distribution of correlation coefficients of unemployment rates over time for counties 250-300 miles apart

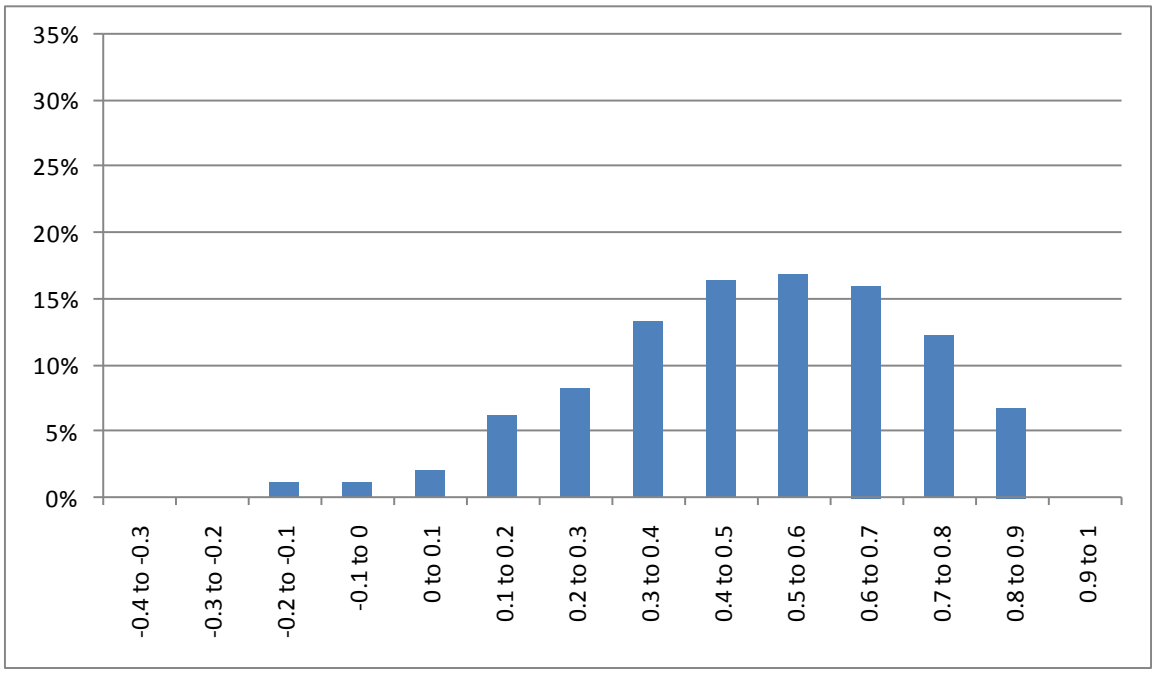
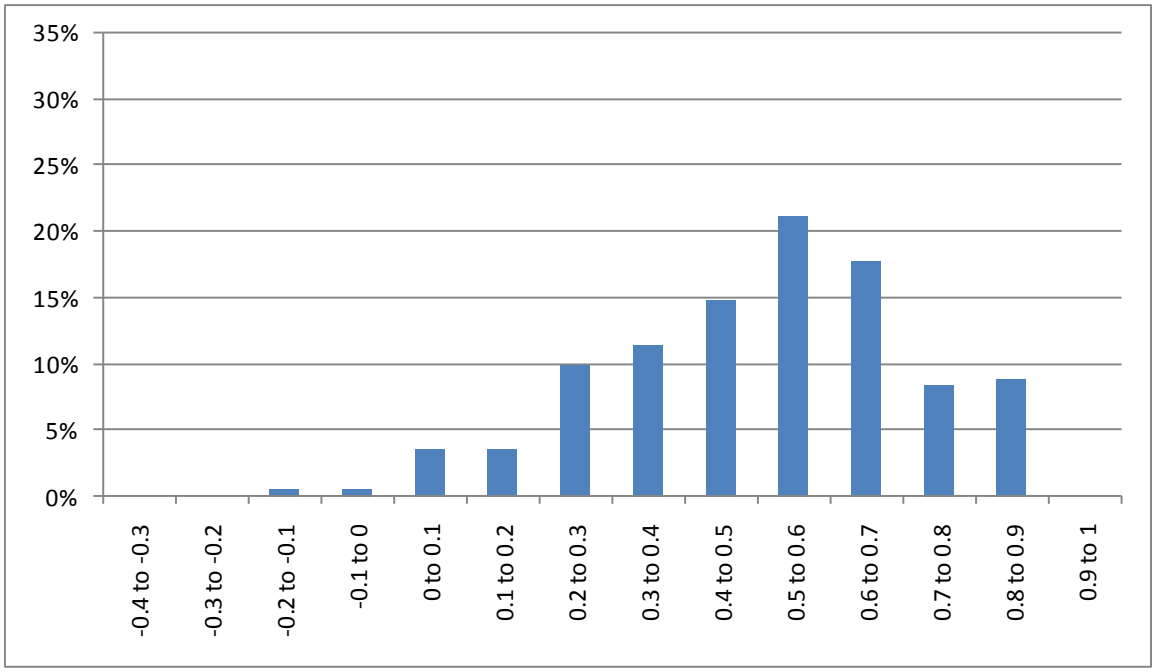


Figure 2.14: Distribution of correlation coefficients of unemployment rates over time for counties 300 or more miles apart



As the distance between counties increases, there are several things to notice. First of all, we can see that the mean correlation generally decreases for greater mileages. This is in line with previous calculations based on contiguity. This also suggests that the more proximate two counties are, the more effective predictor one unemployment rate will be for the other. We also see that while the mean does decrease with increasing distance, it does not fall below 0.5 (with the exception of the 250-300 mileage range), indicating fairly strong positive correlations. This is an indication that changes in county level unemployment in Missouri are very strongly tied to one another, meaning that all counties are following this larger region's rate. It would appear from this that accounting for a more global measure of unemployment that is inclusive of all counties will be appropriate.

In conclusion, it appears that spatial correlation is present at the county level in the state of Missouri. In the following section, I will explore some of the possible alternative factors that might also be relevant, including the metropolitan status of counties, the agricultural concentration of counties, the manufacturing concentration of counties, and the educational concentration of counties. For the next discussion, I group counties of similar makeup together based on these categories and check to see if there is any pattern present with counties that are similar.

2.6 Metropolitan Status of Counties

I will begin by reporting the results of applying Equations 2.1 and 2.2 where the categories are determined by whether or not a county is considered a metropolitan county. If a county meets the Census Bureau's definition of a metropolitan county, then it is considered "Metro." Otherwise it is classified as "NonMetro." In this section, three

categories will be considered for matching purposes. One will consider only pairs that are both Metro counties, one will consider two counties a matching pair both are NonMetro counties, and a third will assign two counties the designation of matching pair only if one is a Metro and the other is not.

Figure 2.15: Spatial covariance of unemployment between counties based on metropolitan status annual from 1990-2006

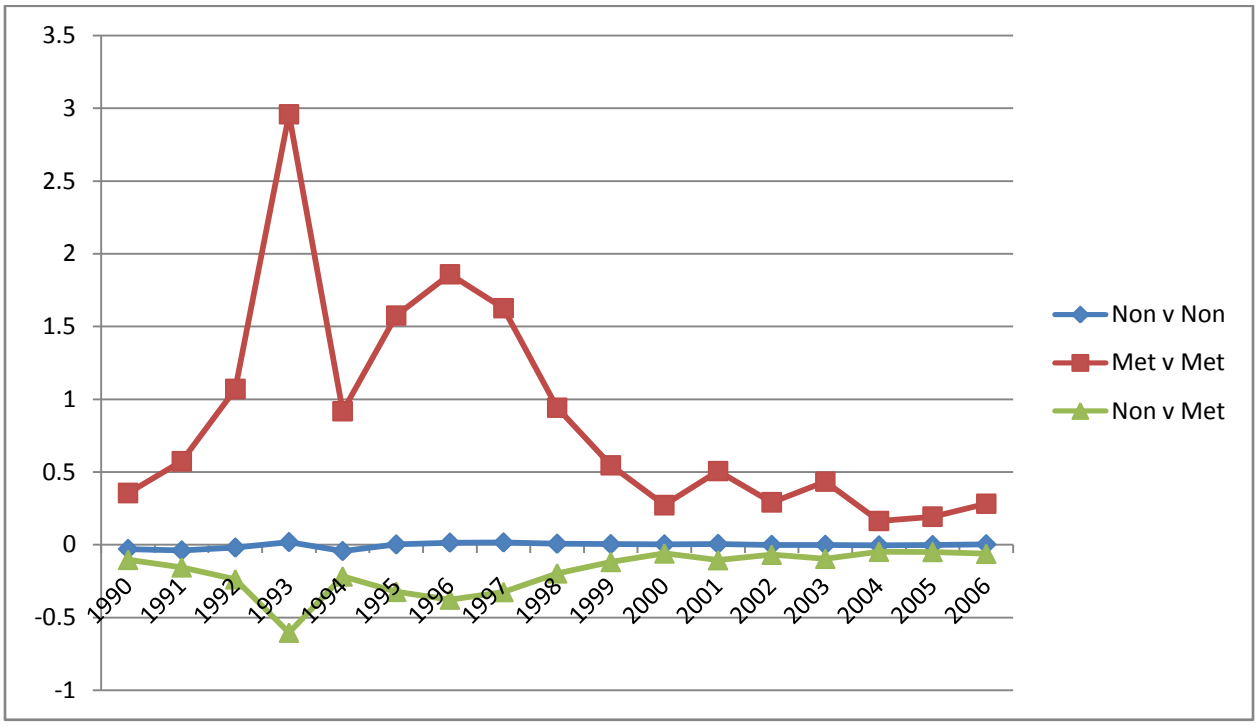


Figure 2.15 presents results based on Equations 2.1 and 2.2.

As we can see from the graph above, the Metro pairs exhibit a high degree of correlation, particularly in the 1990s. The Non-Metro pairs by contrast do not stray very far from zero. The Non-Metro and Metro pairs do deviate a little further from zero and are negative (implying metro and non-metro counties tend to have discrepant unemployment rates relative to the mean); however the deviation from zero is still slight.

Also, the correlation coefficients for unemployment variation over time between counties (as done previously based on mileage and contiguity) is presented for these categories.

Table 2.4: Descriptive statistics of correlations for county unemployment rates over time based on metropolitan status

CORRELATIONS	N of Pairs	Mean	Max	Min	Stdev
NonMet/NonMet	3240	0.54360	0.98050	-0.16727	0.20951
Met/Met	561	0.66276	0.97705	0.13309	0.14893
NonMet/Met	2754	0.56499	0.95031	-0.24465	0.19844

Figure 2.16: Distribution of correlation coefficients of unemployment rates over time for pairs of non-metropolitan counties

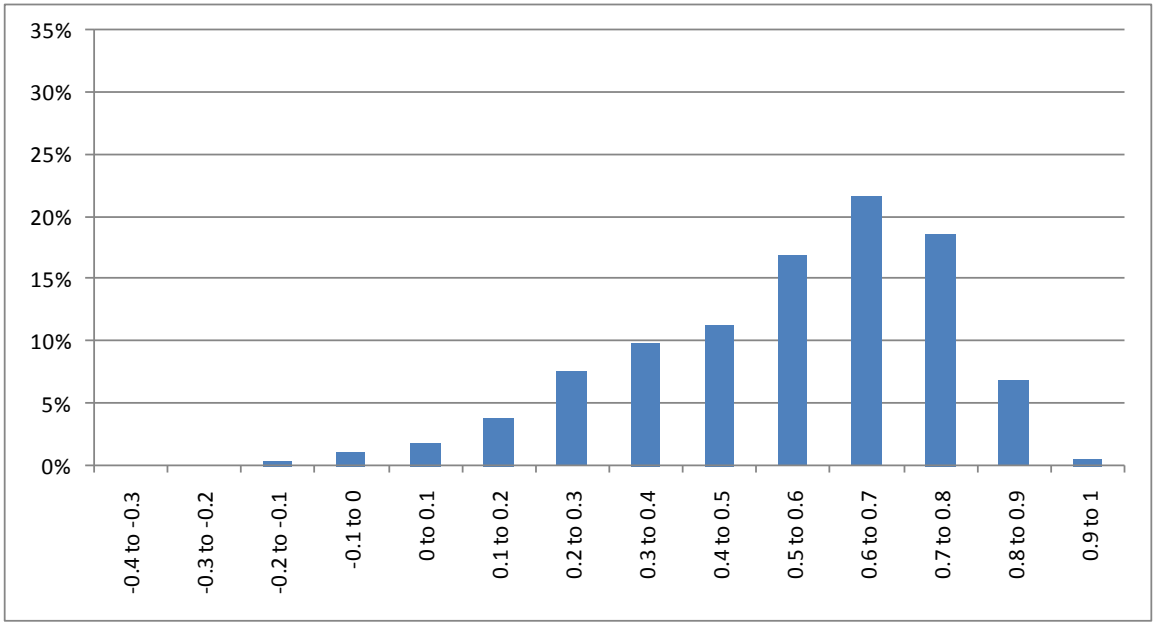


Figure 2.17: Distribution of correlation coefficients of unemployment rates over time for pairs of metropolitan counties

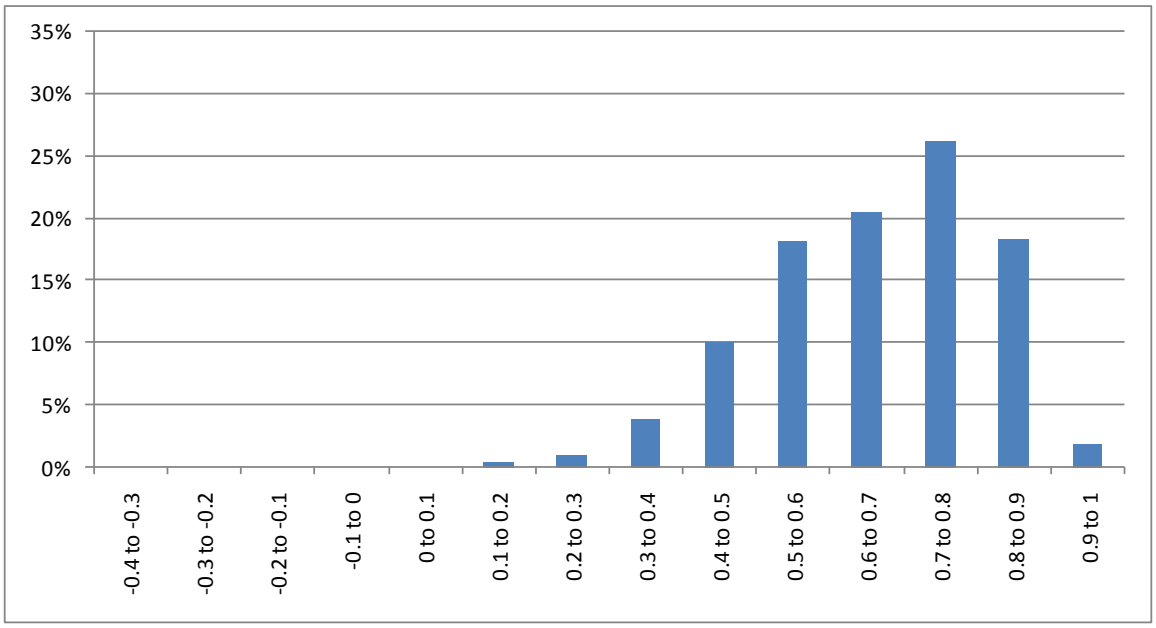
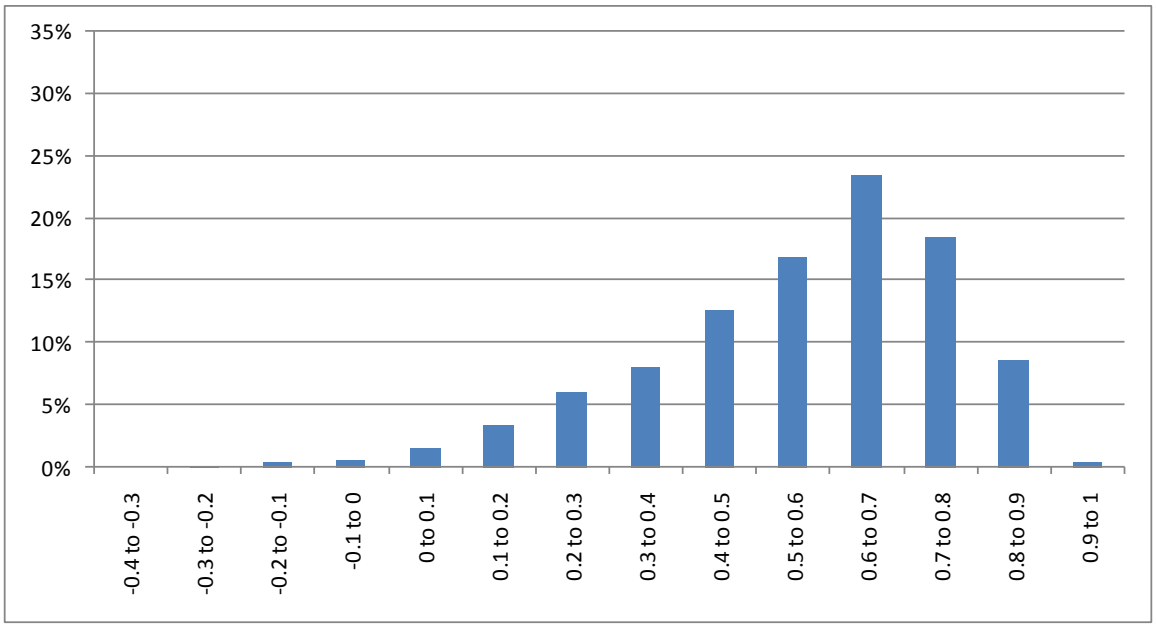


Figure 2.18: Distribution of correlation coefficients of unemployment rates over time for metropolitan/non-metropolitan county pairs



Once again the categories display means that are above 0.6 (which would further suggest that these may reflect larger region unemployment). The Metro and Metro pairs however

show a mean of 0.8 and nothing below 0.2. This would be in line with our covariance measures for a single point in time displayed in Figure 2.15.

2.7 Agricultural Concentration

Another possible factor that may be relevant in explaining unemployment patterns across counties is the industrial profile of the county. To determine whether or not a county is considered an agricultural county, I use the percentage of the labor force that is employed in the agricultural sector in the county's overall labor force (Figure 2.19). There is no natural break in the distribution, no clear point that easily distinguishes the agricultural counties from the non-agricultural counties. Two thresholds were examined for distinguishing between agricultural and non-agricultural counties. I considered both 1% and 3% of the labor force employed in the agricultural sector as the dividing line between an agricultural and non-agricultural county. The 1% threshold is the one that is used for the analysis, and such counties are identified as "Ag." The reason that 1% is chosen over the 3% threshold is because using the 3% threshold would have resulted in a low number of counties (eight) designated as agricultural. We also found that the two thresholds produced very similar results and thus the choice makes little difference in the final conclusions.

The categories for matching pairs are the same as those for the metropolitan counties except that there are three categories: one that matches only pairs of Non-Ag counties, one that matches pairs of Ag counties, and a third that matches Ag and Non-Ag counties.

Figure 2.19: Distribution of agricultural concentration by county

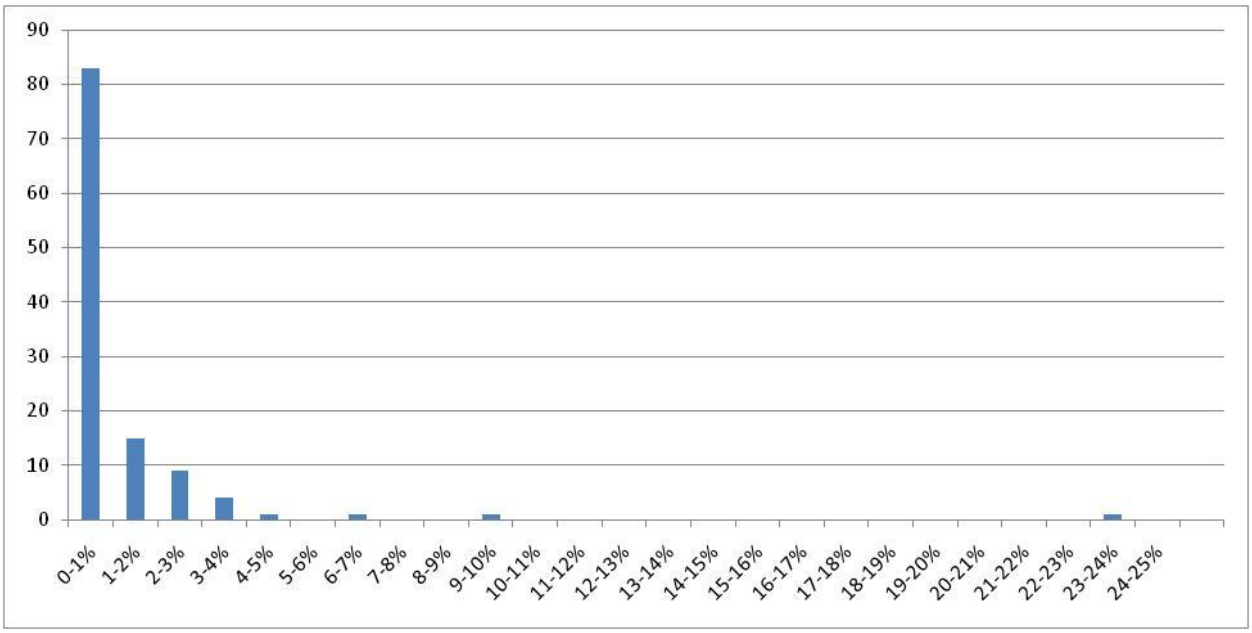


Figure 2.20: Spatial covariance of unemployment between counties based on agricultural concentration annual from 1990-2006

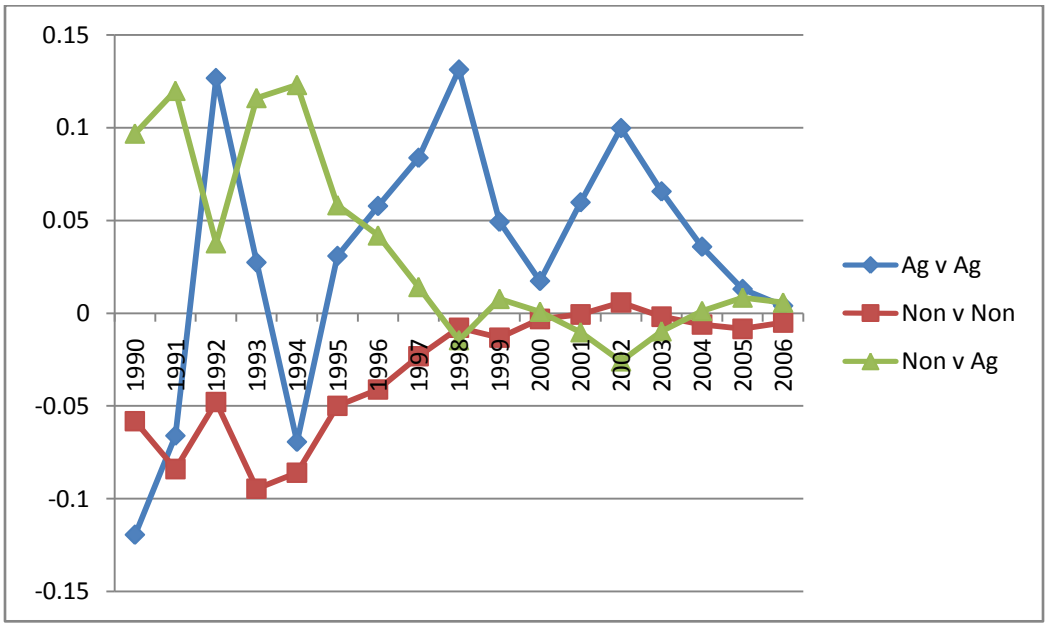


Figure 2.20 presents results based on Equations 2.1 and 2.2.

What we initially see from this category classification is that the over time correlations for the Ag/Ag pairs seem to vary, but the variation in absolute terms is small. The NonAg/NonAg pairs as well as the NonAg/Ag pairs all remain negative (save NonAg/NonAg in 2002) though neither deviate a great deal from zero (absolute value almost always less than 0.1) More information is added to this part of the story by the correlations of the unemployment rates over time. Table 2.5 shows that the average amount of correlation between counties is similar in all three groups listed, which are based on their relative levels of agricultural concentration.

Table 2.5: Descriptive statistics of correlations of unemployment over time for counties based on agricultural concentration

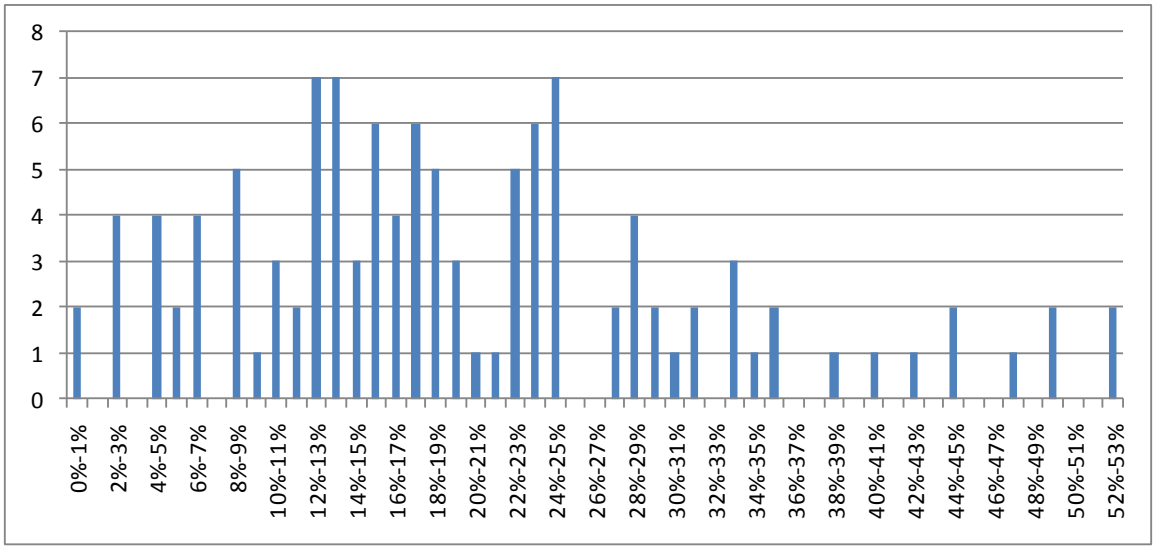
CORRELATIONS	N of Pairs	Mean	Max	Min	Stdev
Ag/Ag	496	0.57207	0.87980	-0.13176	0.18909
NonAg/NonAg	3403	0.56247	0.98049	-0.16661	0.20664
NonAg/Ag	2656	0.56146	0.94654	-0.24465	0.20183

Looking at the distribution of correlations (not presented) confirms that correlations for these three categories are very similar. It does not appear that the presence (or lack of) agriculture identifies counties whose unemployment rates are similar or move together. When using the less restrictive 3% designation as opposed to the 1% threshold, the difference appears to be that the 3% threshold produces somewhat larger proportion of very low correlations for Ag/Ag pairs. Otherwise the distributions are very similar, so the effects on the means are negligible.

2.8 Manufacturing

The next category considered is whether or not a county is a manufacturing county. Again I have used the percentage of the labor force in a county that is employed in the manufacturing sector as the guide for whether or not a county is classified as manufacturing. While the distribution of agricultural employment across counties was smooth, the manufacturing distribution was much more erratic.

Figure 2.21: Distribution of manufacturing concentration by county



Looking at Figure 2.21, there appears to be two "breaks" that can be used. For this purpose, the manufacturing data was broken into three categories. Those counties that have less than 6% of their labor force in manufacturing are classified as "Low." Counties having between 6% and 25% will be called "Medium" and above 25% are "High." The matching of pairs of counties is similar to the previous industrial categories except we have a third classification. The four matching criteria for the categories will be Low/Low, Medium/Medium, High/High, and a fourth that considers two counties a matching pair when each is in a different classification.

Figure 2.22: Spatial covariance of unemployment between counties based on manufacturing concentration annual from 1990-2006

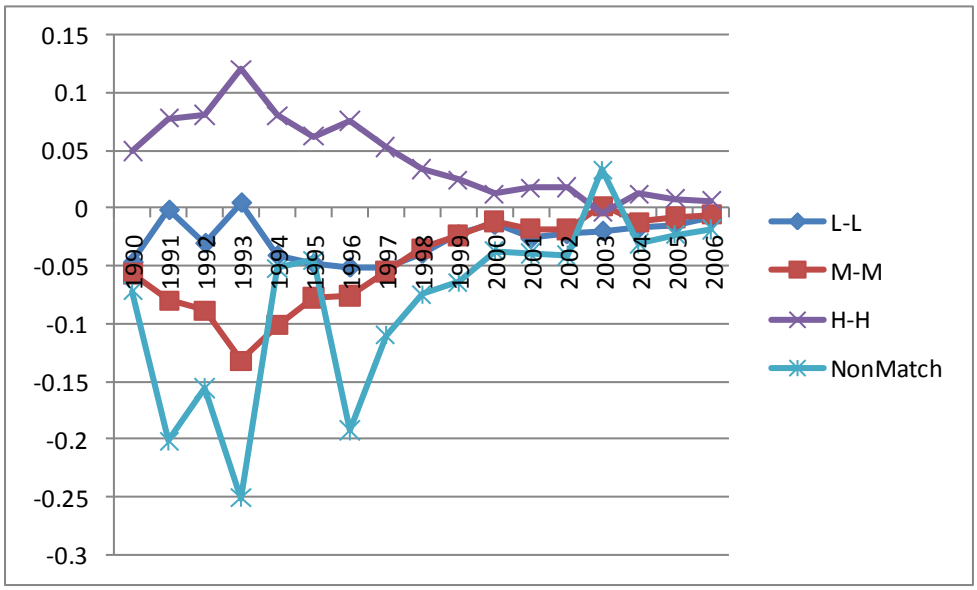


Figure 2.22 presents results based on Equations 2.1 and 2.2.

Perhaps the first comparison that catches the eye is the High/High pairs. Unemployment in counties with high levels of manufacturing are positively related, indicating that the unemployment rates at a given point in time for two manufacturing counties are similar. This degree of relationship weakens as lower concentration counties are compared.

Table 2.6: Descriptive statistics of correlations of unemployment over time for counties based on manufacturing concentration

CORRELATIONS	N of Pairs	Mean	Max	Min	Stdev
Low/Low	120	0.50042	0.97020	-0.12482	0.20091
Med/Med	2556	0.58588	0.98050	-0.17082	0.19392
High/High	351	0.56133	0.93529	-0.09447	0.20949
NonMatch	3528	0.54833	0.97705	-0.24465	0.20692

Figure 2.23: Distribution of correlation coefficients of unemployment rates over time for pairs of low concentration manufacturing counties

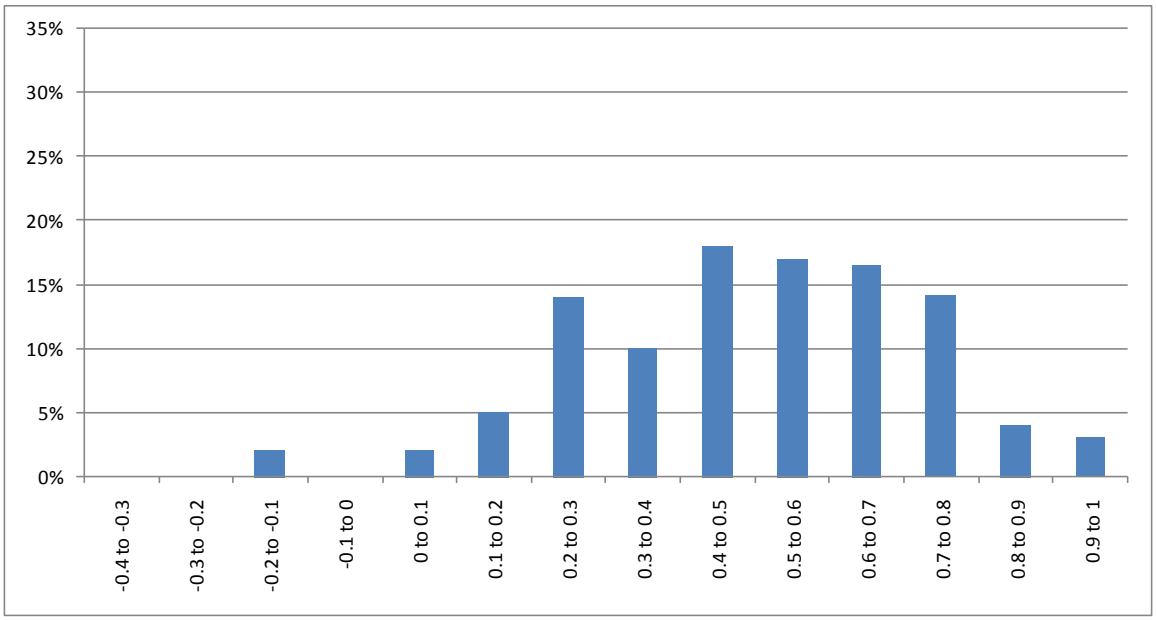


Figure 2.24: Distribution of correlation coefficients of unemployment rates over time for pairs of medium concentration manufacturing counties

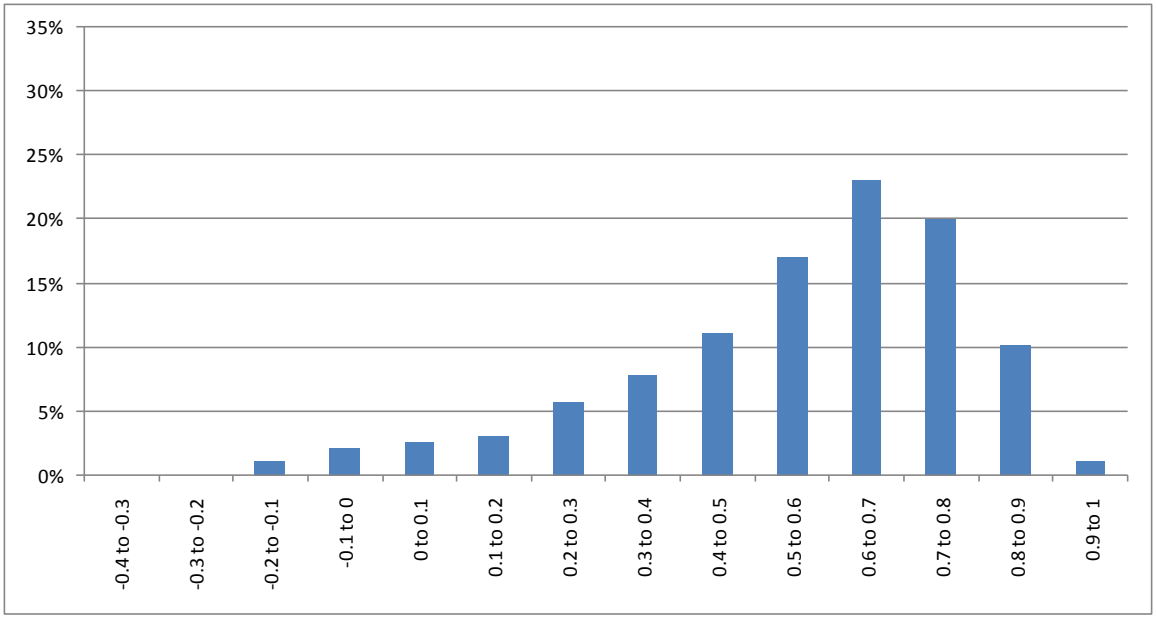


Figure 2.25: Distribution of correlation coefficients of unemployment rates over time for pairs of high concentration manufacturing counties

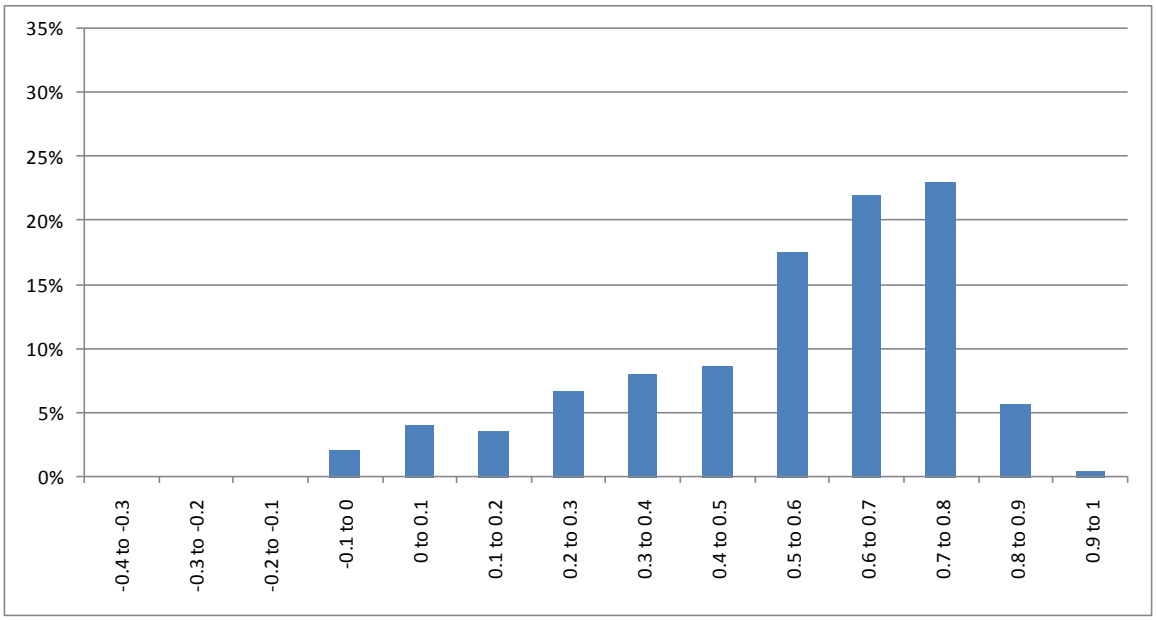
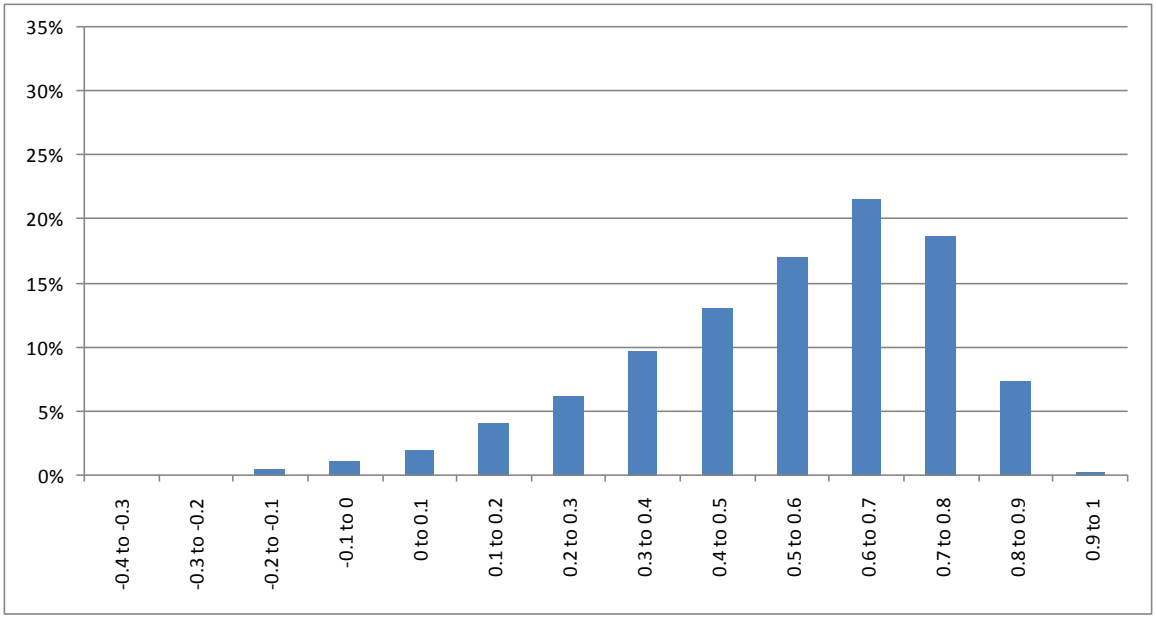


Figure 2.26: Distribution of correlation coefficients of unemployment rates over time for manufacturing counties of mixed concentration pairs



These graphs show that the Med/Med has the highest mean. The Low/Low has the lowest mean of the four categories given here. That is not unreasonable. The common thread between these counties is that they are low in manufacturing and nothing

else. Imagine a county that is low manufacturing and high agriculture while yet another is low manufacturing and high service sector. There is no reason to believe that these two counties are similar in their industrial profile.

Looking at the analysis based on manufacturing concentration, one interesting result is that in the covariance measure, based on unemployment rates at a single point in time, the High/High pairs yield the largest number, while in the correlations, measuring variation over time, Med/Med pairs produce the largest average correlation. This indicates that counties with high levels of manufacturing have similar levels of unemployment, but their unemployment rates do not move together over time. One possible explanation for this result is that diversified counties (such as the Med/Med pairs) follow the national or state unemployment rate more closely, and the high manufacturing counties are more tied to the specific markets that are served by particular industries that are concentrated in a county.

2.9 Education

The final consideration to explore is the concentration of higher education employment in a given county. In the absence of a measure of employment in higher education, student enrollment is used as a proxy for the overall impact of higher education in a county. To classify the counties, the number of students enrolled in a college/university was divided by the labor force size to arrive at a percentage. If the value calculated was 20% or larger, that county was classified as an education-heavy county and labeled "Ed." Those below 20% are labeled "Non-Ed." There will be three categories determined using the same method utilized in the agricultural case. Figure 2.27 presents the result of the covariance function at a given point in time.

Figure 2.27: Spatial covariance of unemployment between counties based on higher education concentration annual from 1990-2006

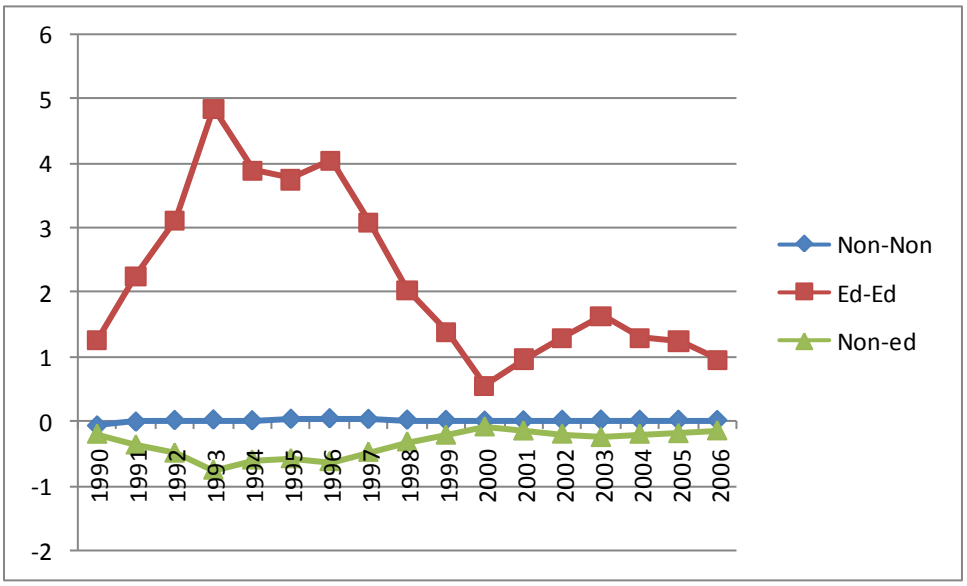


Figure 2.27 presents results based on Equations 2.1 and 2.2.

From this graph, we can see that the pairs of education-heavy counties display extraordinarily strong covariances. The NonEd/NonEd comparison is effectively zero for the entire series. Table 2.7 presents the mean correlations, over time based on this classification, and the subsequent graphs present the distributions of those correlation.

Table 2.7: Descriptive statistics of correlations of unemployment over time for counties based on educational concentration

CORRELATIONS	N of Pairs	Mean	Max	Min	Stdev
NonEd/NonEd	4950	0.55685	0.98050	-0.24465	0.20218
Ed/Ed	105	0.68256	0.95031	0.15743	0.16158
NonEd/Ed	1500	0.57399	0.97705	-0.17082	0.20492

Figure 2.28: Distribution of correlation coefficients of unemployment rates over time for non-educational county pairs

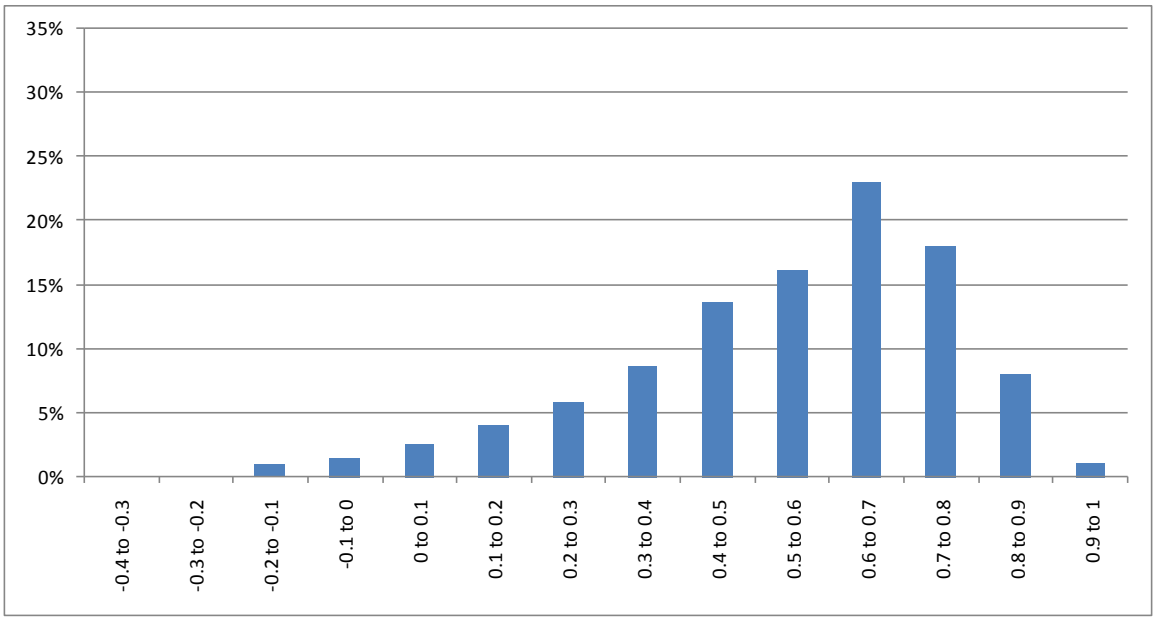


Figure 2.29: Distribution of correlation coefficients of unemployment rates over time for educational county pairs

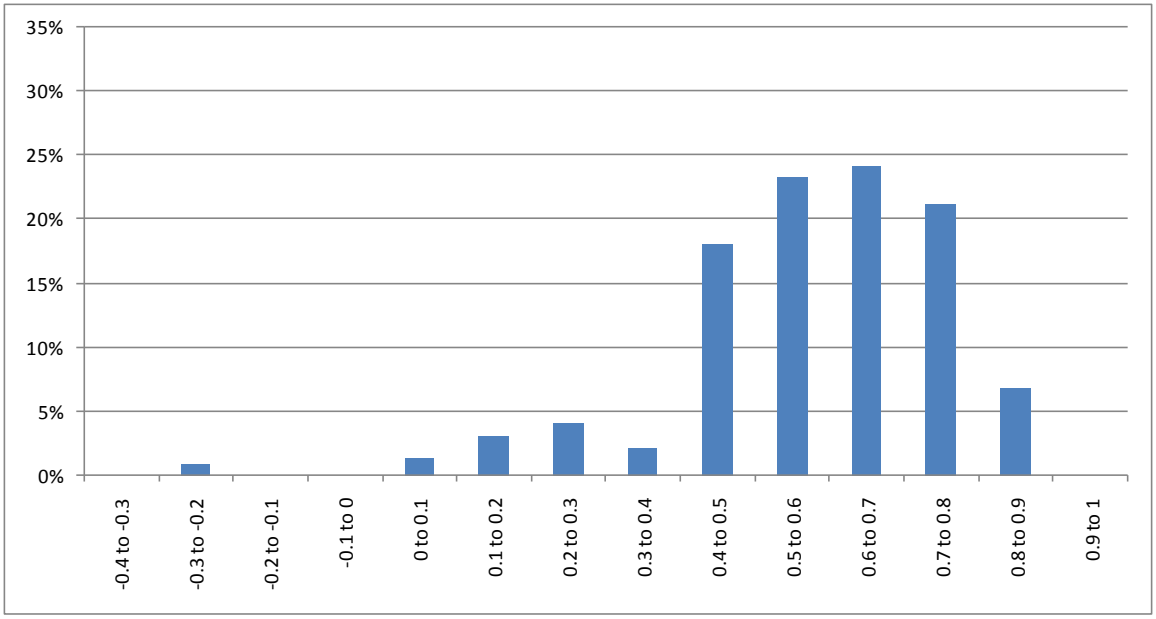
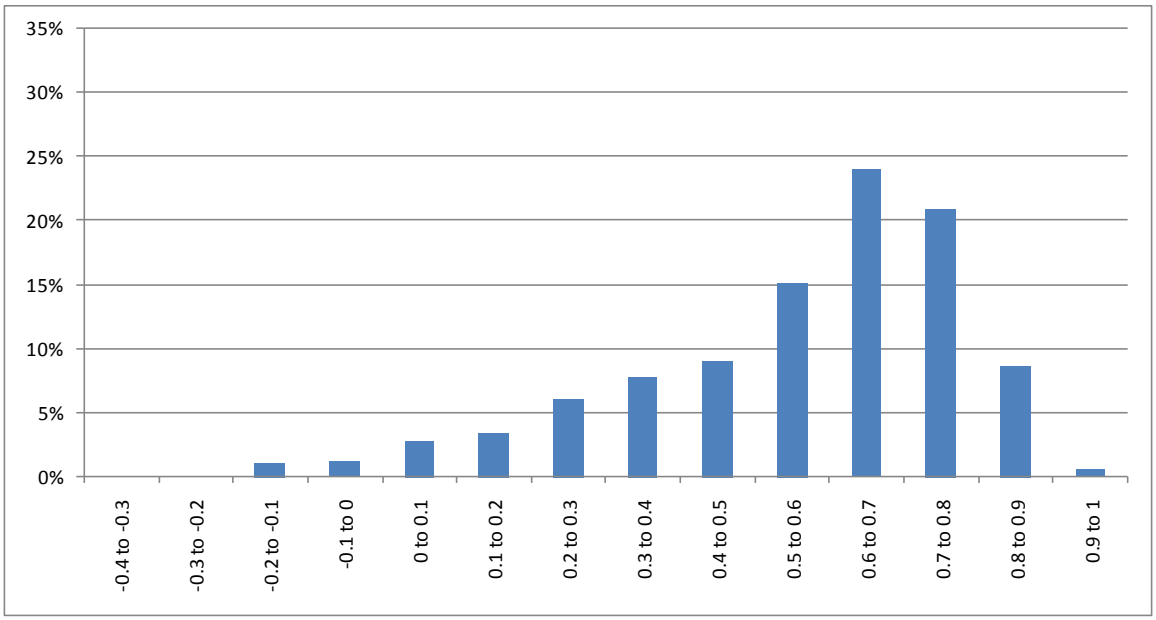


Figure 2.30: Distribution of correlation coefficients of unemployment rates over time for educational/non-educational county pairs



Upon examining the graphs and table, it appears that the education-heavy pairs are strongly related. Not only is there a strong covariance of unemployment at a single point in time for these counties, but also there is a much stronger relationship apparent in the correlation statistics focusing on changes over time, implying that the unemployment for counties with higher educational concentration also move together over time.

2.10 Moran's I

One of the most popular measures of spatial correlation in the literature is Moran's I. Taking x_i as a characteristic of location i (e.g., the unemployment rate), Moran's I is calculated as follows:

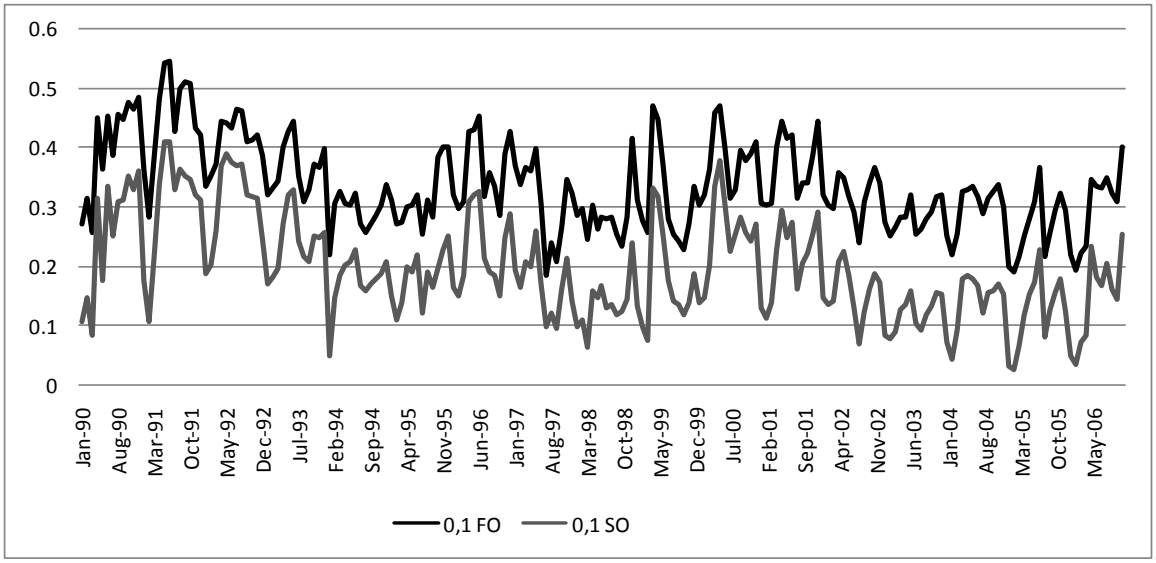
$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where w_{ij} is the weighting matrix (and $w_{ii} = 0$), $\bar{x}$ is the mean for x_i in the sample of locations, and n is the number of locations. The weighting matrix is constructed so that more proximate locations receive larger weights and more distant ones lower weights. The values of Moran's I normally range from -1 to 1, where positive (negative) values indicate that areas close to one another are more (less) likely to have similar values. I have calculated Moran's I for my dataset of Missouri's county-level unemployment rates using several different weighting schemes to be explained below.

2.11 Dichotomous Weights

The first set of results considers weights taking only values of either zero or one, where we use both first-order and second-order contiguity to determine these weights. A first-order neighbor of county i is one that shares a border with county i, and so only counties sharing a border are coded 1. For the second-order neighbor coding, both contiguous counties and those that share a border with a county that is a first-order neighbor are assigned a weight of 1.

Figure 2.31: Moran's I for first-order and second-order neighbors' county unemployment with dichotomous weights (0,1), monthly from 1990-2006

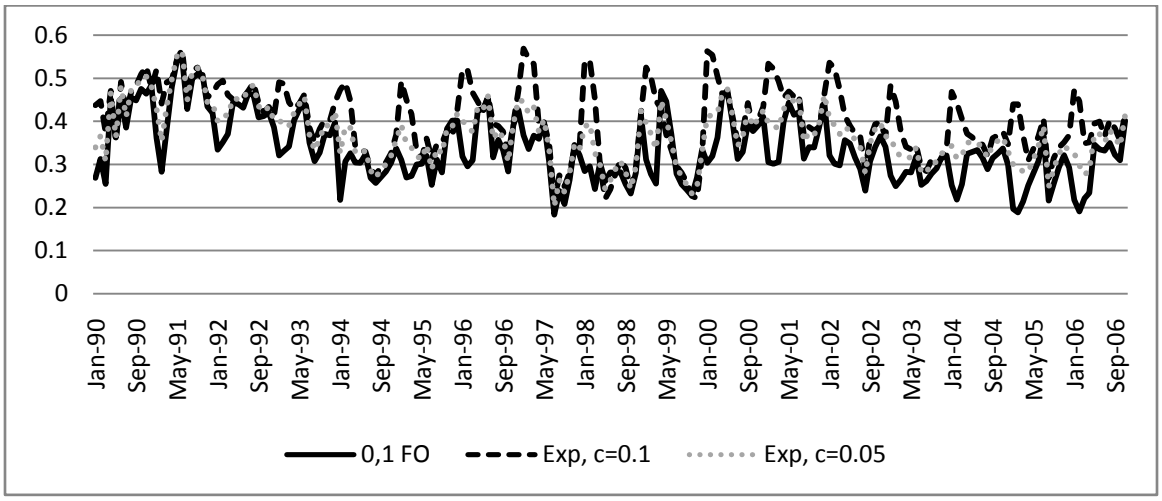


From this graph we can see that the first-order scheme gives higher values of Moran's I than the second-order scheme. This would indicate that neighbors that are further away are less strongly associated with the unemployment rate than those that are closer. This is not a surprise; this is a similar result to previous results found with the correlations at a single point in time.

2.12 Exponential Weights

The next set of analyses separately considers both first-order (FO) and first and second-order neighbors (SO) as before. In this section however the weight assigned is based on an exponential function. Despite its simplicity, the 0,1 weighting function has the potential drawback of treating each neighboring county's unemployment rate as of equal importance when determining the unemployment rate of county i. The exponential weighting function (used by Niebuhr (2003)) will factor the mileage between counties when assigning the weights to the neighbors of county i. The weight for county i and county j is taken as $w_{ij} = e^{(-cd_{ij})}$. In this formula, d_{ij} is the distance in miles between county centroids and c is a scaling constant. Several values of c were considered (see figures for specific values) for all inclusion scenarios. Those that showed statistical significance are presented in Figures 2.32-2.34. Figure 2.32 gives non-zero weights to only first-order neighbors, Figure 2.33 gives non-zero weights to first and second-order neighbors and Figure 2.34 gives non-zero weights to all counties in Missouri except when i = j (i.e., $w_{ii} = 0$). In Figures 2.32-2.34, the Moran's I for first-order contiguity using a 0,1 weighting scheme is included for comparison purposes.

Figure 2.32: Moran's I for first-order neighbors' county unemployment with exponential weights, including 0,1 weight based on FO contiguity, monthly 1990-2006



Note: The weight of Exp, c = 0.001, FO, is nearly identical to the 0,1 FO weight.

Figure 2.33: Moran's I for second-order neighbors' county unemployment with exponential weights, including 0,1 weight based on FO contiguity, monthly 1990-2006

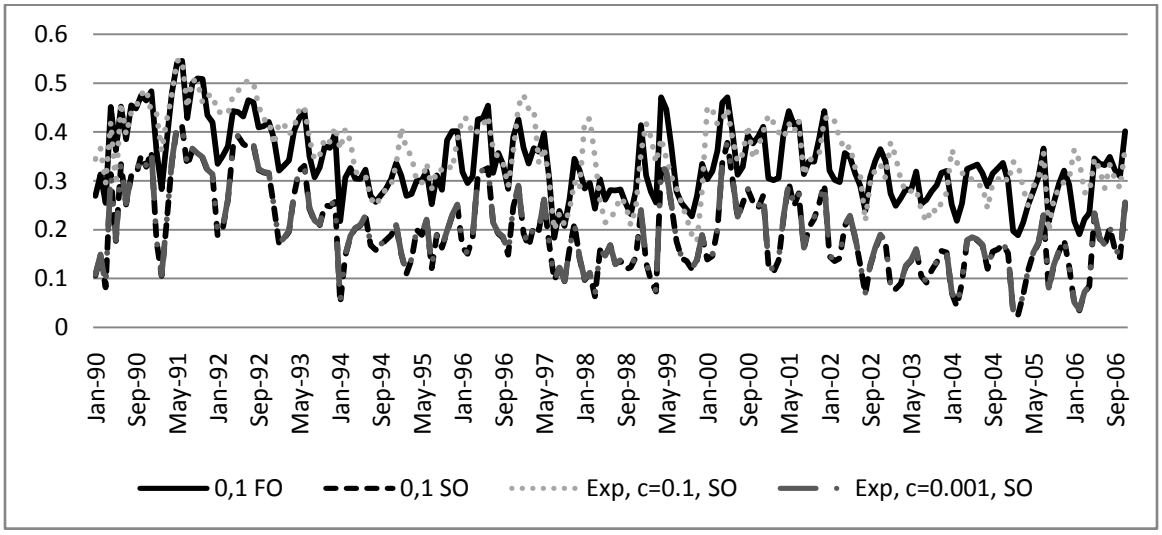
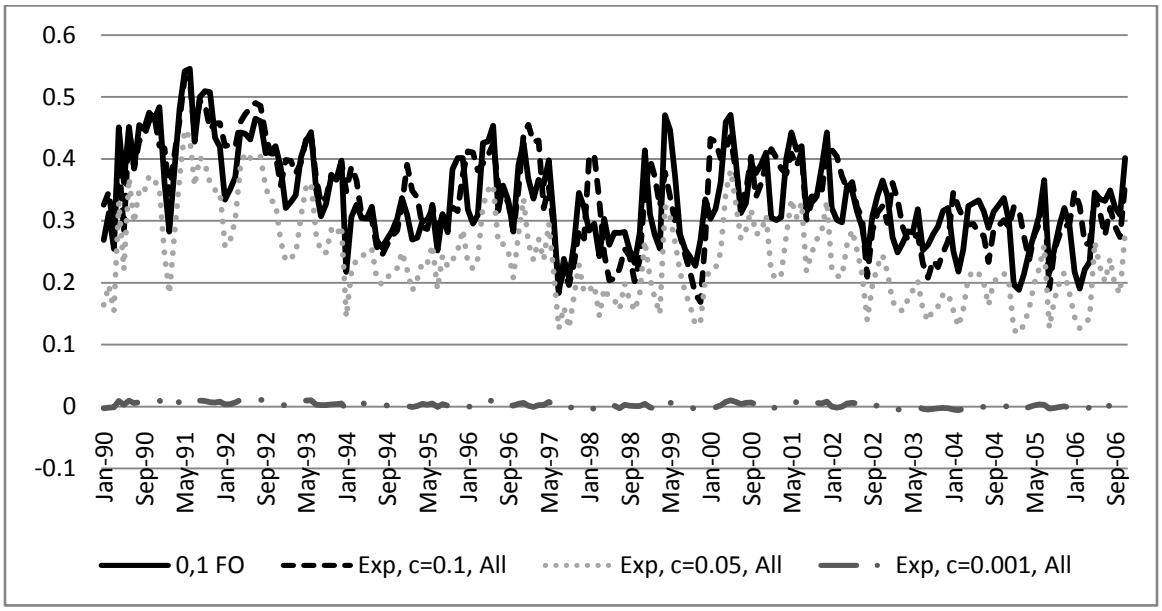


Figure 2.34: Moran's I for all county neighbors' county unemployment with exponential weights, including 0,1 weight based on FO contiguity, monthly 1990-2006



Comparing Figures 2.32 and 2.33, including second-order neighbors reduces the estimated value of Moran's I, even when weights are applied. Another interesting result is in Figure 2.34. Including counties that are neither first-order nor second-order neighbors leads to a lower value of Moran's I for almost every weight scheme and thus less overall correlation between neighbors. The exponential weight where the scaling constant (c) is 0.001 and all counties are included as neighbors, the value of Moran's I is nearly 0 for all months in the sample.

All of the above graphs also include the Moran's I for first-order contiguity using a dichotomous weighting structure. The purpose of this is to see the relative strength of using simple contiguity versus using a distance function combined with contiguity. We can see that the results are mixed. There are times that the 0-1 contiguity measure produces a higher Moran's I, but there are also months where the distance function in combination with contiguity gives a larger Moran's I. It is perhaps surprising that simple

contiguity is not consistently dominated by any distance function combined with contiguity.

Overall, these results indicate that not only does sharing a border matter for the relationship between two counties' unemployment rates but also the simple 0-1 coding based on contiguity generally provides equal or higher values of Moran's I than any of the exponential coding schemes that do not make use of either contiguity measure.

2.13 Inverse Distance Weights

I also considered a weighting scheme based on a power function of distance. In this setup, the non-zero weight will be $w_{ij} = (1/d)^c$, where again d is the distance in miles between county centroids. Several values of c were considered ranging from 0.5 to 3. With all three inclusion scenarios applied (only first-order contiguous, first and second-order contiguous, and all counties in the state), none of the inverse distance weighting schemes produced statistically significant measures of Moran's I.

2.14 Statistical Significance of Moran's I

While knowing the value of Moran's I is useful, it also is important to check for the statistical significance of our calculated values. We have found curiously high values of Moran's I for some of the weighting schemes that we have not presented here. For instance, in the exponential scheme when c = 0.25 we found some values of Moran's I very near to and greater than 1[3]. Two comments should be made about these results. First, none or few of these particular Moran's I values are statistically significant, leading us to conclude that they are not different from zero in a statistical sense. Despite the high value of Moran's I, the failure of statistical significance suggests that the high value is not

[3] Although Moran's I is normally described as being bounded between -1 and 1, it may extend beyond those bounds in a particular finite sample.

capturing a stable factor underlying the spatial structure of unemployment. Most values of Moran's I we found to be not statistically significant are omitted from our graphs.

One case in which Moran's I would have a high value but not be statistically significant would occur if the weighting structure gives relatively high importance to a small number of pairs of counties, so that the very high values of Moran's I driven by just a few pairs. This can be problematic in this particular dataset as there are 115 counties (thus resulting in 6,555 possible pairs) and to have the results driven by a small number would potentially not be a good representation of the spatial structure of the state as a whole. To test the significance of the calculated values of Moran's I, I have used the following Z-statistic (Griffith, 1987). The test for significance of Moran's I is a Z-test with Z-statistic as

$$\text{Equation 2.4} \quad Z_{stat} = \frac{I - E(I)}{StDev(I)}$$

where I is the Moran's I value being tested, $E(I) = -1/(n-1)$, and StDev(I) is the standard deviation of Moran's I. The standard deviation is simply the square root of the variance and the calculation of the variance is as follows:

$$\text{Equation 2.5} \quad Var(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n-1)(n+1)S_0^2}$$

$$\text{Equation 2.6} \quad S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}, i \neq j$$

$$\text{Equation 2.7} \quad S_1 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(w_{ij} + w_{ji}\right)^2, i \neq j$$

$$\text{Equation 2.8} \quad S_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{ij} + \sum_{i=1}^{n} w_{ji}\right)^2$$

Below are plots of the various Z-statistics on those Moran's I's that showed statistical significance.

Figure 2.35: Z-statistics for Moran's I for first-order neighbors' county unemployment with three separate weighting structures

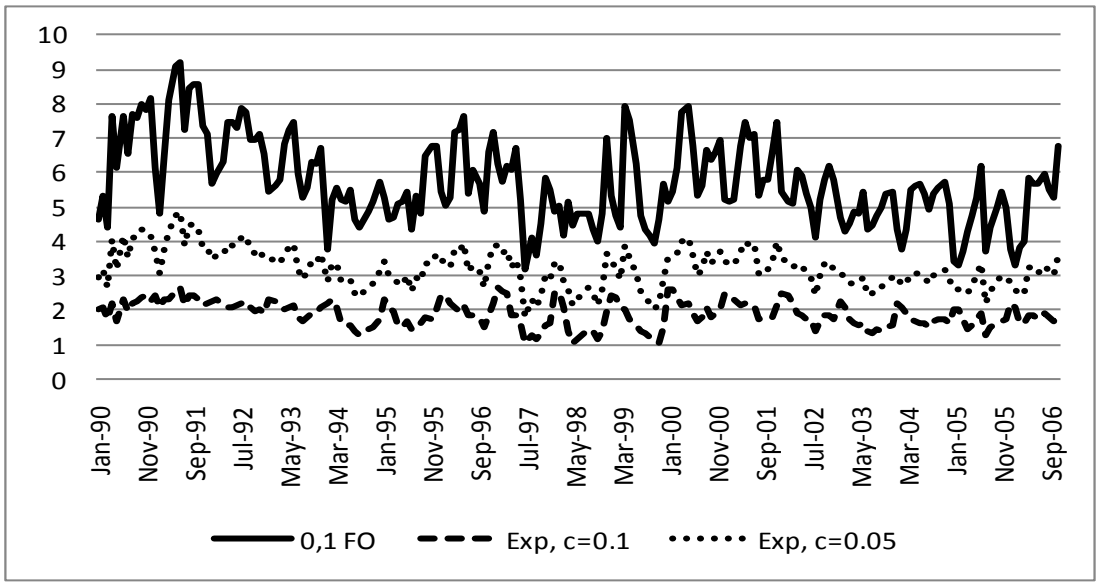
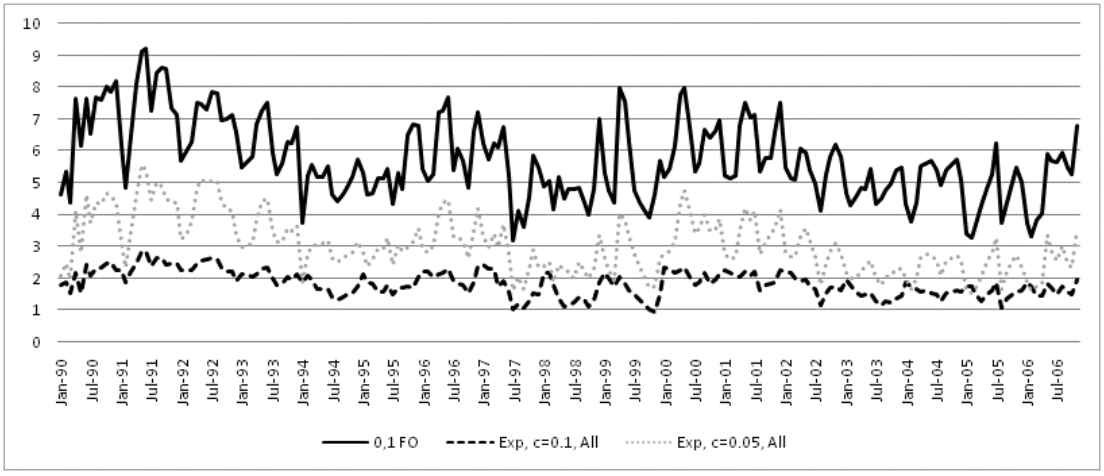


Figure 2.36: Z-statistics for Moran's I for county unemployment with exponential weighting structures and various inclusion scenarios (first-order neighbors and all counties)



Looking at the graphs of the Z-statistics, one common result that will color the remainder of the analysis done here is the strength (in terms of statistical significance) of the 0,1 FO weighting structure. Recall that the two graphs of Z-statistics were limited to those weighting schemes that produced statistically significant results for at least most of

the months in the sample. The inference we can draw from this is that the simple 0,1 FO weights do as well at capturing the spatial structure of county-level unemployment in Missouri as any other weighting structure estimated by this analysis. It is this result that allows future analyses done on this dataset to use the 0,1 FO weighting structure as a starting point.

It is important to remember that the goal of this research is to estimate the underlying social process that drives unemployment and its spatial component. Since the ultimate goal if this research is to build predictive models and test their accuracy and effectiveness, Z-statistics are a useful tool in guiding us towards the appropriate setup of these forecasting models.

2.15 Summary

In this chapter, we have learned several things about the spatial structure of county-level unemployment in the state of Missouri. In the first part of the chapter, the importance of both distance and several industrial characteristics were examined. These results indicated that the distance between two counties was more important in explaining the patterns of unemployment between a pair of counties than other characteristics of the counties. The industrial makeup of a particular county does appear to have an impact in determining spatial correlation between unemployment rates. There is a modest relationship between unemployment rates for counties that have similar industrial structures and it appears that these effects are important enough to keep track of in later analyses. They do not however show the same importance as distance between two counties. On a related note, all of these relationships (whether based on a distance measure or on industrial composition) weaken at later points in time. In every measure of

spatial correlation calculated, the dependence between counties is stronger in earlier months of the dataset than in the later months.

Examination of all the measures of spatial correlation shows that county contiguity captures much of the spatial dependence in county unemployment. This can be seen in the covariance functions calculated in the earlier sections of this chapter, in the correlations examining patterns of variation in unemployment over time, and in the Moran's I values examined most recently in the text.

The calculated values of Z most clearly suggest that county contiguity captures the spatial character of unemployment by county. This becomes important when moving on to building the predictive models later. The highest values of Z were returned from the first-order contiguous weighting schemes, though the second-order contiguous weights produced only slightly weaker Z-values. The Z-values are much weaker in every weighting scheme considered–regardless of the parameter chosen for the exponential or inverse distance relationship–that does not use contiguity. This lends statistical strength to the claim that contiguous counties are where most of the spatial dependence exists in Missouri's county-level unemployment rates.

There are two general results that seem to be consistent throughout the data work. The first one is that distance matters. In every comparison above, more proximate counties tend to have more closely correlated unemployment rates than those that are more distant. This will be important when specifying predictive models. The second point is related to this first point in that while distance does seem to matter it has become less important in more recent years. In those tests that allow comparisons over time, the spatial dependence softens as the data moves from the 1990s to the 2000s. This may

indicate a structural change in the spatial dependence of county-level unemployment in Missouri.

3. ESTIMATION OF MODELS

The previous calculations provided a broad picture of the spatial structure of county-level unemployment in Missouri. Since the ultimate goal of this paper is to study the forecasting accuracy of spatial models, the next step is to construct models that will be used to explore model specifications for the structure of unemployment in Missouri.

The first step in the analysis will be to explore what the data suggest is the most appropriate setup for non-spatial models. I begin working with a first order autoregressive model (AR(1)) and checking the outcomes when we restrict estimated parameters to be the same for all counties in Missouri. It is important to distinguish here our use of the term autoregressive model as often an autoregressive or AR model places this structure in the error of the model. In this research, autoregressive is used to refer to a model where a lagged dependent variable is inserted and the order of the autoregressive structure (e.g., AR(1), AR(2)) indicates how many lagged dependent variables will be included.

To estimate these models, the entire sample is used. These full sample models are estimated using OLS, and no predictions outside the sample are considered. This is in contrast to what is done in chapter 4, where forecast models are produced and out-of-sample forecasting accuracy is calculated. Chapter 4 also considers different sample sizes to estimate values used in the forecasting procedure. It is important to note that while both this chapter and chapter 4 explore many of the same factors (e.g., number of lags included, including counties with similar industrial profiles), there are significant differences in the methods used. In this chapter, the focus is on the r-squared of the OLS

regressions, the amount of variance in the sampling error, and F-tests between nested models.

These models will help to determine if estimated constants and coefficients should be permitted to vary across counties or be restricted to equality. We do find that allowing a different coefficient for each county appears to be the more appropriate model structure. The same tests and procedures are also performed on an AR(2) model. With the extra lag included we have some added possible combinations of restrictions on coefficients to examine. We also conclude that coefficients should be allowed to vary in these models.

Both the AR(1) and AR(2) models are estimated including monthly dummies to capture the seasonality in the unemployment data. These seasonal models are compared to the non-seasonal models to determine the effects of factoring out the seasonality. We also test whether the inclusion of these dummies affects our results regarding the restrictions on the intercepts and coefficients. The results show that including a seasonal component in the models does have an impact on the significance of the coefficient on the second temporal lag, but otherwise has only small effects on estimated coefficients.

Next, various spatial models are estimated. The creation of the spatial variable is done in the following fashion. For a given county i, the first-order spatial variable is constructed using unemployment numbers from all counties that share a border with that county. For each of the first-order neighbors, the number of people employed and unemployed in each county is collected. These numbers are then summed across all neighbors and the unemployment rate of the neighbors as a single group is calculated. This gives a single unemployment rate for the neighboring area as a whole rather than for

each neighboring county. The second-order spatial neighbors consist of all counties that share a border with a first-order neighbor of county i but do not share a border with county i. The number of people employed and unemployed for these counties is again combined in the same way as for first-order neighbors and the unemployment rate is calculated for the group as a whole giving a single unemployment rate for the second-order spatial neighbors. These variables are lagged temporally one month for the models, and counties outside the state of Missouri are included in the spatial variables where appropriate. The first and second temporal lags are included as well as a first-order and second-order spatial lags. We again estimate different models where the coefficients are allowed to vary across counties as well as restricting them to be the same for all counties in Missouri. As with the non-spatial models, we find that allowing separate coefficients across counties is a more appropriate model specification. Seasonal dummies are also included and the same conclusion is reached.

We also introduce the unemployment rate of a larger region as a predictor in the spatial models. Three different measures of unemployment are considered: the U.S. unemployment rate, the Midwest's unemployment rate, and Missouri's unemployment rate. The purpose of this is to determine to what extent the observed spatial model effects are due to conditions that affect all counties. All three measures of regional unemployment are found to add explanatory power to the models. We conclude by estimating a spatial model that incorporates all factors that have been analyzed in this section.

The inclusion of the spatial neighbors provides some interesting results. The first-order spatial variable does have a significant coefficient regardless of whether the

coefficients are constrained to be constant across counties or allowed to vary across counties. The models where coefficients are allowed to vary across counties appear to provide better fit, as indicated by an F-test. This property is true of second-order neighbors as well, though the F-statistics for including second-order neighbors is smaller than the F-statistics for including only first-order neighbors. The inclusion of a seasonal component has an impact on the significance of the spatial lags, but this impact is quite small.

The results also show that including other predictors in the spatial models (e.g., county-characteristic predictors, distinguishing out-of-Missouri counties, regional unemployment) does affect the spatial lag coefficients, but this effect depends on which predictor is included. We have also fitted models that include an extra predictor which measures the unemployment rate in like counties. For example, the agricultural counties include a predictor measuring the unemployment rate in all other agricultural counties in the state. Among these models, measures capturing unemployment in counties with similar levels of manufacturing and educational concentration have the largest impact, while the agricultural concentration's effect is quite small. There is also a modest impact of unemployment in metropolitan counties for other metropolitan counties. Separating out-of-Missouri counties from the Missouri counties in the spatial lag also alters some of the numerical results as does the inclusion of a larger region's unemployment rate. Despite the impact on the coefficients being relatively small in most cases, further testing indicated that including these extra predictors is appropriate in most models. The overall results of this chapter will go a long way in helping to estimate forecasting models in later analysis.

3.1 Non-Spatial AR Models

The general structure of the non-spatial models will be an autoregressive model of order p, with p determined using procedures to be discussed later. Initially, we will estimate models where p = 1, thus including only one temporal lag of unemployment for each county.

There is an initial question of whether each county will require a separate constant term (α_i) or a separate coefficient for its lag (β_i). It is possible that the processes at work for counties are so similar that using 115 constants is not a valuable use of the limited degrees of freedom. To answer this question, five different specifications of an AR(1) model are estimated. The first restricts each county to have the same constant and coefficient on the lagged variable. This model is described by Equation 3.1.

Equation 3.1 $$Y_{it}=\alpha + \beta Y_{i,t-1} + \varepsilon_{it}$$

where Y_{ti} is the unemployment rate for county i at time t. A second model assumes that all counties in Missouri have the same constant but allows different coefficients on lagged unemployment for each county. This model is described by Equation 3.2.

Equation 3.2 $$Y_{it}=\alpha + \beta_i Y_{i,t-1} + \varepsilon_{it}$$

A third allows constants to vary across counties but restricts the coefficient on the lagged value of unemployment to be the same across counties. This is depicted by Equation 3.3.

Equation 3.3 $$Y_{it}= \alpha_i + \beta Y_{i,t-1} + \varepsilon_{it}$$

The fourth model assumes that all counties have both differing constant terms and differing coefficients on the lagged value of unemployment as shown by Equation 3.4.

Equation 3.4 $$Y_{it}= \alpha_i + \beta_i Y_{i,t-1} + \varepsilon_{it}$$

A fifth and final specification includes no temporal lags of unemployment. This simply reports the constant value for each county individually and is listed in Equation 2.5.

Equation 3.5 $$Y_{it} = \alpha_i + \varepsilon_{it}$$

Estimating the parameters on these five models was done using 114 dummy variables for the intercept terms (when the intercepts were allowed to vary) and 115 coefficients (when the coefficients were allowed to vary) simultaneously in a single procedure[4]. In other words, we did not estimate each of these county's models separately, but rather estimated coefficients jointly assuming a single independent error term. Of course, for the unrestricted model, coefficient estimates are exactly the same as those that would be obtained if each county's model was estimated separately, although standard errors and related statistical tests would differ slightly.

In Equation 3.1, both the restricted constant and restricted coefficient were statistically significant ($p < 0.0001$). The second model showed a significant constant term and 115 significant temporal lags. The third model had many (though not all 115) significant constant terms and the single coefficient on the lag was also significant. The fourth model showed each of the 115 lagged values to have significant coefficients ($p < 0.0001$). They also were all positive. Only a few of the constant terms, however, were statistically significant. A quick summary of these models is presented in Table 3.1. For models with multiple intercepts/coefficients, the average of those values is reported (these cases are listed in italics). Otherwise, the single intercept/coefficient of the model is listed.

[4] Estimating parameters simultaneously is done in future cases where coefficients/constants are allowed to vary across counties. The parameter estimates from this method are identical to the results that would have been obtained from estimating each county's equation separately, however, measures of statistical significance could differ. Those cases where both methods were used, the conclusions from both methods were the same in terms of statistical significance.

Table 3.1: Summary of AR(1) models with various restrictions on coefficients and intercepts for county-level unemployment from January 1990-November 2006

	Constant	Coefficient	Std Error of Coefficients	r-squared	Adjusted r-squared	n
Equation 3.1	0.5734	0.8979	0.0028	0.8064	0.8064	23,229
Equation 3.2	0.9912	*0.8102*	*0.0152*	0.8137	0.8128	23,229
Equation 3.3	*-0.4046*	0.8288	0.0036	0.8137	0.8127	23,229
Equation 3.4	*0.1271*	*0.8340*	*0.0516*	0.8184	0.8166	23,229
Equation 3.5	*-2.3187*	N/A	N/A	0.3995	0.3966	23,229

Entries in italics are mean values for coefficients that vary across counties.

Note that in all four estimation structures that have temporally lagged unemployment as a predictor[5], and the coefficient or average coefficient on the lagged term is between 0.81 and 0.89. Additionally, the r-squared and adjusted r-squared are almost identical for all four models. Equation 3.5 shows a large drop in r-squared from the other four models indicating the strength in explanatory power given by the addition of lagged unemployment.

Because we have some models that allow differing values (constant or coefficient) across counties, we can also look at several groups of counties and see how the averages of these values behave as a group. Table 3.2 reports average values of the constant estimate by groups of counties while Table 3.3 gives the average values of the coefficient estimate by the same groups. The counties are grouped together by the characteristics considered in the previous chapter. Agricultural concentration, manufacturing concentration, educational concentration and metropolitan status define the groups analyzed. Any county that has over 1% of its labor force in agriculture is designated an agricultural county. Otherwise the county is non-agricultural. A similar procedure is used for manufacturing and educational concentration, although the

[5] The presence of a unit root was not addressed in this analysis. However, the coefficients on the first lag are always less than 1 and the distance from 1 is large based on the standard error.

thresholds are different. For manufacturing, any county with 6% or less is considered low manufacturing and any county over 25% is considered high manufacturing. For educational concentration, any county with over a 20% ratio of students in higher education to labor force size is high education[6]. A fifth group is added that includes the top fifteen counties in terms of labor force size.

Table 3.2: Average value of the constant term of AR(1) models by county characteristic, January 1990-November 2006

	Top 15 (LF)	Ag	Manuf (H)	Manuf (L)	Educ	Metro
Equation 3.3	-0.2940	0.2096	0.0684	0.3479	-0.2021	0.0172
(Std Error)	(0.1030)	(0.0663)	(0.0697)	(0.0889)	(0.0955)	(0.0658)
Equation 3.4	-0.4364	-0.3589	-0.4101	-0.4315	-0.6705	-0.5413
(Std Error)	(0.0289)	(0.0197)	(0.0214)	(0.0279)	(0.0289)	(0.0191)
N (counties)=	15	32	27	16	15	34

Table 3.3: Average value of the lagged unemployment coefficient of AR(1) models by county characteristic, January 1990-November 2006

	Top 15 (LF)	Ag	Manuf (H)	Manuf (L)	Educ	Metro
Equation 3.2	0.7714	0.8246	0.8161	0.7958	0.759	0.7885
(Std Error)	(-0.005)	(-0.0025)	(-0.0029)	(-0.004)	-0.01	(-0.003)
Equation 3.4	0.8881	0.8276	0.8469	0.8018	0.858	0.8332
(Std Error)	(-0.0184)	-0.0089	(-0.0097)	(-0.0135)	(-0.02)	(-0.0108)

Looking at the results of the constant term from Equation 3.3, we see that the values jump around based on different county characteristic. Recall that in these models, the coefficient on the lag is held constant across counties. When this restriction is removed (i.e., coefficients on the lag are allowed to vary across counties), the constant terms become very similar. This implies that the constant term that is varying across counties is picking up variation that is missed because coefficients on the lag are held

[6]The detailed specifications for these groups can be found in chapter 2.

constant across counties. When this restriction is removed, the constant terms become similar. The average value of the lag coefficient is quite similar regardless of whether or not the constant term is allowed to vary across counties or is restricted to be the same across counties.

In models that allow coefficients to differ across counties, variation in estimates reflects both sampling error and variation in true coefficient values. It is possible to identify the relative importance of these two factors. Using the reported standard errors and the variance of the parameter estimates, we can estimate the extent of sampling error in the models that allow coefficients or constants to vary. If we examine the equation for the variance of the estimate for β_i across counties which is denoted by $\hat{\beta}_i$, we can see the following

$$\mathrm{Var}(\hat{\beta}_i) = \mathrm{Var}(\hat{\beta}_i - \beta_i) + \mathrm{Var}(\beta_i)$$

This reflects the fact that the difference between the unbiased estimate and the actual value should be independent of the actual values. A simple rearrangement of this yields

$$\mathrm{Var}(\beta_i) = \mathrm{Var}(\hat{\beta}_i) - \mathrm{Var}(\hat{\beta}_i - \beta_i)$$

If the square root is taken of this term, we have then

$$\mathrm{StDev}(\beta_i) = \sqrt{\mathrm{Var}(\hat{\beta}_i) - \mathrm{Var}(\hat{\beta}_i - \beta_i)}$$

The left-hand side represents the standard deviation of the true value of β_i, which is not observable but is estimated by $\hat{\beta}_i$. The first term inside the square root on the right is easily calculated from the observed estimates. The second term on the right side of the equation is estimated by the standard errors given by regression analysis. In particular, we substitute the estimate as follows

$$\mathrm{Var}\left(\hat{\beta}_i - \beta_i\right) = \frac{1}{n}\sum_{i=1}^{n}(\text{Standard Error})^2$$

This term represents the amount of variation due to sampling error[7].

This gives us an estimate of the standard deviation of the true values of β_i which is generally unobservable. This calculation would also apply to the intercept terms (α) that are allowed to vary across counties. These results are listed in Table 3.4.

Table 3.4: Estimates of the true variance of α or β, variance of estimates of α or β, and standard error of estimated values of α or β

	Calculated variance of true value	Variance of estimate	(1/n)*Σ(SE^2)
	(Std. Dev. of true value)	(Std. Dev. Of Estimate)	√[(1/n)*Σ(SE^2)]
Equation 3.2 coeff.	0.0023	0.0025	0.0002
	(0.0481)	(0.0506)	(0.0157)
Equation 3.3 const.	0.0947	0.1072	0.0124
	(0.3078)	(0.3274)	(0.1115)
Equation 3.4, const.	0.0856	0.2225	0.1369
	(0.2925)	(0.4717)	(0.3700)
Equation 3.4 coeff.	0.0030	0.0060	0.0029
	(0.0553)	(0.0775)	(0.0543)

What we can see from this table is that a significant amount of variation is expected to be due to variation in the true values of the coefficients (first and fourth rows). This gives weight to the idea of allowing varying coefficients across counties. We can see that the standard deviation across counties amounts to about 6% of the mean value of the estimate. If the more appropriate model was to specify coefficients to be the same for all counties, we would expect a much smaller amount of variation to be present in the true value of β than what is observed here.

[7] Error terms are assumed independent. Although standard errors for coefficients in a single equation are not generally independent, for a large number of dummy variables, as is the case here, independence holds in practice.

In comparing the models listed above, we can identify cases where one model is nested in another. In these comparisons, we conducted a standard restricted F-test. The test statistic is

$$F_{stat} = \frac{(SSE_{restricted} - SSE_{unrestricted})/q}{SSE_{unrestricted}/(n-k)}$$

This test compares the sum of the squared errors (SSE) of the models. The value q is the number of restrictions placed on the restricted model, while n is the number of observations and k is the number of predictors (including the constant), with k measured for the unrestricted model. The results of these comparisons are reported in Table 3.5.

Table 3.5: Pairwise comparison of AR(1) models with coefficients that are constant/vary across counties

	F_{stat}	p-value
3.1 v 3.4	6.615	<0.0001
3.3 v 3.4	5.278	<0.0001
3.1 v 3.3	7.776	<0.0001
3.2 v 3.4	5.192	<0.0001

Based on the results reported in Table 3.5, it appears that the least restricted model, which has a separate constant and coefficient for each county is the best of the group. In all cases, the restricted F-test indicates that using separate coefficients gains in explanatory power more than it costs in terms of degrees of freedom.

3.2 Adding Additional Lags

The next section will consider models that have two temporal lags of county unemployment. Much of this section will closely correspond to analysis from the previous section.

There are eight models that will be estimated. All models will include an intercept term, a first temporal lag, and a second temporal lag; the variation across models will be in whether one or more of these terms will be held constant across counties or allowed to vary. The first model, displayed by Equation 3.6, is the most restrictive of the group, forcing the constant and coefficients to be the same for each county. Notice that Equation 3.6 is an extension of Equation 3.1 where p = 2 instead of 1.

Equation 3.6 $Y_{it}=\alpha+\beta_1 Y_{i,t-1}+\beta_2 Y_{i,t-2}+\varepsilon_{it}$

The remaining models will allow selected coefficients or constants to vary across counties. Equations 3.7-3.13 list the various combinations that will be estimated. While many of these are extensions of previous equations, the specific functional forms of the equations are provided in detail.

Equation 3.7 $Y_{it}=\alpha_i+\beta_1 Y_{i,t-1}+\beta_2 Y_{i,t-2}+\varepsilon_{it}$

Equation 3.8 $Y_{it}=\alpha+\beta_{i1} Y_{i,t-1}+\beta_2 Y_{i,t-2}+\varepsilon_{it}$

Equation 3.9 $Y_{it}=\alpha+\beta_1 Y_{i,t-1}+\beta_{i2} Y_{i,t-2}+\varepsilon_{it}$

Equation 3.10 $Y_{it}=\alpha+\beta_{i1} Y_{i,t-1}+\beta_{i2} Y_{i,t-2}+\varepsilon_{it}$

Equation 3.11 $Y_{it}=\alpha_i+\beta_{i1} Y_{i,t-1}+\beta_2 Y_{i,t-2}+\varepsilon_{it}$

Equation 3.12 $Y_{it}=\alpha_i+\beta_1 Y_{i,t-1}+\beta_{i2} Y_{i,t-2}+\varepsilon_{it}$

Equation 3.13 $Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \beta_{i2} Y_{i,t-2} + \varepsilon_{it}$

Clearly the most restrictive model is 3.6, which requires the intercept and both of the coefficients for the lags to be constant across the 115 counties in the sample. Table 3.6 presents a summary of the models estimated. When the parameter estimate was allowed to vary, the average of the 115 estimates is reported and is presented in italics.

Table 3.6: Summary of AR(2) models with various restrictions on coefficients and intercepts for county-level unemployment from January 1990-November 2006

	Constant	First Lag	Second Lag	r-squared	Adj r-squared	n
Equation 3.6	0.6169	0.9945	-0.1065	0.8185	0.8192	23,113
Equation 3.7	*-0.4389*	0.9543	-0.1445	0.8265	0.8261	23,113
Equation 3.8	1.0824	*0.9370*	-0.1462	0.8271	0.8263	23,113
Equation 3.9	0.9883	0.9547	*-0.1451*	0.8281	0.8272	23,113
Equation 3.10	1.0394	*0.8659*	*-0.0664*	0.8368	0.8354	23,113
Equation 3.11	*0.2156*	*0.9875*	-0.1756	0.8331	0.8312	23,113
Equation 3.12	*0.4682*	0.9219	*-0.0787*	0.8398	0.8384	23,113
Equation 3.13	*0.3685*	*0.8837*	*-0.0517*	0.8456	0.8435	23,113

Entries in italics are mean values for coefficients that vary across counties.

Compared to the AR(1) models previously estimated, the first lag's coefficient is greater, although the increase is modest[8]. There is also a fairly tight range in which the reported estimate for the first temporal lag remains (between 0.86 and 0.99). All are statistically significant at conventional levels. The range is slightly wider than the AR(1) cases. The second lags all are negative and small. This would indicate that these models are picking up a time trend in the data[9]. Their significance is also less clear. In all cases where the second lag's coefficient is restricted to be constant across the counties it is statistically significant ($p < 0.0001$). When this parameter estimate is allowed to vary across counties, the story changes. Table 3.7 summarizes the frequency of statistically significant second lag coefficients based on three different alpha levels.

[8] Unit root tests were not performed in this analysis.

[9] Note that $\beta_1 Y_{t-1} + \beta_2 Y_{t-2}$ can be written as $(\beta_1 + \beta_2)Y_{t-1} - \beta_2(Y_{t-1} - Y_{t-2})$, so the coefficient $-\beta_2$ can be interpreted as a trend effect.

Table 3.7: Number of statistically significant coefficients on the second temporal lag in AR(2) models when the coefficient varies across counties for selected alpha levels

	0.1	0.05	0.01
Equation 3.9	115	115	115
Equation 3.10	34	27	19
Equation 3.12	37	25	10
Equation 3.13	52	36	26

Three of the four specifications have a consistent trend. Equation 3.9 is an interesting case. This model allows only the coefficients on the second lag to vary; both the intercept term and coefficient on the first lag are constrained to be constant across counties. The finding that the coefficient is statistically significant for every county suggests that the second lag is picking up variation across counties that the first lag fails to capture (as it is held constant). In all other models listed in Table 3.7, the first lag's coefficient is allowed to vary. In these cases it is capturing most (though not all) of the variation across counties. This coincides with the conclusions based on the AR(1) models, which showed that allowing coefficients to vary across counties improved prediction.

In a similar fashion to what was done with the AR(1) models, we present estimates of parameters for counties by selected county characteristic groups. For the constant term and the coefficient on the first temporal lag of unemployment, the values are similar to those found in tables 3.2 and 3.3 and are not presented here. The second temporal lags however show some differences between characteristic groups. In percentage terms, variation is much larger than that for the first lag coefficient, although differences are similar in magnitude.

Table 3.8: Average value of the coefficient of the second temporal lag of AR(2) models by county characteristic, January 1990-November 2006

Second Lag	Top 15 (LF)	Ag	Manuf (H)	Manuf (L)	Educ	Metro	Con	1st	2nd
Equation 3.9	0.1798	0.1301	0.1383	0.1684	0.1927	0.1635			
(Std Error)	(0.0049)	(0.0026)	(0.003)	(0.0041)	(0.0051)	(0.003)	C	C	V
Equation 3.10	0.0153	0.0876	0.0678	0.1261	0.0742	0.0031			
(Std Error)	(0.0393)	(0.0151)	(0.0174)	(0.0222)	(0.0349)	(0.0201)	C	V	V
Equation 3.12	0.0023	0.092	0.0642	0.1457	0.0475	0.0556			
(Std Error)	(0.0172)	(0.0084)	(0.0092)	(0.0128)	(0.0162)	(0.0101)	V	C	V
Equation 3.13	0.0793	0.0848	0.0479	0.1298	0.0204	0.0251			
(Std Error)	(0.0391)	(0.0153)	(0.0175)	(0.0226)	(0.0349)	(0.0202)	V	V	V
n(counties) =	15	32	27	16	15	34			

The far right columns of Table 3.8 refer to the constant term (Con), the coefficient on the first temporal lag (1st) and the coefficient on the second temporal lag (2nd). A "C" indicates that value is held constant across counties. A "V" indicates that value is allowed to vary across counties.

We can also estimate the variation in the true value of the β using the same procedures as indicated above. This time, however, we have more models per parameter estimate and an extra parameter (coefficient on the second temporal lag) to examine. For the estimates on the constant and the coefficient on the first temporal lag, the variation in the value of parameters again is larger than any variation caused by error. Because this is not fundamentally different from previous equations, the numerical estimates are not presented.

Table 3.9: Estimates of the true variance of β_2, variance of estimates of β_2, and standard error of estimated values of β_2 for AR(2) models

	Calculated variance of true value	Variance of estimate	(1/n)*Σ(SE^2)
Second Lag	(Std. Dev. of true value)	(Std. Dev. Of Estimate)	√[(1/n)*Σ(SE^2)]
Equation 3.6	0.0194	0.0292	0.0097
(Std Dev)	(0.1394)	(0.1709)	(0.0987)
Equation 3.7	0.0159	0.0255	0.0096
(Std Dev)	(0.1262)	(0.1598)	(0.0980)
Equation 3.8	0.0021	0.0024	0.0002
(Std Dev)	(0.0469)	(0.0495)	(0.0159)
Equation 3.11	0.0046	0.0072	0.0026
(Std Dev)	(0.0681)	(0.0852)	(0.0513)

In all cases listed on Table 3.9, the variance due to differences in the underlying parameter is larger than that variance that is due to sampling error. This coincides with the view that allowing parameter estimates to vary across counties is appropriate.

Once again, we can use our restricted F-tests to determine if the increased explanatory power when we allow coefficients to differ across counties is worth the lost degrees of freedom taken by the increased number of coefficients. In this case, however, we have several more comparisons available. Each of the comparisons considers restrictions on a single parameter. For example, we compare Equation 3.13 (where all three parameter estimates vary) to Equation 3.10 (where only the intercept is held constant and both coefficients are allowed to vary) as there is only one parameter being restricted.

Each of these comparisons shows gains from allowing coefficients to vary across counties. In each case we show statistically significant gains by letting each county have its own coefficient. This is further evidence that a separate model for each county (where the coefficients would naturally vary) is more appropriate.

Within the models estimated in this section to this point, we can see that a common theme is building. In all cases examined to this point, it was shown to be statistically more appropriate to estimate models allowing coefficients and intercepts to vary across counties. This is shown not only by the restricted F-tests performed on the models estimated in this chapter but also the analysis of the variance of the estimated values. Both of these measures constructed from models estimated to this point have suggested that there is sufficient variation between the counties in coefficient estimates to justify estimating separate parameters for each county.

3.3 Incorporating Seasonality

Due to the characteristics of unemployment, we know that monthly unemployment measures are subject to seasonality. Unemployment is generally lower in December due to the holiday retail employment and higher in May/June due to the large number of college graduates entering the labor force. We also generally see higher unemployment in the winter due to the weather, which causes many construction and other outdoor workers to be out of work. The next section of analysis will incorporate a seasonal component into the models that were estimated in previous pages.

We will begin by reexamining the AR(1) models. This time however we will insert eleven monthly dummies to capture any seasonality that may be affecting the results. Because we will be working with the same equations, their functional form will not be repeated here. We will be using the same equation labels/titles; we will simply add an "S" to the title indicating it is the model that captures seasonal effects. Each equation merely adds $\sum_{j=1}^{11} \theta_j D_j$, where D_j will be either 0 or 1 depending on what month is being analyzed at the time, and the θ_j is the coefficient for each dummy. The first step

here is to estimate the parameters of the AR(1) models with the seasonal dummies included in the model. Table 3.10 shows the results of this estimation.

Table 3.10: Summary of seasonal AR(1) models with various restrictions on coefficients and intercepts for county-level unemployment from January 1990-November 2006

	Constant	Coefficient	r-squared	Adjusted r-squared	n
Equation 3.1S	0.9511	0.9130	0.8442	0.8442	23,229
Equation 3.2S	1.2605	*0.8401*	0.8492	0.8484	23,229
Equation 3.3S	*-0.3556*	0.8482	0.8497	0.8489	23,229
Equation 3.4S	*-0.0512*	*0.8709*	0.8541	0.8526	23,229
Equation 3.5S	-2.3188	N/A	0.4562	0.4531	23,345

Entries in italics are mean values for coefficients that vary across counties.

All of the columns represent the same values as before. Looking at the average value of the coefficients, we do not see a large difference in any of the coefficient values between the seasonal and non-seasonal models (compare with table 3.1). The seasonal models have slightly larger coefficients which are different for each county, but these differences are no larger than 0.04. The r-squared has also improved from adding the seasonal dummies, but the improvement is small.

Table 3.11: Coefficients on the seasonal dummies for various seasonal AR(1) models.

	3.13S	3.6S	Dummies Only
Jan	0.6798	0.6296	1.1528
Feb	-0.3708	-0.4911	1.0517
Mar	-0.7128	-0.8085	0.6074
Apr	-1.2219	-1.2878	-0.2835
May	-0.6198	-0.5691	-0.3936
Jun	0.1021	0.1150	0.2085
Jul	-0.4029	-0.4716	0.1893
Aug	-0.8517	-0.8151	-0.1864
Sep	-1.0239	-0.9077	-0.6312
Oct	-0.8471	-0.7202	-0.8520
Nov	-0.2648	-0.1765	-0.5049
Dec	-	-	-

Table 3.11 reports the coefficients on the seasonal dummies in the most restrictive (3.6S) and the least restrictive (3.13S) of the models estimated. We can see that there is not a large amount of difference in these estimates for the two extremes. It is worth keeping in mind that these seasonal coefficients are derived from models that include temporal lags. For example, we can infer that February's unemployment is lower than January's because of the negative coefficient. We cannot however conclude anything about the comparison between February and December because the model's prediction for February includes information from January, due to the inclusion of the temporal lag. The far right column of Table 3.11, which is where the dummies are the only predictor in the model, provides a direct comparison between average unemployment by month.

We continue the analysis of the seasonal AR(1) models presenting estimates for separate groups of counties. Tables 3.12 and 3.13 report these results.

Table 3.12: Average value of the constant term of seasonal AR(1) models by county characteristic, January 1990-November 2006

	Top 15 (LF)	Ag	Manuf (H)	Manuf (L)	Educ	Metro
Equation 3.3S	-0.5440	-0.3188	-0.3643	-0.3832	-0.5957	-0.4809
Std Error	(0.0260)	(0.0177)	(0.0193)	(0.0251)	(0.0260)	(0.0172)
Equation 3.4S	-0.5545	0.0506	-0.0721	0.1465	-0.4523	-0.2371
Std Error	(0.0924)	(0.0594)	(0.0625)	(0.0797)	(0.0856)	(0.0590)
n(counties) =	15	32	27	16	15	34

Table 3.13: Average value of the coefficient term of seasonal AR(1) models by county characteristic, January 1990-November 2006

	Top 15 (LF)	Ag	Manuf (H)	Manuf (L)	Educ	Metro
Equation 3.2S	0.8092	0.8520	0.8451	0.8274	0.7996	0.8229
Std Error	(0.0045)	(0.0022)	(0.0026)	(0.0036)	(0.0047)	(0.0027)
Equation 3.4S	0.9460	0.8573	0.8755	0.8442	0.9167	0.8866
Std Error	(0.0165)	(0.0080)	(0.0087)	(0.0121)	(0.0154)	(0.0097)
n(counties) =	15	32	27	16	15	34

The relationship between the average values of the coefficients by groups is similar to those reported in the models without seasonal dummies. The seasonal model produces slightly larger coefficients; however these differences are small as well.

The next step is to determine if we are gaining anything by allowing coefficients to vary. We will do this in the same way as with both sets of non-seasonal models. We will estimate the variation present in the true value of the parameter as well as conduct restricted F-tests to determine if varying coefficients is more appropriate. We start by calculating the variance of the true values.

Table 3.14: Estimates of the true variance of parameters, variance of estimates of the parameters, and standard error of estimated values of parameters

	Variance of true value	Variance of estimate	Sampling Error
	(Std. Dev. of true value)	(Std. Dev. of estimate)	(Std. Dev. Sampling Error)
Equation 3.2S	0.0015	0.0017	0.0002
coefficients	(0.0387)	(0.0412)	(0.0142)
Equation 3.3S	0.0879	0.0979	0.0101
Constants	(0.2965)	(0.3131)	(0.1002)
Equation 3.4S	0.1392	0.2491	0.1101
Constants	(0.3729)	(0.4991)	(0.3318)
Equation 3.4S	0.0031	0.0054	0.0023
Coefficients	(0.0557)	(0.0741)	(0.0487)

We can see that as in the previously reported models a significant amount of variation is present in the underlying parameters. This supports using separate intercepts and coefficients for each county.

We also performed restricted F-tests, following the same procedure as above. Each of our comparisons indicate that allowing coefficients and intercepts to vary is giving us more value in explanatory power than it costs in degrees of freedom.

To summarize our seasonal AR(1) results, we see that adding the seasonal component does not change our fundamental results. We still see evidence that it is valuable to have coefficients and intercepts vary across counties as opposed to restricting them to be the same for all 115 counties. We see only slight differences in the parameter estimates.

We now consider adding a seasonal component to the AR(2) models. We will also again be using the same equations as listed in the non-seasonal section of the text except that we will add monthly dummies to the models to capture the seasonality.

Table 3.15: Summary of seasonal AR(2) models with various restrictions on coefficients and intercepts for county-level unemployment from January 1990-2006

	Constant	First Lag Coefficient	Second Lag Coefficient	r-squared	Adjusted r-squared	n	Seasonal
Equation 3.6S	0.9355	0.9382	-0.0238	0.8543	0.8542	23,113	-0.5002
Equation 3.7S	*-0.3426*	0.9041	-0.0541	0.8591	0.8583	23,113	-0.4514
Equation 3.8S	1.2199	*0.8997*	-0.0546	0.8590	0.8582	23,113	-0.4469
Equation 3.9S	1.1050	0.9072	*-0.0355*	0.8610	0.8603	23,113	-0.4676
Equation 3.10S	1.2033	*0.7431*	*0.1125*	0.8714	0.8700	23,113	-0.4900
Equation 3.11S	*-0.0079*	*0.9580*	-0.0865	0.8643	0.8628	23,113	-0.4419
Equation 3.12S	*0.1907*	0.8619	*0.0555*	0.8741	0.8728	23,113	-0.4967
Equation 3.13S	*0.0970*	*0.7630*	*0.1368*	0.8804	0.8785	23,113	-0.5030

Entries in italics are mean values for coefficients that vary across counties.

The coefficients on the first lag are different from the non-seasonal models, but the changes from the non-seasonal models are slight. We are also seeing a similar range of coefficients on the seasonal dummies in these cases as compared to those produced by the seasonal AR(1) models. The second lag coefficients are, however, affected by the addition of seasonal dummies much more than the coefficients on the first lag. A few of them now have become positive. To further explore these effects, we have determined the number of coefficients that are statistically significant at various alpha levels (as done

in the non-seasonal AR(2) cases) to see if these are affected. The numbers in parentheses are the number of those significant coefficients that are negative.

Table 3.16: Number of statistically significant coefficients on the second temporal lag in seasonal AR(2) models when the coefficient varies across counties for selected alpha levels

	0.1	0.05	0.01
Equation 3.9S	80(75)	74(70)	54(52)
Equation 3.10S	60(2)	51(1)	37(1)
Equation 3.12S	76(12)	62(11)	28(6)
Equation 3.13S	81(9)	66(7)	42(7)

The results of this tell an interesting story. When season is controlled, we see in general more statistically significant coefficients on the second lag, suggesting that the seasonality in the data was making some coefficients look insignificant. However, we do not see the second lag capturing the "leftover" variation in Equation 3.9S to the same degree as in Equation 3.9. This can be seen from the fact that in Table 3.7 far more second lag coefficients were significant; in fact, all 115 were significant at all three listed levels of alpha in Equation 3.9. These two models have fixed coefficients on the first temporal lag of unemployment. The conclusion here is that the seasonal coefficients are capturing some of the county-specific variation. That being said, the seasonal model still implies that the coefficients on the second temporal lag (allowed to vary across counties) is picking up some variation that is being missed by the coefficient on the first temporal lag (constrained to be constant across counties).

We next looked at the values of the varying coefficients and intercept when the counties are broken up into groups, using the same groups as above (results not presented). The coefficients on the first temporal lag are extremely close in value to those calculated for the non-seasonal AR(2) models. It would appear that they were

relatively unaffected by any seasonality in the data. The second lag coefficients, however, are affected in accord with results presented above.

We now turn to our analysis of whether or not the estimated coefficients and intercepts for each county are statistically justified. Recall in the non-seasonal AR(2) models the results indicated that allowing parameters to vary was appropriate. Now that we have included seasonal dummies, we will check to ensure that this conclusion is unaffected. We first calculated estimates of the variances of the true values of the parameters. As with previous scenarios, we found that substantial variation is present across the true values of the parameters. This again indicates that allowing coefficients to vary across counties is appropriate, implying that the inclusion of seasonal dummies did not affect our results.

We conducted the same pairwise comparisons based on F-tests as previously tested in the non-seasonal AR(2) case. As in previous restricted F-tests, the data indicate that it is more appropriate to allow coefficients to vary across counties.

To quickly summarize our findings so far, we have learned that a first temporal lag is highly appropriate. Coefficient estimates for the first temporal lag are also fairly robust to whether or not seasonal dummies or further lags are included. This indicates that including a first lag will likely be a component of every county's model. When considering second lags, however, the evidence is less clear. The second lag's coefficient is influenced by seasonal dummies and whether or not the seasonal dummies are included, not all of these coefficients are found to be significant.

The results also show that estimating separate parameters for each county appears to be the preferred structure. We see this in all four sets of models that were estimated.

This is indicated not only by the results of the F-tests but also the analysis of the variance of parameters across counties.

Another consideration in building the forecasting models was to determine the appropriate number of temporal lags that would be required in the county models. The two most popular methods of model selection are the Akiake Information Criteron (AIC) and the Bayesian Information Criteron (BIC). These measures are as follows:

$$AIC = 2k + n\left\{\ln\left(\frac{RSS}{n}\right)\right\}$$

$$BIC = \ln(\sigma_e^2) + \frac{k}{n}\ln(n)$$

In these equations, n is the sample size, k is the number of predictors, RSS is the residual sum of squares, and σ_e^2 is the variance of the error. In both cases, the model that produces the lowest value for the measure is taken to be the most appropriate model. The following table shows the number of lags implied by these tests in the seasonal models run by counties.

Table 3.17: Number of county models where the given number of lags were specified by AIC and BIC

Lags	AIC	BIC
1	23	24
2	43	46
3	21	22
4	16	15
5	12	8

Two things immediately are apparent from the table. First, we can see the results for the two measures provide similar results. In fact, for only ten counties do the AIC and BIC produce different numbers of lags. Of those ten, five of them show a difference of one lag, with the other five having a difference of two lags (four times) or three lags (once). This would indicate that the results are fairly robust with respect to the choice of model

selection criterion. Second, we see that the criteria specify one to three temporal lags in four-fifths of the counties. The actual calculated values for both measures for these three lags are all very close together. We also found that the number of lags implied by these two measures does not differ by the county size.

Based on these model selection criteria, additional models were estimated using the full sample. Using BIC as the guide (since it appears to choose a simpler model - fewer lags than the AIC), we have grouped counties together by the number of temporal lags suggested by the BIC. Within these groups, we restrict the coefficients to be constant across the counties included in the model. The goal is to see what the general pattern of these lag coefficients is. The results are in the Table 3.7.

Table 3.18: Regression results of various autoregressive models of order p (p = 1 to 4) using all observations

	AR(1)	AR(2)	AR(3)	AR(4)
Intercept	0.6307	1.0544	0.8863	0.8866
First Lag	0.9439	0.9476	0.8783	0.9795
Second Lag		-0.0469	0.0477	-0.0040
Third Lag			0.0089	-0.1723
Fourth Lag				0.1124
r^2	0.8967	0.8379	0.8763	0.8518
Counties	24	46	23	22

These regressions have produced some very interesting results. First, the coefficient on the first temporal lag appears to be quite similar across models with different numbers of temporal lags. All coefficients except for the third lag in AR(3) and the second lag in AR(4) are statistically significant. It is also interesting that the coefficient on the third lag in the AR(3) model is quite small. This is a curious result

because this was estimated using only those counties suggested to have three temporal lags by BIC.

3.4 Spatial Models

At this point, we move on to investigating spatial models. While it is possible that we will wish to include more than two temporal lags in our spatial models in the future, we focus now on some diagnostics and work with either one or two temporal lags. Because we have found that allowing coefficients to vary across counties is a more appropriate setup for our models, we will use this principle as we progress through this section. We will also include estimates for models that are the most restrictive, i.e., we will estimate parameters for models that restrict all coefficients to be equal across counties. This provides an overall summary of the relationships in the data. The major difference between these models and the AR models previously estimated is the inclusion of a spatial lag.

The succeeding models (Equations 3.14 to 3.25) are ordered by increasing complexity. The first (3.14) has only a constant, a single temporal lag, and a spatial lag. All three coefficients are restricted to be constant across counties. Next, a second temporal lag is introduced. This also has a coefficient that is restricted across counties. We then move to allowing coefficients to vary across counties in models first without and then with the second temporal lag. Equations 3.20 to 3.25 identify models that are similar to the prior one except that they introduce a second spatial lag. In initial models, the second spatial lag is restricted to be constant across counties, while in later models the second spatial lag coefficient is allowed to vary across counties.

The spatial lags used in these models are estimated in the following way. For all first-order spatial neighbors (those counties that are adjacent to county i), a single variable has been constructed. This was done by aggregating the total number of unemployed persons in these neighboring counties and dividing by the total labor force. This gives us a single measure of unemployment for the neighboring areas taken together, which is denoted by X_i^{FO}. This procedure was also used to calculate an unemployment rate for all second-order spatial neighbors (counties adjacent to first-order neighbors but not adjacent to county i) which is denoted by X_i^{SO}. The equations estimated in this section are listed below.

Equation 3.14 $$Y_{it} = \alpha + \beta_1 Y_{i,t-1} + \pi X_{i,t-1}^{FO} + \varepsilon_{it}$$

Equation 3.15 $$Y_{it} = \alpha + \beta_1 Y_{i,t-1} + \beta_2 Y_{i,t-2} + \pi X_{i,t-1}^{FO} + \varepsilon_{it}$$

Equation 3.16 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-k} + \pi X_{i,t-1}^{FO} + \varepsilon_{it}$$

Equation 3.17 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \pi_i X_{i,t-1}^{FO} + \varepsilon_{it}$$

Equation 3.18 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \beta_{i2} Y_{i,t-2} + \pi X_{i,t-1}^{FO} + \varepsilon_{it}$$

Equation 3.19 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \beta_{i2} Y_{i,t-2} + \pi_i X_{i,t-1}^{FO} + \varepsilon_{it}$$

Equation 3.20 $$Y_{it} = \alpha + \beta_1 Y_{i,t-1} + \pi X_{i,t-1}^{FO} + \omega X_{i,t-1}^{SO} + \varepsilon_{it}$$

Equation 3.21 $$Y_{it} = \alpha + \beta_1 Y_{i,t-1} + \beta_2 Y_{i,t-2} + \pi X_{i,t-1}^{FO} + \omega X_{i,t-1}^{SO} + \varepsilon_{it}$$

Equation 3.22 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \pi_i X_{i,t-1}^{FO} + \omega X_{i,t-1}^{SO} + \varepsilon_{it}$$

Equation 3.23 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \pi_i X_{i,t-1}^{FO} + \omega_i X_{i,t-1}^{SO} + \varepsilon_{it}$$

Equation 3.24 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \beta_{i2} Y_{i,t-2} + \pi_i X_{i,t-1}^{FO} + \omega X_{i,t-1}^{SO} + \varepsilon_{it}$$

Equation 3.25 $$Y_{it} = \alpha_i + \beta_{i1} Y_{i,t-1} + \beta_{i2} Y_{i,t-2} + \pi_i X_{i,t-1}^{FO} + \omega_i X_{i,t-1}^{SO} + \varepsilon_{it}$$

The results of these estimates are listed in the following table. As with previous tables, those values in italics represent average parameter estimates where parameters are not restricted to be constant across counties.

Table 3.19: Summary of estimates of various spatial models

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Second Spatial Lag	r-squared	Adjusted r-squared
3.14	0.4865	0.8816		0.0338		0.8067	0.8067
3.15	0.5153	0.9772	-0.1087	0.0399		0.8189	0.8189
3.16	*0.2432*	*0.7759*		0.0877		0.8192	0.8174
3.17	*0.0100*	*0.7510*		*0.1327*		0.8226	0.8199
3.18	*0.5002*	*0.8308*	*-0.0590*	0.0881		0.8464	0.8441
3.19	*0.3472*	*0.8140*	*-0.0686*	*0.1284*		0.8498	0.8467
3.20	0.5230	0.8810		0.0438	-0.0164	0.8068	0.8068
3.21	0.5512	0.9767	-0.1088	0.0498	-0.0162	0.8190	0.8189
3.22	*0.0259*	*0.7477*		*0.1030*	0.0394	0.8227	0.8200
3.23	*0.0097*	*0.7285*		*0.0801*	*0.0870*	0.8241	0.8205
3.24	*0.3650*	*0.8109*	*-0.0696*	*0.0961*	0.0430	0.8499	0.8468
3.25	*0.2791*	0.7930	*-0.0763*	*0.0730*	*0.0899*	0.8513	0.8476

Entries in italics are mean values for coefficients that vary across counties.

In the models that restrict spatial coefficients to be the same across all counties, all of the spatial parameter estimates are significant for both first-order and second-order neighbors. In all of the cases where the first spatial lag is allowed to vary across counties, we find over 30 coefficients significant at an alpha level of 0.10. That number decreases as we lower the alpha level to 0.01, but never drops below 10 significant coefficients. For the second spatial lag the number of significant coefficients is smaller than for the first spatial lag, which is not a surprise.

At this point we determine if allowing the coefficients to vary across counties is worth the loss of degrees of freedom. As with the non-spatial models estimated earlier, we will utilize the restricted F-test to check this.

Table 3.20: Results of restricted F-tests between pairs of spatial models

Base Model	New Model (Addition)	Adjusted r-squared		Significance F
		Base	New	
3.14	3.15 (2nd temporal lag)	0.8067	0.8189	<0.0001
3.16	3.17 (varying π)	0.8174	0.8199	<0.0001
3.18	3.19 (varying π)	0.8441	0.8467	<0.0001
3.20	3.21 (2nd temporal lag)	0.8068	0.8189	<0.0001
3.22	3.23 (varying ω)	0.8200	0.8205	0.0001
3.24	3.25 (varying ω)	0.8468	0.8476	<0.0001

Each of these comparisons show a highly statistically significant difference between the two models compared. These results indicate that we gain statistical power by allowing coefficients to vary.

We also considered varying numbers of lags for spatial models as was done with non-spatial models using AIC and BIC as our model selection criteria. We examined all counties with one through five temporal lags and a single first-order spatial lag that was lagged temporally one period. While our results for these models are not identical to the non-spatial models (Table 3.17), they are very similar. The numbers in each of the columns are only altered slightly.

3.5 Spatial Models with County Characteristic Predictors

One extension of the spatial models that we now explore is the inclusion of a predictor that measures the unemployment rate of other similar counties identified by selected characteristics. It is constructed in a similar way to the spatial variable, but the criterion for including a county is based on the concentration of manufacturing, education, or agriculture in a given county. We also separately fit models for those counties that are classified as metropolitan counties and use all other metropolitan counties to construct this extra predictor. They are based on models that have been run

previously so we will continue to use the same equation titles, but each of these includes an extra predictor.

Table 3.21: Summary of estimates of spatial models including a predictor measuring unemployment in all other agricultural counties (32 counties)

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Ag	r-squared
Equation 3.14	0.4162	0.9137		-0.0175	0.0192	0.8314
Equation 3.15	0.3889	0.8445	0.0780	-0.0081	0.0048	0.8411
Equation 3.17	*-6.2773*	*0.6403*		*0.0648*	*0.1756*	0.6691
Equation 3.19	*-6.2507*	*0.6632*	*-0.0287*	*0.0811*	*0.1602*	0.6761

Entries in italics are mean values for coefficients that vary across counties.

Table 3.22: Summary of estimates of spatial models including a predictor measuring unemployment in all other high manufacturing counties (27 counties)

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Manuf	r-squared
Equation 3.14	0.5099	0.9025		0.0370	-0.0267	0.8345
Equation 3.15	0.4751	0.8922	0.0126	0.0376	-0.0253	0.8466
Equation 3.17	*0.3023*	*0.7242*		*-0.0429*	*0.2185*	0.8501
Equation 3.19	*0.4183*	*0.7890*	*-0.0643*	*-0.0284*	*0.1974*	0.8638

Entries in italics are mean values for coefficients that vary across counties.

Table 3.23: Summary of estimates of spatial models including a predictor measuring unemployment in all other educational counties (15 counties)

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Educ	r-squared
Equation 3.14	0.3486	0.8915		-0.0091	0.0356	0.8228
Equation 3.15	0.3218	0.8891	0.0130	-0.0064	0.0273	0.8323
Equation 3.17	*0.4576*	*0.7241*		*0.0104*	*0.1782*	0.8410
Equation 3.19	*0.4781*	*0.7694*	*-0.0407*	*-0.0011*	*0.1765*	0.8495

Entries in italics are mean values for coefficients that vary across counties.

Table 3.24: Summary of estimates of spatial models including a predictor measuring unemployment in all other metropolitan counties (34 counties)

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Metro	r-squared
Equation 3.14	0.4163	0.9137		-0.0175	0.0192	0.8314
Equation 3.15	0.3889	0.8446	0.0780	-0.0082	0.0049	0.8411
Equation 3.17	*0.1314*	*0.6698*		*0.0229*	*0.2245*	0.8515
Equation 3.19	*0.1578*	*0.6844*	*-0.0177*	*0.0410*	*0.2028*	0.8610

Entries in italics are mean values for coefficients that vary across counties.

The coefficients on all predictors are quite small when they are constrained to be the same across counties in all cases. The only statistically significant coefficient constrained to be constant is the educational variable in Equation 3.14 ($p = 0.01$). When it is allowed to vary across counties, the average coefficient value is greater than the single constrained estimate. The high manufacturing counties have the most coefficients that are significant; Equation 3.17 has 11 (40.7%) significant coefficients and Equation 3.18 has 12 (44.4%) significant. Educational counties have the next highest incidence with 6 (40%) on Equation 3.17 and 4 (26.6%) on Equation 3.18. The metropolitan predictor has 9 (26.4%) and 7 (21.8%), respectively. Both of the agricultural models have fewer than 3% of their predictors significant when coefficients are allowed to vary across counties. This would indicate that there is some county-level variation that is being captured by the additional variable but only in certain cases.

We also conducted tests to determine if the addition of predictors was worth the cost of the degrees of freedom. For all groups' 3.14 and 3.15, the addition of the group predictor was tested and in each case the results showed that adding a group predictor was appropriate. We also tested the assumption of the extra group-specific predictor

should be held constant across counties or be allowed to vary across counties. Once again, the values that resulted from this test showed statistical significance, implying that allowing these extra predictors' coefficients to vary is the appropriate setup.

3.6 Spatial Models Accounting for Out-of-Missouri Counties

Another similar extension that we can perform focuses on the counties that border other states and measure how much impact out-of-Missouri counties are having on Missouri counties. We have used two methods to break up the spatial variable into Missouri and non-Missouri parts. One way is to simply construct them analogously; we have simply summed all unemployed persons in the contiguous Missouri (non-Missouri) counties for county i and divided that by the labor force in those counties giving us the unemployment rate.

We have also considered that the relative importance of counties may vary by population, and therefore should be weighted differently. For example, effects in the highly populous St. Louis County should have a much more significant impact on neighboring St. Charles County than would an effect coming from Madison County (IL), which is much smaller in population than St. Louis County. To capture this effect, we have weighted the Missouri counties as a percentage of the labor force size of the Missouri and non-Missouri counties for county i and done the same weighting for the non-Missouri counties[10]. In both sets of estimates that follow, only counties that contain out-of-Missouri counties as a predictor are considered (47 counties).

[10] The weight for the Missouri counties will be $^{LFMO}/_{LF}$, where LFMO = labor force in Missouri counties contained in spatial variable i and LF = labor force in spatial variable i as a whole. Non-Missouri counties get a weight of $^{LFNON}/_{LF}$, where LFNON = labor force in the non-Missouri counties in spatial variable i.

Table 3.25: Summary of estimates of spatial models separating Missouri and non-Missouri neighbors

	Constant	First Temporal Lag	Second Temporal Lag	MO neighbors	Non-MO neighbors	Adjusted r-squared
Equation 3.14	0.5873	0.8483		0.0201	0.0222	0.7498
Equation 3.15	0.7193	1.0447	-0.2351	0.0304	0.0239	0.7744
Equation 3.17	*-0.3824*	*0.7521*		*0.0813*	*0.0718*	0.7742
Equation 3.19	*-0.6797*	*0.8225*	*-0.0849*	*0.0913*	*0.0506*	0.8138

Entries in italics are mean values for coefficients that vary across counties.

Table 3.26: Summary of estimates of spatial models separating and weighting Missouri and non-Missouri neighbors

	Constant	First Temporal Lag	Second Temporal Lag	MO neighbors	Non-MO neighbors	Adjusted r-squared
Equation 3.14	0.6325	0.8521		0.0208	0.0417	0.7496
Equation 3.15	0.7693	1.0495	-0.2350	0.0293	0.0513	0.7740
Equation 3.17	*-0.2008*	*0.7985*		*0.0525*	*0.2166*	0.7664
Equation 3.19	*-0.5598*	*0.8607*	*-0.0851*	*0.0776*	*0.0495*	0.8083

Entries in italics are mean values for coefficients that vary across counties.

As with the industrial characteristic models, we observe a much stronger effect from the coefficients that are allowed to vary across counties than those that are restricted to be constant across counties. We also see similar results from the two different weighting structures imposed on counties. The exception is a large increase in the Non-Missouri neighbors' coefficient in Equation 3.17. However, when the models estimated in Table 3.25 and 3.26 are compared to the models that leave the Missouri and Non-Missouri counties together in the same variable considering only the counties in Missouri that border other states, we see the original setup produces a higher adjusted r-squared. The original specifications in 3.14, 3.15, 3.17, and 3.19 (Table 3.19) all have adjusted r-squared values of greater than that of their counterparts which separate Missouri and Non-Missouri counties. When checking the statistical significance of the coefficients, the

MO neighbors and Non-MO neighbors have significant coefficients when these are held constant across counties. When they are allowed to vary, very few in either group (MO or Non-MO) in either specification where they are separated (weighted or non-weighted) are statistically significant, which is likely due to collinearity.

3.7 Seasonal Spatial Models

As with the previous set of models, we next incorporate seasonality into our estimates. We utilize the same spatial equations as before, except we now include eleven monthly dummies to capture the seasonality in the data. The results of these seasonal spatial models are listed in the following table; to denote the presence of monthly dummies we include an "S" in the title of the equation.

Table 3.27: Summary of estimates of various seasonal spatial models

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Second Spatial Lag	r-squared	Adjusted r-squared	Seasonal Coeff
3.14S	*0.8009*	*0.8849*		*0.0610*		0.8452	0.8451	-0.5114
3.15S	*0.7690*	*0.9134*	*-0.0309*	*0.0678*		0.8555	0.8554	-0.5143
3.17S	-0.2517	0.7632		0.1798		0.8586	0.8565	-0.4818
3.19S	0.0256	0.6916	0.1075	0.1602		0.8841	0.8817	-0.5064
3.20S	*0.7678*	*0.8856*		*0.0522*	*0.0150*	0.8452	0.8451	-0.5145
3.21S	*0.7235*	*0.9148*	*-0.0314*	*0.0558*	*0.0206*	0.8555	0.8554	-0.5184
3.23S	-0.1697	0.7361		0.1037	0.1300	0.8603	0.8574	-0.4905
3.25S	0.0739	0.6719	0.0966	0.0874	0.1229	0.8856	0.8826	-0.5100

The changes we observe here between seasonal and non-seasonal models are consistent across alternative models, and correspond with results for the non-temporal model. The first temporal lag is very similar for the two types (seasonal and non-seasonal) of models, and the second temporal lag is more likely to differ. We are more interested here in the behavior of the spatial coefficients. The first spatial lag coefficients are larger in the seasonal models, but this increase is small. The second spatial lag is also slightly larger

in the seasonal models. As for the statistical significance of the parameter estimates, we also see a modest increase in the number of coefficients individually that are significant. This indicates that a county's neighbors are more relevant in explaining unemployment if season is taken into account. Since the neighbor unemployment rates are lagged, this result suggests that the impact is to some degree masked in the non-seasonal models by the difference in unemployment across months.

What we have seen in our spatial models to this point indicates two things that are worth mentioning. First, we observe in the spatial models that allowing varying coefficients across counties is statistically justified, a result consistent with our findings for the non-spatial models. This would lead us to believe that estimating our forecasting models in the future will call for separate models for each county. We also can see that while seasonality is present in the data, the effects of the seasonality are rather small. While our models should include a seasonal component, we expect they will be fairly robust to any excessive seasonality in the data.

3.8 Models Including Larger Region Unemployment

As we are including spatial neighbors as predictors of county unemployment, it is important to determine if the spatial correlation that is observed is a result of a relationship between the counties or the result of other factors affecting all counties in a similar way. If there are other factors affecting all counties, a measure of unemployment for a larger region that includes all counties should capture this. We consider the lagged unemployment rates for the United States, for the Midwest, and for Missouri and we constrain their coefficients to be constant across counties giving us a single impact estimate for each model. We add each of them as predictors to our spatial models to

determine what effects are present. Because we are using variations of previously stated equations, we will not be restating them. Instead we will add a label to the equation titles indicating which larger region is being used. The national unemployment rate will have a label of "US", the Midwest unemployment rate will have "MW" and the Missouri unemployment rate will have "M" added to the equation. The results of these estimates are in the following tables.

Table 3.28: Summary of estimates of various spatial models including the U.S. unemployment rate

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Second Spatial Lag	r-squared	Adjusted r-squared	National Coeff
3.14US	0.1218	0.8791		0.0119		0.8076	0.8076	0.0900
3.15US	0.0379	0.9812	-0.1170	0.0114		0.8204	0.8204	0.1186
3.17US	*0.0463*	*0.7339*		*0.0550*		0.8250	0.8224	0.1855
3.19US	*0.4079*	*0.8223*	*-0.1072*	*0.0345*		0.8534	0.8504	0.2318
3.20US	0.1321	0.8767		0.0343	-0.0455	0.8080	0.8079	0.1124
3.21US	0.0497	0.9803	-0.1194	0.0379	-0.0537	0.8209	0.8209	0.1454
3.23US	*-0.0231*	*0.7176*		*0.0560*	*0.0062*	0.8264	0.8229	0.2016
3.25US	*0.2452*	*0.8082*	*-0.1145*	*0.0438*	*-0.0098*	0.8549	0.8512	0.2564

Table 3.29: Summary of estimates of various spatial models including the Midwest unemployment rate

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Second Spatial Lag	r-squared	Adjusted r-squared	Midwest Coeff
3.14MW	0.1260	0.8802		0.0084		0.8078	0.8077	0.0993
3.15MW	0.1181	0.9806	-0.1144	0.0120		0.8202	0.8201	0.1102
3.17MW	*0.2299*	*0.7292*		*0.0223*		0.8258	0.8232	0.2257
3.19MW	*0.6189*	*0.8208*	*-0.1096*	*0.0068*		0.8536	0.8507	0.2547
3.20MW	0.1419	0.8779		0.0323	-0.0499	0.8082	0.8082	0.1255
3.21MW	0.1354	0.9798	-0.1164	0.0375	-0.0532	0.8207	0.8206	0.1383
3.23MW	*0.0841*	*0.7093*		*0.0413*	*-0.0365*	0.8273	0.8238	0.2583
3.25MW	*0.3880*	*0.8029*	*-0.1192*	*0.0307*	*-0.0510*	0.8555	0.8518	0.2988

Table 3.30: Summary of estimates of various spatial models including the Missouri unemployment rate

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Second Spatial Lag	r-squared	Adjusted r-squared	State Coeff
3.14M	0.0983	0.8792		0.0029		0.8080	0.8080	0.1136
3.15M	0.0946	0.9804	-0.1155	0.0064		0.8204	0.8204	0.1242
3.17M	*0.2422*	*0.7136*		*-0.0208*		0.8273	0.8247	0.3038
3.19M	*0.6331*	*0.8156*	*-0.1261*	*-0.0408*		0.8555	0.8525	0.3438
3.20M	0.1103	0.8761		0.0291	-0.0581	0.8086	0.8085	0.1479
3.21M	0.1077	0.9794	-0.1181	0.0342	-0.0616	0.8211	0.821	0.1608
3.23M	*0.0530*	*0.6921*		*0.0189*	*-0.0895*	0.8292	0.8257	0.3697
3.25M	*0.3526*	*0.7990*	*-0.1409*	*0.0058*	*-0.1108*	0.8578	0.8542	0.4283

With respect to the first temporal lags, there does not appear to be a large difference between the models (National, Midwest, Missouri). Nor do the second temporal lag coefficients vary by much. They are smaller when the larger region's lagged unemployment is included, but the difference does not appear to be very big. The most volatile coefficient estimates are those on the first spatial lag. The first spatial lag coefficient tends to decrease when adding a larger region unemployment rate; some even become negative. This would indicate that the inclusion of a larger region's

unemployment rate is capturing the spatial correlation present in the data. Based on this, we could conclude that the spatial correlation in Missouri county data is in large part due to greater shocks affecting all counties equally and not due to county-level spatial correlation. To check this property, we need to count how many spatial coefficients are significant before and after adding a larger region's unemployment. Tables 3.31 to 3.33 show the number of significant coefficients on the first spatial lags in four model specifications. These specifications all allow the first spatial lag coefficients to vary across counties and we check the effects of including the three larger measures of lagged unemployment as well as omitting any measure.

Table 3.31: Number of significant coefficients on first spatial lags in various models, $\alpha = 0.10$

	Base	National	Midwest	Missouri
3.17	54	33	34	41
3.19	54	38	42	45
3.23	27	28	25	30
3.25	27	34	34	37

Table 3.32: Number of significant coefficients on first spatial lags in various models, $\alpha = 0.05$

	Base	National	Midwest	Missouri
3.17	45	24	27	36
3.19	45	29	32	37
3.23	21	19	18	23
3.25	21	22	22	29

Table 3.33: Number of significant coefficients on first spatial lags in various models, $\alpha = 0.01$

	Base	National	Midwest	Missouri
3.17	27	13	14	17
3.19	27	19	18	24
3.23	18	7	9	9
3.25	10	7	11	14

The results show that the effects of including a larger region's lagged unemployment appear much stronger when allowing coefficients to vary across counties. Also, including Missouri's state unemployment rate actually decreases the number of significant coefficients on first spatial lags. One other property of note is the behavior of the coefficient on the larger region's unemployment rate. In all cases, this coefficient is larger when coefficients on the temporal and spatial lags are allowed to vary. These differences are quite large as it results in an increase of three times the magnitude.

We can also statistically test the extra variable for the larger region's unemployment rate and determine if the inclusion of the larger region unemployment was worth the degrees of freedom. Because of the consistency of the previous results, we will test the various forms of Equation 2.25. The results indicate that including the larger region unemployment is statistically important in all three cases. While it is tempting to conclude that Missouri's might be the best of the three, all F-statistics are so large that any of the three is effective at capturing the variation that is being picked up by the larger region's inclusion. Despite the lack of a clear "winner" with respect to which larger

region to include, we can conclude that a larger region may be useful to include as a predictor. We now conclude our estimation in this chapter by showing models that combine the seasonal component with the larger region's unemployment. We have chosen to use Missouri as the larger region because of the slightly larger r-squared and adjusted r-squared values. Even though the difference is small, they are the result of the same amount of degrees of freedom lost as the other two larger regions' unemployment (national, Midwest). Table 3.34 reports the parameter estimates of the most restrictive and least restrictive models including both temporal lags, both spatial lags, seasonal dummies, and Missouri's unemployment rate.

Table 3.34: Summary of estimates of various seasonal spatial models including the Missouri unemployment rate

	Constant	First Temporal Lag	Second Temporal Lag	First Spatial Lag	Second Spatial Lag	Seasonal	State Coeff	r-squared	Adjusted r-Squared
3.21S,M	0.5265	0.9134	-0.0332	0.0479	-0.0026	-0.5036	0.0733	0.8559	0.8558
3.25S,M	*0.1095*	*0.6630*	*0.0774*	*0.0477*	*0.0179*	-0.4312	0.2024	0.8866	0.8836

In comparing the two model specifications, we can see that we have gained in both r-squared and adjusted r-squared. Given our previous results, it is likely that we would find the lost degrees of freedom from allowing coefficients to vary would be more than compensated by the increase in explanatory power.

3.9 Summary

In this chapter we have gained insight with respect to building our forecasting models for Missouri counties. Perhaps the most important theme of this chapter is the effect of the inclusion of spatial variables in models for county-level unemployment. To address this question, models using the full sample with spatial variables were estimated and those results were compared to models that omitted spatial variables. This question

has been explored with such variations as the treatment of coefficients (whether they vary across counties or are restricted to be constant across counties), the inclusion of monthly seasonal dummies, and the inclusion of a larger region's unemployment rate (Missouri, Midwest, United States). While some modifications were made with respect to the specifics of each type of model, the conclusion that including a spatial variable adds to predictive power is consistent in all various methods of estimating these models. To this point, it is clear that including a spatial variable as a predictor for county-level unemployment improves the predictive accuracy of the model.

Another question that has been addressed in this chapter is whether models should allow coefficients to vary across counties or if coefficients should be held constant across counties. To answer this question, we have estimated various models (both non-spatial and spatial) that have allowed the coefficients to vary across counties as well as be restricted. By checking the parameter estimates, the amount of variation between county parameter estimates, and the results of a series of restricted F-tests, we have determined based on information from models estimated to this point (all of which used the full sample) that explained variance is greater when counties have their own coefficients for each predictor included in their models. This conclusion was reached for not only the non-spatial AR models but also the spatial models that included neighboring counties as predictors. This conclusion was robust with respect to including seasonal components in both the non-spatial and spatial models.

We have also examined and found that explained variance increases when a larger region's unemployment is included in the spatial models. While it is not clear which of the regional variables is most effective, it is apparent from the results of the statistical

tests that not including one of these would be a mistake. We concluded by illustrating the estimates of the most complex spatial model estimated that may serve as a baseline for the forecasting models to come.

4. FORECASTING ACCURACY

In the previous chapter, we estimated various specifications of models considering differing numbers of temporal and spatial lags. In some models, we held coefficients constant across counties and in others we allowed them to vary across counties. We also fitted separate models for groups of counties identified by characteristics. In all models estimated in the previous chapter, the common thread is that the entire sample of available data was used. In this chapter, we move to forecasting models based on sample sizes that are smaller than in the previous chapter, allowing for the calculation of forecasting accuracy based on out-of-sample forecasts. The arrangement of this chapter differs from the previous chapter because chapter 3 was more exploratory in nature. The purpose of those analyses was to determine which specifications should be focused on when constructing forecasting models. Chapter 4 uses these results from Chapter 3 to narrow the focus to those specifications that appear to be most appropriate.

The main focus of this chapter is the predictive accuracy of forecasting models. The base model estimated in this chapter includes seasonal dummies, although in the final analysis we consider models that omit them. We start by comparing the forecasting accuracy of several non-spatial model specifications using models that are estimated using 60-month samples to predict the unemployment in the subsequent month. To predict the next month's unemployment, we then drop the first month and add the most recent month to the model and re-estimate a new model from these updated 60 observations. This process is repeated for each month moving forward giving us static

forecasts to later measure forecasting accuracy. Models with one and two temporal lags are considered as well as models allowing coefficients to vary across counties and those that restrict coefficients to be constant. To determine the forecasting accuracy of the models, we will use root mean square error (RMSE) and mean absolute error (MAE). The expressions for these two measures are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{actual} - x_{forecast})^2}{n}}$$

$$MAE = \frac{\sum_{i=1}^{n}(|x_{actual} - x_{forecast}|)}{n}$$

These models are compared both in terms of their predictive accuracy for all counties taken together and separately for each county[11]. The results for these non-spatial models end up being quite complex. The one common theme is that models with constant coefficients across counties are generally better predictors than those that allow the coefficients to vary.

Next, spatial models are considered. These models are very similar to the non-spatial models with the exception that a spatial variable is included as a predictor. We make the same comparisons between these specifications as in the non-spatial models, but also compare these models to the non-spatial models already estimated. We find consistent patterns when comparing spatial to non-spatial models. In the models allowing coefficients to vary, non-spatial models are superior to their spatial counterparts. When coefficients are restricted to be constant across counties, the spatial models' accuracy measures are nearly identical to the non-spatial models'. There are however cases where the spatial model does produce better forecasts than its non-spatial

[11] Formal tests such as those developed by Diebold and Mariano (1995), Giacomini and White (2006), and Clark and McCracken (2001) were not performed in this analysis.

counterpart, in particular in the models where coefficients are held constant across counties.

We then consider models with different numbers of months used to estimate the parameters. We consider models that use both 40 months and 100 months to see the effects of both increasing and decreasing the sample size. The results of this comparison between the original 60 month forecasts and the new forecasts based on different sample sizes show a clear trade-off between model complexity and sample size. In the models using a smaller number of months, a larger number of predictors damage the accuracy of the model, so the simplest models tend to predict best. If more months are used to estimate the model, the more complex models' (in terms of number of predictors) predictive accuracy improves significantly relative to the simpler models. We finish our overall model comparisons by allowing a different number of lags for each county (as prescribed by BIC) and see if customized models improve prediction. Our results indicate that such an approach does not help forecasting accuracy.

We then consider separate groups based on county characteristics used in previous chapters. We find first that for the most part the best model is the same for the various groups of counties. In all four of our groups (metropolitan status, manufacturing concentration, agricultural concentration, and educational concentration) the best model is much the same. We also have included a predictor in these groups that matches the characteristic group of interest; for example, in the metropolitan counties a new independent variable constructed from the unemployment rates of only other metropolitan counties was included in the model. We find this group-specific predictor actually damages the accuracy of forecasts for the most part.

We also consider how important it is for a predictor to be a Missouri county. To determine this, we run models and produce forecasts for counties on Missouri's border, splitting the spatial variable into two, one variable including only Missouri counties and the other including only non-Missouri counties. We find that including non-Missouri counties helps accuracy, but separating them from Missouri counties is also helpful for predictive accuracy. Our final set of analyses focuses on the question of including seasonal dummies in our models. We find that leaving them out hurts predictive power dramatically.

4.1 Forecast Accuracy of Non-Spatial Models

Given the superiority of the seasonal models, we limit our consideration to seasonal models below[12]. We will start by looking at non-spatial models that include one and two temporal lags of unemployment. We will also use the lagged unemployment rate for Missouri as a way to incorporate a broader regional measure of unemployment. At this point, we need to introduce some new notation. In the previous chapter, equations 3.1 and 3.4 were similar in that they were both models for county unemployment that used a constant term and a single temporal lag as a predictor.

Equation 3.1 $$Y_{it}=\alpha+ \beta_1 Y_{i,t-1} + \varepsilon_{it},$$

Equation 3.4 $$Y_{it} = \alpha_i+ \beta_{i1} Y_{i,t-1} +\varepsilon_{it}$$

The difference between the two is that 3.1 restricted the coefficients to be the same for all 115 counties while 3.4 allowed them to be different for each county. Because we want to keep together comparable equations that allow coefficients to vary

[12] Our results show that seasonal models consistently outperform non-seasonal models in this comparison. These results are presented in section 4.7. We suppress the seasonal variables in the formal equations listed below.

across counties and those that constrain them to be the same, we now refer to both versions using the notation of Equation 4.1.

Equation 4.1 $Y_t=\alpha+ \beta_1 Y_{t-1}+\varepsilon_t$

What we have done is suppress any subscripts referring to county i. Our models will still be estimated for each county by assuming coefficients are held constant across counties and by allowing coefficients to vary across counties. We will distinguish between the two specifications (constant or varying coefficients) by using "Constant" or "C" to indicate the coefficients are being held constant across counties or by using "Vary" or "V" to indicate the coefficients are allowed to vary across counties. At this time we also estimate these two variations of Equation 4.2.

Equation 4.2 $Y_t=\alpha+ \beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\varepsilon_t$

We also add Missouri's unemployment rate at time t – 1 as a predictor in some cases in these models; the presence of this added variable is indicated by "M" in the equation title.

The left side of Table 4.1 presents the root mean square error. The right side of the table compares the predictive power of the model separately for each county. The "Model" label is a generic label designed to refer only to the models listed in the table below. In Table 4.1, "Vary" ("Constant") models are only compared to "Vary" ("Constant") models. We consider below the impact of allowing coefficients to vary across counties. The count is the number of counties for which the first model produces a stronger forecast than another. For example, in "2 v 1" in the row "Vary," the value of 75 indicates that in 75 (or 65%) counties model 2 (4.1M) has a lower RMSE than model 1 (4.1) when coefficients are allowed to vary across counties.

Table 4.1: Root mean square error of selected one-month ahead non-spatial forecast models using a 60-month sample

Model	1	2	3	4				
	4.1	4.1M	4.2	4.2M	2 v 1	3 v 1	4 v 2	4 v 3
Vary	0.6841	0.6681	0.6774	0.6716	75 (65%)	61 (53%)	42 (37%)	74 (65%)
Constant	0.5631	0.5680	0.5699	0.5738	42 (37%)	53 (46%)	62 (54%)	55 (48%)

The overall results here are quite complex. In the cases where coefficients are held constant, it appears that adding a new predictor to the existing model actually hurts the accuracy. The comparisons for individual counties are generally consistent with the overall difference, with a majority of counties showing declines in predictive accuracy with the inclusion of additional variables. The exception is when the second lag is added to the model containing the Missouri unemployment measure (4 v 2); this improves accuracy in over half of all counties even though the accuracy based on overall RMSE declines. When the coefficients are allowed to vary, the results are generally reversed. In this specification, it helps to add new predictors; the exception to this is in the 4 v 2 model comparison. In the varying coefficients case, the overall measure of accuracy favors 4.1M, in contrast to the constant coefficients case which slightly favors 4.1.

Table 4.2: Mean absolute error of selected one-month ahead non-spatial forecast models using a 60-month sample

Model	1	2	3	4				
	4.1	4.1M	4.2	4.2M	2 v 1	3 v 1	4 v 2	4 v 3
Vary	0.4707	0.4584	0.4656	0.4607	69 (60%)	62 (54%)	38 (33%)	65 (57%)
Constant	0.4510	0.4516	0.4529	0.4534	47 (59%)	56 (49%)	57 (50%)	65 (57%)

As we can see from Table 4.2, the pattern of results is very similar to those presented in the previous table when we use the MAE, rather than the RMSE.

We are also interested in knowing how the forecasting accuracy is affected by the time period. In the previous chapters we saw that the 1990s had a much more volatile unemployment rate than the 2000s. This is evident in the forecasting accuracy as well. If we look at the RMSE and MAE for the two periods separately, we observe that in all model specifications the forecasts on the months in the 2000s are more accurate than those in the 1990s. Based on the RMSE, the average forecast accuracy for the 2000s is greater in 85% of counties than the 1990s forecast measure; overall the RMSE measure for the 2000s is 75% of the value for the 1990s.

Another consideration is whether or not allowing coefficients to vary across counties or remain constant across counties helps predictive accuracy. The results in Tables 4.1 and 4.2 indicate that allowing coefficients to vary across counties hurts the predictive accuracy. This can also be checked by comparing the accuracy of each county individually rather than the combined RMSE (or MAE) for all 115 counties. Table 4.3 shows the results of comparing predictive accuracy by county. As in the right panel of Tables 4.1 and 4.2, the number listed in the table indicates the number of counties (out of 115) for which allowing coefficients to vary provides more accurate predictions than the model restricting coefficients to be constant across counties.

Table 4.3: Pairwise comparisons of root mean square error and mean absolute error between selected non-spatial varying and constant coefficient one-month ahead non-spatial forecast models using a 60-month sample

	Vary v Con, RMSE	Vary v Con, MAE
4.1	50 (43%)	56 (48%)
4.1M	55 (48%)	61 (53%)
4.2	50 (43%)	59 (51%)
4.2M	52 (45%)	59 (51%)

While there was a large difference in the model accuracy when aggregated across counties, in over 40 percent of the counties model accuracy is reversed. This would indicate that in the cases where the varying coefficients are worse in accuracy, they are significantly worse. The cases where the varying coefficients models are better show only a minor improvement by allowing coefficients to vary.

The finding that constant coefficient models are more accurate would seem to be inconsistent with results of previous chapter. In those analyses, our comparisons indicated that allowing coefficients to vary across counties was preferable to restricting them to be the same. In contrast to results here, those results are based the full sample. It is possible that altering the sample size to build the forecasting models alters the value of allowing coefficients to vary across counties. We will explore this issue below. It is also true that this is the first time that any out-of-sample results have been presented. All previous results not only used the entire dataset but also (and as a consequence of using the full dataset) made no predictions nor drew any inference outside of the sample. This also may contribute to the differing results.

4.2 Forecast Accuracy of Spatial Models

We now move to analysis of models that include a spatial variable as a predictor. In this section, we add a first-order spatial variable to the equations that we used in section 4.2. We again suppress the county subscript as we will be estimating both sets of models (coefficients that are held constant and coefficients that vary across counties) for both Equations 4.3 and 4.4. (As above, seasonal dummies are included in the specification but suppressed in the equations.)

Equation 4.3 $$Y_t=\alpha+ \beta_1 Y_{t-1}+ \pi X_{t-1}^{FO}+\varepsilon_t$$

Equation 4.4 $$Y_t=\alpha+ \beta_1 Y_{t-1}+\beta_2 Y_{t-2} + \pi X_{t-1}^{FO}+\varepsilon_t$$

Once again, the lagged unemployment rate for Missouri is included as a predictor in models indicated by an "M". The following two tables present forecast accuracy estimates for four spatial models.

Table 4.4: Root mean square error of selected one-month ahead spatial forecast models using a 60-month sample

Model	1	2	3	4				
	4.3	4.3M	4.4	4.4M	2 v 1	3 v 1	4 v 2	4 v 3
Vary	0.9332	1.0092	1.0396	1.0432	56 (49%)	18 (16%)	34 (30%)	57 (50%)
Constant	0.5641	0.5693	0.5324	0.5754	19 (17%)	114 (99%)	57 (50%)	14 (13%)

Table 4.5: Mean absolute error of selected one-month ahead spatial forecast models, 60-month sample

Model	1	2	3	4				
	4.3	4.3M	4.4	4.4M	2 v 1	3 v 1	4 v 2	4 v 3
Vary	0.6715	0.6938	0.7508	0.7869	61 (53%)	21 (19%)	32 (28%)	59 (52%)
Constant	0.4500	0.4521	0.4215	0.4542	47 (41%)	114 (99%)	57 (50%)	13 (20%)

These results show some clear patterns. Based on the reported RMSE for the group as a whole, it appears that holding coefficients constant across counties is preferable. It is also true that adding predictors actually hurts forecasting accuracy, with the exception of the second temporal lag in the constant coefficient case (Equation 4.4 v. Equation 4.3, Constant). This again contradicts earlier results based on analysis of all counties. We return to this later.

Moving to the pairwise comparisons, it appears, in the spatial models where the coefficients are allowed to vary, adding a second lag reduces predictive accuracy in most counties. Over 80% of counties show a decline in accuracy when a second lag is added in models without the lagged Missouri unemployment rate, and the number is 70% in

models with the lagged Missouri unemployment rate. The same conclusion holds in the models where coefficients are held constant and the lagged Missouri unemployment rate is present. The one place this is not true is when coefficients are held constant and Missouri's rate is absent. In this situation, in only one county does adding the second lag hurt accuracy (Howell County in RMSE, Platte County in MAE). According to these numbers, the second temporal lag matters, but it is most appropriately estimated constraining the coefficient to be the same.

One interesting result here is that the spatial model that has two temporal lags and holds constant coefficients across counties produces more accurate forecasts on average than any other model presented to this point. It is also consistently better in the county-by-county comparisons than the two models for which comparisons are made. What this indicates is that although including spatial measures often does not improve model predictions, a model specification that includes a spatial variable produces more accurate predictions than any of the others.

We can also explore whether holding coefficients constant or allowing them to vary improves predictions, viewed in terms of the separate counties. This will give us more information regarding which structure is most appropriate.

Table 4.6: Pairwise comparisons of root mean square error and mean absolute error between selected spatial varying and constant coefficient one-month ahead spatial forecast models using a 60-month sample

	Vary v Con, RMSE	Vary v Con, MAE
4.3	20 (18%)	21 (18%)
4.3M	22 (19%)	20 (18%)
4.4	9 (7%)	8 (7%)
4.4M	9 (7%)	7 (6%)

From these results, it appears that allowing coefficients to vary across counties improves predictions in no more than 22 counties (19% of the total). It is possible that this is due to the high amount of variation in the spatial variables combined with the small sample size. The coefficients on the spatial variables are much smaller than those on the county's own temporal lag. Because of the small impact of the spatial variable and the small number of observations used to construct the model, it is not serving as an accurate predictor of the next period's unemployment rate.

Now that some forecasting estimates for both spatial and non-spatial models have been calculated, some comparisons between the two are possible. Table 4.7 compares some of the spatial and non-spatial models' accuracy.

Table 4.7 Root mean square error and mean absolute error of selected one-month ahead forecast models using a 60-month sample

Model	1	2	3	4		
	4.1,V	4.1,C	4.3,V	4.3,C	3 v 1	4 v 2
RMSE	0.6841	0.5631	0.9332	0.5641	0 (0%)	45 (39%)
MAE	0.4707	0.4510	0.6715	0.4500	0 (0%)	43 (37%)

The table shows that when coefficients are allowed to vary across counties, the non-spatial model is definitively more accurate. In particular, the RMSE and MAE of the models is higher for the spatial models (4.3,V) than the non-spatial models (4.1,V). The county-by-county comparisons reported in the two right columns of the table also imply that the spatial models are inferior. When coefficients are held constant across counties, the spatial models approach the non-spatial models in terms of accuracy. In fact, the two are nearly identical in terms of the average RMSE and MAE. The non-spatial is still slightly better on a county-by-county basis, but the difference between the two specifications narrows.

To this point, we have observed that models with coefficients restricted to be the same across counties perform better than models that allow coefficients to vary. This is true in both the non-spatial and spatial models that we have estimated. It is also true that when coefficients are allowed to vary, the spatial models are generally outperformed by the non-spatial models. This relationship disappears when the coefficients are restricted to be constant across counties. In these cases, the forecasting accuracy is usually very close for the spatial and non-spatial models, and several of the spatial models actually outperform the non-spatial models.

4.3 Forecasting Accuracy and Sample Size

The next section explores what happens when the sample size used to estimate models is changed. We expect that increasing the number of observations will improve forecasting accuracy insofar as estimation error in model parameters is important. On the other hand, if structural changes occur, longer periods may provide estimates that do not reflect current relationships, reducing prediction accuracy. We begin by increasing the number of observations to 100. This decreases the number of values for which accuracy can be calculated. Such a comparison gives a larger weight to the observations in the 2000s. We reported above that these observations have less variation in them. Because of this, the accuracy of the forecasts in this limited sample will be better than those presented above due to the smaller variation in the data used. Hence, when we undertake direct comparisons between models fitted with different numbers of months, we will fit models predicting unemployment for the same set of counties.

We begin by considering to what degree the ranking of models changes with the estimation sample. The following is a comparison of models based on the 100-month

forecasts. In the notation below, "V" implies that coefficients vary across counties, and "C" implies that coefficients are held constant across counties. As in previous tables, entries in the rightmost columns identify the number of counties in which the first model listed outperformed the second model.

Table 4.8: Root mean square error and mean absolute error of selected one-month ahead forecast models using 100-month sample

Model	1	2	3	4				
	4.1,V	4.1,C	4.3,V	4.3,C	2 v 1	3 v 1	4 v 2	4 v 3
RMSE	0.6330	0.5444	0.6060	0.5423	59 (52%)	83 (72%)	86 (75%)	61 (54%)
MAE	0.4381	0.4387	0.4259	0.4365	53 (46%)	81 (70%)	79 (69%)	49 (43%)

Here we see that the spatial models are now outperforming the non-spatial models in both the overall measure of forecasting accuracy and the pairwise comparisons. Also, while the overall measure of accuracy still favors the constant coefficients model, this is true for only about half of the counties for both spatial and non-spatial models.

If we estimate the forecast accuracy based on the 60-month models applied to the same sample used to estimate the 100-month sample, we can compare the accuracy of models based on the two sample sizes[13].

Table 4.9: Root mean square error of selected one-month ahead forecast models using 100-month samples and pairwise comparisons

RMSE	100-month		60-month			
Model	1	2	3	4		
	4.1	4.3	4.1	4.3	3 v 1	4 v 2
Vary	0.6330	0.6060	0.6182	0.8784	76 (67%)	3 (3%)
Constant	0.5444	0.5423	0.5401	0.5389	106 (93%)	101 (88%)

[13] The pairwise comparison results for MAE are nearly identical to those of the RMSE and are not presented hereafter.

The results here indicate that when coefficients are held constant, the 60-month forecasts are more accurate on a county-by-county basis. We can see however that the overall measures of accuracy are so close that these individual county differences are not likely to be statistically significant. What is quite interesting is the large improvement seen in the spatial models when coefficients are allowed to vary as we increase the number of observations used to estimate the model from 60 to 100 months. Not only is the RMSE of the 100-month sample 75% of the size in the 60-month samples but also the number of counties for which the spatial outperforms is nearly the full 115 (Howell, Mississippi and Wright being the three exceptions).

We next turn to measures of predictive accuracy of 40-month models. These estimates use the same months forecasted in the 100-month forecasts, so valid comparisons in predictive accuracy can be made.

Table 4.10: Root mean square error of selected one-month ahead forecast models, 40-month, 60-month, and 100-month forecasts, 100-month sample

Months in Forecast						
	40-month		60-month		100-month	
	4.1	4.3	4.1	4.3	4.1	4.3
Vary	0.8393	0.8083	0.6182	0.8784	0.6330	0.6060
Constant	0.4924	0.5714	0.5401	0.5389	0.5444	0.5423

In general, we can see that the best average fit uses the simplest model and with the shortest sample period. It also is the model that allows no variation across counties.

What we can see as a general pattern is the tradeoff between the complexity of the model and the number of observations used to calculate parameter estimates, where complexity refers to both the number of predictors as well as whether or not coefficients are allowed to vary. The shorter time period (40-months) produces much more accurate

forecasts if the coefficients are held constant. It also produces better forecasts with a smaller number of predictors. As we increase the number of observations, we observe the difference in accuracy between the more complex and simpler models decreases. That being said, it is nearly always true that the model with constant coefficients will outperform the corresponding model with varying coefficients.

At this point, we have presented estimates from several models using varying sample sizes. If we consider the best models in terms of forecasting error by sample size, we find that Equation 4.1 (single lag, no spatial component) performs best for the 40-month model, Equation 4.4 (two lags, spatial component) is best for the 60-month model, and the Equation 4.3 (single lag, spatial component) is best for the 100-month model. In all three cases, these models' specifications require constant coefficients across counties. Table 4.11 presents the RMSE of these specifications for each sample size. The months predicted are the same, so that the effect of varying sample sizes can be compared.

Table 4.11: RMSE for selected models (100-month sample)

	40s best	60s best	100s best
	4.1, C	4.4, C	4.3, C
40 month	0.4924	0.5639	0.5714
60 month	0.5401	0.5324	0.5389
100 month	0.5444	0.5541	0.5423

As previously stated, we can see that the more complex models perform better when more data are used for estimating parameters. As we shrink the sample size, less complicated models perform better.

One further comparison that can be made is by isolating the best model specification for each sample size and then slightly altering it and comparing across sample sizes.

Table 4.12: RMSE for best 40-month model and selected comparison models, 100-month sample

Months in Forecast	4.1, C	4.1, V	4.3, C
40 month	0.4924	0.8393	0.5714
60 month	0.5401	0.6182	0.5389
100 month	0.5444	0.6330	0.5423

From this table, based on comparing the first and the third columns, we can infer that the addition of a spatial lag with the coefficient constrained to be constant across counties hurts forecasting accuracy only for the 40-month model and helps the other two sample sizes. In the 40-month forecasts, allowing coefficients to vary (i.e., adding 114 variables) causes a large decline in the forecasting accuracy.

We also compare the model specifications for the best 60-month model and 100-month model; these results are in the succeeding Tables 4.13 and 4.14.

Table 4.13: RMSE for best 60-month model and selected comparison models, 100-month sample

Months in Forecast	4.4, C	4.4, V
40 month	0.5639	0.9197
60 month	0.5111	0.9450
100 month	0.5541	0.9024

Table 4.14: RMSE for best 100-month model and selected comparison models, 100-month sample

Months in Forecast	4.3, C	4.3, V	4.4, C
40 month	0.5714	0.8083	0.5639
60 month	0.5389	0.8784	0.5111
100 month	0.5423	0.6060	0.5541

These results show a common theme, which is that allowing coefficients to vary always increases the forecasting error and thus reduces the fit of the model. This is shown in

Tables 4.13 and 4.14, and can be seen in Table 4.12, which was a comparison of the models close to that which fit the 40-month model best.

4.4 Forecasting Accuracy and Varying Number of Lags

We have also calculated forecasting results for county models where each county has its own number of temporal lags. In chapter 3 we selected the number of lags separately for each county based on the results of BIC. Here we present 60- and 100-month forecasts based on results of these previously estimated models. We need to introduce Equations 4.5 and 4.6, which simply include a third temporal lag to the non-spatial and spatial model, respectively.

Equation 4.5 $$Y_t=\alpha+ \beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\beta_3 Y_{t-3} + \varepsilon_t$$

Equation 4.6 $$Y_t=\alpha+ \beta_1 Y_{t-1}+\beta_2 Y_{t-2} +\beta_3 Y_{t-3}+ \pi X_{t-1}^{FO}+\varepsilon_t$$

Building from the previous table in chapter 3, the counties identified as best estimated with one or two lags, i.e., with Equations 4.1 or 4.2 (4.3 or 4.4 in the cases of spatial models) are estimated here with those lags, while the remainder were estimated with three lags, i.e., using Equation 4.5 (or 4.6). For Table 4.15, the coefficients are allowed to vary across counties and thus each county's model has its own coefficient for each lag. To produce the results in Table 4.16, the coefficients are held constant. To do this, all counties (n = 24) that had one temporal lag only were estimated as a group and the coefficients were held constant. We also grouped together counties that had two temporal lags (n = 46) and three temporal lags (n = 45) and calculated the RMSE of these three groups. This column is labeled 4.X to reflect that the number of lags is not the same across counties. When a spatial variable is included, the measure carries the label Spatial 4.X. We include both RMSE values as well as pairwise comparisons.

Table 4.15: Root mean square error of 60- and 100-month forecasting models and pairwise comparisons for models with coefficients that vary by county, 100-month sample

	1	2	3	4		
Vary	4.X	4.1	4.3	Spatial 4.X	2 v 1	4 v 3
60-month	0.7088	0.6182	0.9332	0.9262	82 (72%)	75 (66%)
100-month	0.6273	0.6330	0.6060	0.6390	8 (7%)	12 (10%)

Table 4.16: Root mean square error of 60- and 100-month forecasting models and pairwise comparisons for models with coefficients that are held constant across counties, 100-month sample

	1	2	3	4		
Constant	4.X	4.1	4.3	Spatial 4.X	2 v 1	4 v 3
60-month	0.7471	0.5401	0.5389	0.6133	75 (66%)	17 (14%)
100-month	0.6823	0.5444	0.5423	0.5983	74 (65%)	20(17%)

For the most part, the results are consistent. The 4.X specification has more error than 4.1 in the 60-month forecasts with coefficients allowed to vary but slightly less than 4.3 in the same specification (Table 4.15). The pairwise results are consistent with this conclusion. When the coefficients are held constant across counties, the 4.X and Spatial 4.X (Table 4.16) have higher values for RMSE than their counterparts, which do not have a varying number of lags. The three models (4.1, 4.3, and Spatial 4.X) all produce somewhat similar results for the 100-month case when coefficients are restricted to be constant across counties according to the forecasting error. That being said, those specifications without a varying number of temporal lags by county seem to outperform those that do have a varying number of temporal lags by county. From these results, we infer that having a different number of temporal lags by county is not helpful in terms of forecasting gains.

4.5 Forecasting Accuracy and County Characteristics

One further step we can take in the analysis is to look at the county characteristics used in previous chapters. We can calculate forecasting accuracy based on whether or not a county is part of a metropolitan area, whether or not it is agricultural, and so on. The groups we consider here are the same as those in previous chapters: We will break the counties up according to the following four dimensions: metropolitan status, and employment in agriculture, education, and manufacturing. The divisions will be the same as before. Metropolitan area, and the agricultural and educational concentration categories will be broken into two groups based on high or low concentration, whereas manufacturing will be broken into three groups for high, medium, and low concentration.

The following tables present results for selected model specifications. These specifications were chosen because they showed the lowest values of RMSE for the full complement of counties in various comparisons thus far in the analysis. This particular setup will reveal differences (if any) in how these models perform when forecasting for our characteristic groups. In this table, "40", "60", or "100" indicates the number of months used to estimate the model while "V" or "C" indicates if the coefficients are allowed to vary across counties or are restricted to be constant. The number in parentheses in the top rows indicates the number of counties in that group. The models in the succeeding tables are ordered by prediction success from best to worst.

Table 4.17: RMSE of counties of selected model specifications based on metropolitan status, 100-month sample

Model	Metro (34)	Non-Metro (81)
4.1, C, 40 months	0.4937	0.4919
4.4, C, 60 months	0.5099	0.5086
4.3, C, 100 months	0.5319	0.5435
4.2, V, 60 months	0.5518	0.6253
4.1, V, 60 months	0.5941	0.6280
4.3, V, 60 months	0.9622	0.8407

Table 4.18: RMSE of counties of selected model specifications based on agricultural concentration, 100-month sample

Model	Ag (32)	Non-Ag (83)
4.1, C, 40 months	0.4871	0.4944
4.4, C, 60 months	0.4943	0.5145
4.3, C, 100 months	0.5281	0.5446
4.2, V, 60 months	0.5971	0.6074
4.1, V, 60 months	0.5992	0.6253
4.3, V, 60 months	0.8216	0.8993

Table 4.19: RMSE of counties of selected model specifications based on educational concentration, 100-month sample

Model	Educ (15)	Model	Non-Ed (100)
4.2, V, 60 months	0.4402	4.1, C, 40 months	0.4913
4.1, V, 60 months	0.4491	4.4, C, 60 months	0.5078
4.1, C, 40 months	0.4995	4.3, C, 100 months	0.541
4.4, C, 60 months	0.5167	4.2, V, 60 months	0.6255
4.3, C, 100 months	0.5336	4.1, V, 60 months	0.6397
4.3, V, 60 months	0.7647	4.3, V, 60 months	0.8942

Table 4.20: RMSE of counties of selected model specifications based on manufacturing concentration, 100-month sample

Model	Hman (27)	Mman (72)	Lman (16)
4.1, C, 40 months	0.4921	0.4883	0.5110
4.4, C, 60 months	0.5041	0.5035	0.5405
4.3, C, 100 months	0.5388	0.5316	0.5830
4.2, V, 60 months	0.6295	0.5564	0.7515
4.1, V, 60 months	0.6436	0.5598	0.7984
4.3, V, 60 months	0.9078	0.8367	1.0020

In three of the four groups, we see that the order from best predictions to worst predictions is consistent across groups. For example, in all three manufacturing concentration groups the best model is 4.1, C, 40 months, while the worst is 4.3, V, 60 months. Positions two through five are also consistent for both sets of models in each group. This is true for all groups except education. It is also the case that constant coefficients hold the top three spots and varying coefficients hold the bottom three. Education provides a slightly different outcome. Here, the order of the models is not the same for "Educ" and "Non-Educ." The high education counties have nonspatial models as their top three specifications, while the bottom three are spatial models. This group is the only one that deviates from selecting the constant coefficients as the top three models. This could be due to either the small number of counties that are designated as educational counties or the higher average income typically observed in educational counties.

Another comparison that can be made between groups is to determine if including similar counties as predictors helps the forecasting accuracy of the models. It is possible that including other metropolitan counties as predictors for a metropolitan county regardless of geographic proximity might improve the predictions of the model. We have

constructed forecasts from 60-month models for each of these groups, including a variable that is constructed from all other counties that match the characteristic of the county in question. For example, consider St. Louis County and its metropolitan status. We have included a predictor that is the unemployment rate for all other metropolitan counties in Missouri but does not include St. Louis County. They were created in the same fashion as the spatial variable; the "Metro" variable is the sum of the number of unemployed in all metropolitan counties divided by the sum of all labor forces in the same counties. This has been done for all characteristic groups. The number in parentheses in Table 4.21 indicate the number of counties used in each group.

Table 4.21: RMSE for 60-month models for characteristic groups using predictor based on characteristic using 60-month sample

	4.1, without	4.1, with	4.3, without	4.3, with
Educ (15)	0.2107	0.7251	0.6020	0.6256
Manuf (27)	0.4713	0.9359	0.9152	1.0912
Agri (32)	0.4736	1.1634	0.8265	1.1576
Metro (34)	0.4917	0.8312	1.0109	0.9331

In all cases except one, the inclusion of the characteristic group's variable hurts predictive accuracy. The one exception here is Equation 4.3 in the metropolitan case. Here it does improve prediction. For the most part, however, there does not seem to be a great advantage to adding other similar counties as predictors.

4.6 Forecasting Accuracy and State Border

Another county characteristic that has not been explored to this point is whether or not a county includes non-Missouri counties as spatially proximate predictors. All spatial variables to this point have included proximate counties without regard to state, so that counties along Missouri's boundaries included counties outside the state in the

spatial measure. In this section we separate the neighboring counties in Missouri from those that are in other states. The "Missouri" model includes only Missouri counties as neighbors, the "All (Separately)" model includes both Missouri and non-Missouri counties but separates them as two distinct predictors, and "All (FOLag)" is the same structure as previous spatial models where both Missouri and non-Missouri counties are combined. We also weight the separated spatial predictors based on the labor force size of the area as done previously; this is denoted with "All (Separate and Weighted)." The following results consider only those counties in Missouri that include counties outside the state as spatial predictors.

Table 4.22: RMSE for 60-month models for characteristic groups using predictor based on being in Missouri using 60-month sample

	4.3
Missouri Only	1.0216
All (Separately)	0.6425
All (Separate and Weighted)	0.5586
All (FOLag)	0.8063

There appears to be an advantage to including non-Missouri counties. The Missouri only model is outperformed by both other structures. There does however appear to be an advantage to separating non-Missouri counties from Missouri counties and weighting them based on labor force.

4.7 Contribution of Seasonal Measures

We continue by comparing the results of four seasonal models to the results of their corresponding non-seasonal models. To this point, all models have contained monthly seasonal dummies. This section confirms that seasonal dummies were a necessary addition. The seasonal models contain a monthly dummy for 11 months with

December the omitted month. The four models use a single temporal lag that is allowed to vary across counties, a single temporal lag restricted to be constant across counties, a temporal lag and spatial lag allowed to vary across counties, and a temporal lag and spatial lag restricted to be constant across counties. We also have three combinations of seasonal coefficients. The "Non-seasonal" models omit any seasonal coefficients, "Seasonal (Cons)" include seasonal coefficients but restrict them to be constant across counties, and "Seasonal (Vary)" include seasonal coefficients but allow them to vary across counties. These results are presented in the table below.

Table 4.23: Root mean square error of seasonal and non-seasonal forecasting models and pairwise comparisons without seasonal coefficients, with seasonal coefficients that vary across counties, and with seasonal coefficients that are constant across counties

	4.1, C	4.3, C	4.1, V	4.3, V
Non-Seasonal	0.9932	0.9929	0.9795	1.174
Seasonal (Cons)	0.5631	0.5641	1.1071	1.1123
Seasonal (Vary)	0.8732	0.8747	0.6182	0.9332

With only one exception, the seasonal models have a lower forecasting error, and the difference is particularly large in the cases where the coefficients are held constant. Looking across counties, all models show a majority of counties where the seasonal model outperforms the non-seasonal model. The results show that restricting seasonal coefficients to be constant while lag coefficients vary across counties produces worse forecasts than either holding everything constant across counties or allowing everything to vary. It appears that seasonal dummies are appropriate and should be held constant (allowed to vary) across counties when temporal lag coefficients are also held constant (allowed to vary) across counties. Seasonal models fit the data much better than non-

seasonal models regardless of how accuracy is measured, and seasonal models are much better for the more complex models considered.

4.8 Summary

The topic of this chapter is measuring forecasting accuracy of models and determining what factors affect the results. First, non-spatial models were estimated using both coefficients that vary across counties and coefficients that were held constant across counties. The conclusion for this first set of models was that coefficients held constant led to better predictions. We then considered including a temporally lagged spatial variable in the models. While there were specific counties where this was not the case, the overall measures of accuracy indicated that models constraining coefficients to be constant across counties were more accurate.

The next analysis examined various sample sizes and their effects on forecasting accuracy. While the previous models were built using 60-month sample sizes, here we estimated models using a larger sample (100-months) and smaller sample (40-months) and their effects. We found that as the complexity of the models increased, so did the optimal sample size. Smaller sample sizes produced better forecasts with the least complex model, while the more complex models performed significantly better with larger sample sizes, illustrating the tradeoff between sample size and model complexity.

We then considered county characteristics and their effects on forecasting accuracy. We checked the relationship between forecasting accuracy and the characteristics of each county (metropolitan status, agricultural concentration, manufacturing concentration, and educational concentration) and found that there was little or no improvement when similar counties were included as predictors and that in

most cases it decreased forecasting accuracy. We also checked whether or not distinguishing spatial neighbors by presence in Missouri made any difference in predictive accuracy. While including non-Missouri counties does help accuracy, it was the case that allowing effects to differ for out-of-state proximate counties was only moderately helpful. There were slight improvements to forecasting accuracy by imposing weighting structures on the non-Missouri counties. The chapter ended by checking the effects of omitting seasonal coefficients from our models. We found that seasonal coefficients should be included and should only be held constant across counties if the lag coefficients are also held constant across counties.

5. CONCLUSION

This study began by exploring the spatial structure of county-level unemployment in Missouri. The spatial structure of Missouri's unemployment was explored using two main approaches. The initial part of chapter 2 was dedicated to measuring the spatial relationship between counties at a single point in time, examining how unemployment in county i related to unemployment in county j based on physical distance between the two counties, industrial similarities, and metropolitan status. Several measures of spatial correlation were calculated, including the most commonly used measure of spatial dependence, Moran's I. The analysis next considered the variation over time in unemployment between counties, in effect focusing on the relationship between the changes in unemployment between counties. It is important to note that neither set of analyses allowed for the unemployment at one point in time to affect the unemployment at a different time. All of these measures considered unemployment in county i at time t compared to unemployment in county j at time t.

We concluded from this section that geographic distance between counties is inversely related to the correlation between unemployment rates at the county level; however the importance of distance has decreased since the turn of the century.

In chapter 3 we moved to estimating spatial and non-spatial autoregressive models predicting monthly county-level unemployment. The results in chapter 2 showed that contiguity was the most appropriate way to capture spatial relationships between county unemployment, and we have therefore used contiguity as the weight in spatial measures in subsequent chapters. Chapter 3 focused on two main aspects of these

models: which predictors should be included in the model (number of temporal lags, number of spatial lags, a larger region's unemployment rate, seasonal dummies), and whether or not coefficients on the predictors should be constrained to be constant across counties. Our models were estimated using OLS, and the primary methods of evaluation were the reported values of r-squared and statistical significance on the estimated parameters. The general conclusion was that estimated coefficients should be free to vary across counties, and that using spatial models produces better results.

Chapter 4 turned to tests of forecasting accuracy based on out-of-sample forecasts. The forecasting accuracy of these models was measured using root mean square error (RMSE) and mean absolute error (MAE). Our results were robust to the choice of these measures. When selecting which models to estimate, the conclusions from chapter 3 were critical. Chapter 3 suggested that limiting our analysis to models that allowed all coefficients to vary across counties and those that restrained all coefficients to be constant across counties was likely to identify the most important differences across models, given that hybrid models showed few constant patterns. This result was the basis for the choice of models estimated in chapter 4.

Initially, we estimated coefficients using a 60-month sample and used those estimates to forecast for the subsequent month. We fitted several spatial and non-spatial models to determine which models produced the most accurate predictions for out-of-sample forecasts. These initial results focused on the effect of adding multiple temporal lags and a larger region's unemployment rate to both spatial and non-spatial models.

We also considered forecasting models that were estimated using 40 months and 100 months of data to determine what effects a larger or smaller sample size had on the

predictions. We found that the simpler model worked much better for the smaller sample sizes, while more complex models were relatively more successful in larger samples. The accuracy of these simpler models based on smaller samples was also compared to the accuracy of more complex models derived from larger samples. Generally, the forecasting accuracy of the best models based on samples with 40, 60, and 100 months was very similar, implying that, in the range of sample sizes considered here, the costs and benefits tended to cancel.

The results from chapter 2 provided very strong evidence of a spatial relationship between Missouri counties, yet subsequent analyses indicated that the spatial component is often of little value when estimating prediction models. Exploring this point, it is possible that the spatial relationships between counties are not causal in nature. Nearby counties' unemployment rates may move together even if the unemployment rate in one county has no causal impact on proximate counties. An alternative explanation would be that a one-month lag is too long a time frame for prediction. If a county's unemployment reacts and adjusts more quickly than the one month time period, an employment shock in one county will not lead to unemployment in a neighboring county in the succeeding month. Any spatial impact that remains in the succeeding month may be too small to have any predictive influence.

There is, of course, an apparent contradiction between the results reported in chapters 3 and 4. Analyses in chapter 3 indicated both that spatial models were preferred to non-spatial models, and that spatial models allowing coefficients to vary across counties were preferred. In contrast, the forecasting models considered in chapter 4

indicated that constraining coefficients to be the same across counties produced better results and that spatial models often did not perform better than non-spatial models.

The most salient difference between the analyses in chapters 3 and 4 is the number of observations used to estimate parameters. Each and every model estimated in chapter 3 was fitted using the full complement of available data. This allowed us to get a clear picture of the spatial structure during the time period of the dataset. In contrast, no model in chapter 4 used the full dataset and all results from chapter 4 were based on out-of-sample forecasts. The largest model was fitted with 100 observations, which is less than half of the dataset used in chapter 3. It is this difference that is likely both most responsible for the difference in results and most useful to us in forming conclusions.

In chapter 3, we observe the more complex models generally performing better in all of the diagnostic measures used. In particular, chapter 3 showed that including multiple temporal lags, multiple spatial lags, and an extra predictor for larger region unemployment all contributed statistically in terms of predictive power. It was also shown that having separate coefficients for each county was preferred to holding them constant across counties. These models were estimated using the full sample of observations available. In contrast, such complex models were rejected by the results in chapter 4.

The results reflect the tradeoff between the sample size used to estimate a model's coefficients and model complexity when measured by forecasting accuracy. In chapter 4, we often found that models based on smaller numbers of months predicted as well or better than those based on longer periods. One implication is that there may have been changes in these relationships over time, so that a shorter period was more able to capture

these. The chapter 3 analyses, which use the entire sample, implied that spatial correlation is both present in the data and of importance in building prediction models. Viewed in these terms, the failure of the spatial models in these small sample sizes does not imply that spatial relationships do not exist. Rather this indicates that with county-level unemployment in Missouri, a larger number of periods is necessary to capture the nature of the spatial structure present. We see this reaffirmed by observing that the largest sample size forecasting models show the spatial models "catching up" in terms of predictive accuracy.

This also suggests that the impact of spatial correlation between counties does not change too dramatically over the life of the sample. If the changes in spatial relationships were large, it would not be possible to observe an impact of spatial measures using the models based on longer time series. What we are observing is that when forecasting unemployment at the county level, more complex models tend to improve predictive accuracy only in the larger sample sizes. Should only a limited amount of data be available, then a simpler model is preferred.

An extension beyond this paper would construct forecasts from even larger sample sizes. Our 100-month samples had the property that forecasts were built mostly from observations in the 1990s and produced forecasts about unemployment rates in the 2000s. Recall that we saw a diminished spatial importance as well as a much smaller variance in county-level unemployment in the 2000s than before. Such a structural break is expected to reduce the models' forecasting accuracy. Despite this shortcoming, the 100-month model showed spatial correlations to be present and significant enough that including a spatial measure in some models produced better predictions. When more

observations are available so that larger models can be constructed, revisiting this issue could produce interesting results.

Since the analyses in chapters 3 and 4 are structured similarly, it seems appropriate to spend a little more time breaking down how these chapters compare. Both chapters analyzed the effects of allowing coefficients to vary across counties versus restricting them to be the same across counties. As noted above, the conclusions from the chapters on this question were different: chapter 3 showed that allowing coefficients to vary across counties was unequivocally better while chapter 4 showed that constant coefficients were better more often.

There are also ways in which the results from chapters 3 and 4 are consistent. Both chapters considered models including seasonal dummies and some models that omitted seasonal dummies, and both concluded that seasonal dummies were a necessary inclusion into the models. Chapters 3 and 4 are also similar insofar as both considered the effects of including or omitting other predictors such as county characteristics, the treatment of non-Missouri counties as separate predictors, and the inclusion of the unemployment rate of a larger region. In each of these cases, the results from chapters 3 and 4 led to the same conclusions.

Both chapters also found that adding predictors for county characteristics and out-of-Missouri counties can improve the model, but only if the model allows coefficients to vary across counties. When coefficients are restricted to be constant across counties, the addition of these predictors damages the predictive power of the model. Both Chapters 3 and 4 showed that the strongest predictor was the first temporal lag and that subsequent temporal lags of a county's own unemployment rate added little predictive power.

Attempts to develop models that allowed each county to have differing numbers of lags produced minimal improvements in overall predictive power.

On the other hand, because of the difference in the summary results from both chapters it is useful to highlight the differences in what they considered. Chapter 3 considered a much wider set of specifications in both the non-spatial and spatial models than did chapter 4 with respect to the treatment of the coefficients. As noted above, chapter 4 considered only models in which all coefficients in a model were permitted to vary across counties and models in which all coefficients were held constant across counties, while chapter 3 considered hybrids where some coefficients in a specification were held constant across counties with others being allowed to vary. Since there were no discernible consistent patterns revealed from the extra comparisons considered in chapter 3, the focus of chapter 4 was narrowed to the selected models estimated.

Another difference is that chapter 4 did not use a second spatial lag in any of the prediction models. Chapter 3 did consider spatial lags and had some specifications that allowed the coefficient to vary across counties and others that held it constant across counties. As the results of chapter 3 indicated that the second spatial lag was not important, it was omitted from consideration in chapter 4.

In summary, we find that while there is evidence of a spatial structure of county-level unemployment in Missouri, it is not an important enough factor to provide any gains when estimating prediction models. One explanation for this result is the possibility that one month is too long of a time frame to accurately measure the speed at which an employment shock in one county is realized in a neighboring county. This explanation would be consistent with the results of chapter 2 showing a strong

contemporaneous spatial relationship that weakened when temporally lagged spatial variables were included (chapters 3 and 4).

A second possible explanation for these results is that the underlying relationship is not suited for providing an increase in predictive power. Our results indicate that forecasts based on fewer months are more accurate. The shorter period provides better forecasts only when the model is relatively simple. This may be due to the fact that a model based on fewer prior months will likely have noisier estimates and thus a smaller number of predictors leads to less noise and better forecasts.

The failure of spatial models is also partly due to the fact that the previous period's unemployment rate for the county itself is an extremely useful variable to include in a prediction model due to its high level of accuracy. For predictions more than one month ahead, these results could differ.

It is appropriate to make clear the limitations of the study. The conclusions drawn from this research are limited to monthly unemployment rates and forecasts that consider unemployment rates only a single month ahead. No other time-period of unemployment was used at any point and no forecasts were made beyond the succeeding month. Should other measures of unemployment be used or more distant forecasts be made, the results may differ. Also, these results were based on counties in a single state, Missouri. Extending these results to other states or perhaps even to a nationwide level was not considered here, and thus the resulting conclusions could conceivably change if those changes were made.

That being said, the limitations of the study do not preclude us from drawing some important conclusions. This research showed that there are clear tradeoffs between

sample size and model complexity when estimating models for prediction. This result is easily seen at various points in the paper. We also can see from this work that spatial models can be quite complex in many ways (e.g., weighting structure, method of including spatial neighbors) and this complexity magnifies the difficulty in selecting a model that makes proper tradeoff. These general conclusions are likely to be present in any spatial models estimated for prediction regardless of the size and scope of the underlying sample.

BIBLIOGRAPHY

Anselin, Luc and Anil Bera. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." *Handbook of Applied Economic Statistics* (1998): 237-291.

Arbia, Giuseppe. *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence.* New York: Springer, 2006.

Baller, Robert, Anselin, Luc, Messner, Steven, Deane, Glenn, and Darnell Hawkins "Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects." *Criminology* 39(3) (2001): 561-590.

Banerjee, Sudipto, Carlin, Bradley, and Alan Gelfand, *Hierarchical Modeling and Analysis for Spatial Data.* Florida: Chapman and Hall, 2004.

Bronars, Stephen and Dennis Jansen, "The Geographic Distribution of Unemployment Rates in the U. S." *Journal of Econometrics* 36 (1987): 251-279.

Clark, Todd and Michael McCracken, "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics* 101 (2001): 85-110.

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

Conley, Timothy and Giorgio Topa, "Socio-economic Distance and Unemployment." *Journal of Applied Economics* 17 (2002): 303-327.

Cracolici, Maria, Cuffaro, Miranda, and Peter Nijkamp, "Geographical Distribution of Unemployment: An Analysis of Provincial Differences in Italy." *Growth and Change* 38(4) (2007): 649-670.

Cressie, Noel. *Statistics for Spatial Data*. New York: Wiley, 1993.

Di Giacinto, Valter "A Generalized Space-Time ARMA Model with an Application to Regional Unemployment Analysis in Italy." *International Regional Science Review* 29(2) (2006): 159-198.

Diebold, Francis, and Roberto Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13 (1995): 253-265.

Elhorst, J. Paul, "Dynamic Models in Space and Time." *Geographical Analysis* 33(2) (2001): 119-140.

Feasel, Edward, and Mark Rodini, "Understanding Employment Across California Counties." *Economic Inquiry* 40(1) (2002): 12-30.

Genton, Marc and Xavier de Luna, "Spatio-temporal Autoregressive Models for U. S. Unemployment Rate." *Advances in Econometrics: Spatial and Spatiotemporal Econometrics* 18 (2004): 283-298.

Giacomini, Raffaella and Halbert White, "Tests of Conditional Predictive Ability." *Econometrica* 74(6) (2006): 1545-1578.

Griffith, D.A. 1987. *Spatial Autocorrelation: A primer.* Association of American Geographers, Resource Publications in Geography.

Kapoor, Mudit, Kelejian, Harry, and Ingmar Prucha, "Panel Data Models with Spatially Correlated Error Components." *Journal of Econometrics* 10 (2006)

Keller, Wolfgang and Shiue, Carol. "The Origin of Spatial Interaction." *Journal of Econometrics* 10 (2007)

Longhi, Simonetta, Nijkamp, Peter, and Jacques Poot, "Spatial Heterogeneity and the Wage Curve Revisited." *Journal of Regional Science* 46(4) (2006): 707-731.

Niebuhr, Annekatrin, "Spatial Interaction and Regional Unemployment in Europe." *European Journal of Spatial Development* 5 (October 2003): 1-26.

Patacchini, Eleonora. and Yves Zenou, "Spatial Dependence in Local Unemployment Rates.", *Journal of Economic Geography* 7 (2007): 169-191.

Pfeifer, Philip and Stuart Deutsch, "A STARIMA-Model Building Procedure with Application to Description and Regional Forecasting." *Transactions of the Institute of British Geographers* 5(3) (1980): 330-349.

Semple, Keith and Green, Milford. *Classification in Human Geography.* Spatial Statistics and Models, Dordecht: D. Reidel, 1986.

Stetzer, F. "Specifying Weights in Spatial Forecasting Models: The Results of Some Experiments.", *Environment and Planning A* 14 (1982): 571-584.

Whittle, P. "On Stationary Processes in the Plane." *Biometrika* 41 (1954): 434-449.

Zhou, Mo and Joseph Buongiorno, "Space-time Modeling of Timber Prices." *Journal of Agricultural and Resource Economics* 31(1) (2006): 40-56.

VITA

Dusty Sweet was born in Paris, Illinois in 1980. He graduated from Casey-Westfield H.S. in 1998 and from Eastern Illinois University in 2003 receiving a B.S. in Management and B.A. in Economics. In August of 2003, he enrolled as a graduate student at the University of Missouri-Columbia.

While at MU, he worked for five years as a teaching assistant to Dr. Sharon Ryan for Principles of Microeconomics. He taught the same class in the summer semester of 2007 and began at St. Louis Community College-Wildwood teaching Economics and Statistics in the Fall of 2007.