5-12-2010

A Local Analysis of Regional Differences in Economic Indicators and Electoral Outcomes

Mukul Sharma
University of Pennsylvania, mukuls@sas.upenn.edu

A Local Analysis of Regional Differences in Economic Indicators and Electoral Outcomes

Keywords

economy, region, voter, elections, voter preference, economics, United States, American elections, Social Sciences, Political Science, John Lapinski, Lapinski, John

Disciplines

American Politics | Behavioral Economics | Models and Methods | Other Political Science | Political Economy

A Local Analysis of Regional Differences in Economic Indicators and Electoral Outcomes

Mukul Sharma

Candidate for Bachelors of Arts in Biology and Political Science (Distinction)

Advisor: Dr. John Lapinski, Associate Professor of Political Science, University of Pennsylvania

Senior Honors Thesis, University of Pennsylvania, Department of Political Science, Spring 2010

To Mom and Dad… who have supported me throughout all of my endeavors – academic, personal, professional, and everything in between.

Table of Contents

Contents

List of Tables

List of Figures

1. Introduction

The very idea of democracy from its infancy to present day form has been predicated on the idea that citizens – in response to varying economic, social, and indeed, even political conditions – have a measurable impact in how their nation is governed. This ideal – to which politicians give countless platitudes and political scientists devote entire lifetimes to studying – is perceived to be a fundamental aspect of the political process in the United States. Since this country's founding, elections have been the means to an end – a process that, while certainly imperfect, has allowed this representative democracy to flourish and has shed light on its citizens' desires. Elections' importance, and the sheer transformative change of which they are instrumental in producing, have made this process a point of inquiry for politicians and political scientists alike.

Discussions regarding elections often entail opposing points of view – indeed, arguments, by their very nature, are the essence of political science. However, when educated individuals happen to discuss electoral dynamics, they are often faced with two separate issues – what motivates voters during election cycles and how one can predict voting behavior with even remote accuracy. Such discussions have been in the academic discourse for over forty years. Namely, the work of Kramer 1971 has sparked a flurry of activity in this field, with the express purpose of identifying voters' key motivations and preferences. The question of "why people vote the way they do" has been a key driver in American politics – each piece of research has produced works with differing stances on this issue.

(1.1) The Importance of Broad Voter Choice

The importance of broader voter choice must be further underscored. The understanding of voting patterns in the United States is dependent on a thorough understanding of voters'

motivations. Countless studies – many of which will be further elaborated upon in this analysis – have been performed to illustrate voter motivations in response to particular stimuli. However, a regional approach to variations in voter preferences has been remarkably absent from literature. Indeed, a regional approach to such an analysis could perhaps elucidate significant geographic variations in response to stimuli across the United States. It can be postulated that, to a certain extent, one could find substantial differences in the relationship between specific indicators – various measurements of changing economic conditions, for example – and electoral outcomes between regions in the United States. This is to essentially infer that there are regional factors at work either stabilizing or polarizing the degree of relatedness between independent indicators and electoral outcomes.

This study is an exercise in studying one of many drivers of voter choice – economics. This is a topic that has been qualitatively studied for over 70 years, but since the early 1970s has become a more quantitative pursuit. As discussed later in Chapter 2, this study aims to build on the work of Gerald H. Kramer's major work of 1971 – a model that successfully linked economics and voter choice in elections. By demonstrating that – in an analysis of per capita personal income, consumer cost-of-living, real income, and unemployment – per capita income has a statistically significant relationship to election returns, Kramer was able to provide an initial foundation to study the linkage between economics and politics. A more rigorous explanation of the fundamentals of Kramer's model is presented in Chapters 2 and 4.

(1.2) The Importance of Economics

Broad voter motivation can be analyzed from several different perspectives – economic, party-affiliation, or even social. The breadth and depth of potential analysis is virtually limitless, which is one of the reasons why this topic fascinates so many in the political science field.

However, the breadth of this topic makes it imperative that those performing quantitative analyses explicitly limit their work to specific realms. The purpose of this study is to elucidate the relationship between *economic* indicators and *political* outcomes on a local basis. While there are several other modifications that can be made to this fundamental line of inquiry – such as the inclusion of other potential political variables, like party affiliation – the scope and limit of this study is solely how economic changes on a local level fundamentally affect voter behavior.

The many alterations that can be made to this study are infinite, as demonstrated in Chapter 2. The ability to vary not only the universe of study, but also the fusion of both economic and political factors makes this topic fascinating. For example, as discussed in Chapter 2, Pollard et al add several *political* variables to the model presented in Kramer 1971. This change in scope presents an entirely new field to study – entirely new relationships that can be explored. Such a methodology is intriguing, but must be limited for the purpose of this study. The sheer number of different permutations and variations to modeling that can be performed make it imperative to explicitly declare the sole factors that will be studied – this study examines how economic factors relate to election returns, or voter behavior.

Moreover, the fundamental question of economics as a driver of electoral results is also of paramount importance in the context of this study. As discussed in Chapter 2, there is a tremendous amount of literature examining the relationship between the economy and voter behavior. In virtually all studies, *some basic level* of correlation between these two factors has been found – the economic metrics with the highest correlation to voter behavior are still very much debatable, but a general relationship is present in literature as discussed in Chapter 2.

On a purely qualitative level, a relationship between the state of the economy – either local or national – and voter behavior makes some intuitive sense. From an incumbent's point of

view, an economy in a generally "poor" condition can be a relative disadvantage. Voters feel some sense of dissatisfaction with the state of the economy – even as a proxy for the state of the nation – and vote based on this perceived unhappiness. It is important to note that this is a general dogma; as previously stated, voter behavior is multifaceted. There are a number of issues that contribute to why voters vote the way that they do. However, the study of the relationship between economic conditions and voter behavior in a more quantitative context – which will be further explained in Chapter 4 – can provide some insight.

(1.3) Regional Differences in Correlation

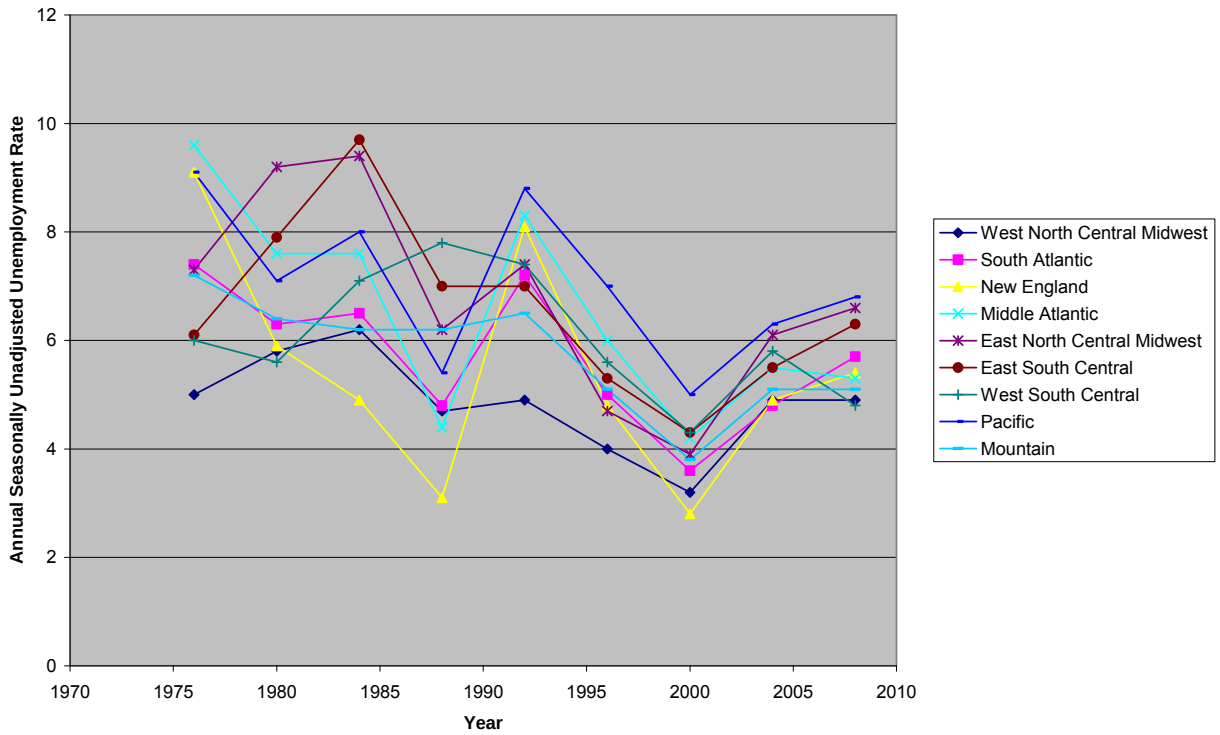
One of the elemental aspects overlooked by many political scientists performing broad national research is that not all regions of the United States are equal. There are key demographic, economic, and social differences among the many regions in this country. Therefore, it is particularly insensitive to this fact to simply aggregate, analyze, and present all regions – particularly, *all voters* – in the country as patently identical for the purpose of a quantitative analysis. Such a broad-level of analysis ignores particulate variations across the United States and serves to bury subtleties that could potentially provide significant insights.

Perhaps a more thorough look at even one isolated region in the context of the broader United States can be instructive. Indeed, one region that has a particularly rich history and wide array of diversity – demographic and economic – is the Midwest, also defined as the West North Central Midwest by the United States Census Bureau. This region encompasses the states of Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota and is emblematic of the key differences among regions throughout the United States. Historically, the Midwest has been a key region of support for both parties in the United States' two-party system. While during the 1960s through the 1980s, this region experienced significant population loss

and high unemployment – both of which have since stabilized. Recently, this region has actually trailed other areas of the country in unemployment, as well. Clearly there are several distinguishing factors – demographic and economic – that separate this region from others in the United States. The particular cultural, demographic, and economic facets of the West North Central Midwest can possibly have varying electoral impacts when compared with other regions of the United States.

Certainly, the previous discussion regarding the West North Central Midwest region further underscores the argument that not all regions should be treated identically via national-level analysis. For example, several unemployment indicators – such as the seasonally unadjusted historical unemployment rate – vary considerably across regions as demonstrated in Figure 1.1:

Figure 1.1: Historical Unemployment Rate by Region



In Figure 1.1, it is clear that some regions have significantly lower historical unemployment rates – the West North Central Midwest, included.

If analysis was performed on a strictly national level, this context would perhaps be lost. Only through a regional parsing is one able to fully analyze the marked differences in unemployment across all regions of the United States. Indeed, there are marked regional differences among demographic, economic, and social indicators – these differences can best be resolved utilizing a form of local analysis which will be discussed in greater detail in Chapter 2.

Such an observation begets the question of a regional difference – that is, is there a difference in how economically motivated individuals vote by region? Do the mechanisms that Kramer presents (and that are fully explained in Chapters 2 and 4) actually work the same across the United States? This study seeks to further examine this topic. However, certain *a priori* expectations can be defined. The reasoning behind why Kramer's defined mechanisms *could* in fact demonstrate a regional bias lies in the fundamentals of each region that have not thus far been considered in analysis – namely, that the demographic, economic, and social differences that exist in regions have some level of combinative effect on how an economically motivated voter acts. It is clear that there are fundamental differences across regions of the United States. Intuition would dictate that these differences can – and perhaps, *should* – have some level of impact on the relationship between economic factors as perceived by a voter, and his or her electoral reaction to such factors.

However, contrarian arguments to the aforementioned expected results can just as easily be argued. Many researchers could easily argue that numerous factors, including the universe of data studied or indicators used simply would not be enough to parse out such a relationship. Similarly, it could be that differences across regions are simply too small to be measured with

any sense of accuracy – that is, there is not enough of a difference to demonstrate a statistically significant variation across regions. Regardless of such contrarian assertions, the expectations of this study are that the inherent regional differences that exist throughout the United States well affect the relationship between economic indicators and voter action to some degree.

(1.4) Study Goals and Novel Treatments

As initially stated, this study clearly does not seek to fully elucidate the rationale of voters nor does it fully seek to *completely* explain the electoral dynamics of any region. Indeed, a thorough examination of individual-level voter behaviors will not specifically be performed. An individual-level analysis of voting behavior is out of this study's scope – it is more related to psychological motivators of *individuals*, rather than furthering an aggregate analysis of macro-level voting motivators for smaller groups of individuals. However, it is absolutely clear that voting is, at its very core, an individual process – aggregate-level analysis of voters' behavior relies on the assumption that voters must absorb and process some sense of economic conditions in their locality and vote based on this analysis.

Rather, the goal of this study is to shed particular light on the electoral dynamics of particular regions as defined by the United States Census Bureau while utilizing a novel approach to analysis. As demonstrated in Chapters 1.1 and 1.4, Kramer's model is an analysis of how national economic conditions affect a political party's share of the vote in congressional elections. This is a top-down analysis of how broad national economic conditions affect electoral outcomes. This study builds upon such analysis by bringing the scope of study to a more local level. As discussed in Chapters 2 and 4, this study departs from Kramer's methodology by investigating the relationship between economic conditions and electoral outcomes on a *local* level. From that analysis, a secondary aspect of this study is to investigate

regional differences in this correlation throughout the United States. This method of analysis is particularly novel because, as discussed in Chapter 2, there has not been either a localized study of the correlation between economic indicators and electoral outcomes nor has there been an analysis of geographic variations for such a correlation. Also particularly unique to this study is the fusion of both macro-level, aggregate voting analysis with what is effectively a micro-level scale – while aggregate-level economic and voting behavior will be analyzed, it will be done at a *county*-level, in order to test for greater accuracy. Such an analysis has been chosen as being representative of a voter's perception of his or her "local economy" – how he or she perceives the economy on a local scale rather than a national scale.

After a presentation of existing models and research regarding aggregate-level voting models – specifically, those relating economic changes to electoral outcomes – an independent model will be presented modeling the correlation between economic indicators and electoral outcomes within all regions in the United States. Indeed, the novelty of this study is rooted in a *local* analysis of economic relationships with electoral outcomes, and determining a *regional* difference in the quantitative relationship between economic indicators and election results.

(1.5) Analytical Expectations

The goal of this study is to effectively evaluate two statements – the degree of correlation between local economic changes electoral outcomes on a regional basis and *how* changing economic conditions affect electoral outcomes. It is first hypothesized that there are *regional differences* in how economic indicators affect electoral outcomes. There are fundamental differences in the demographic, economic, and social aspects to regions in the United States – therefore, fundamental differences in how various regions react to changing economic conditions should also be seen.

Secondly, the actual electoral outcomes from changing local economic conditions will be evaluated. Because the model that is used for this study incorporates an incumbency weight for the party in office during each Presidential election per the methodology presented in Chapter 4, it is expected that the incumbent party will suffer losses when average wage is increasing (indicating a rise in inflation), per capita income is decreasing, or unemployment is increasing – all negative economic conditions. Indeed, when evaluating three common economic indicators – namely, average wage as a proxy for local inflation, per capita income, and unemployment – it is hypothesized that several relationships will be determined. It is expected that when weighted for incumbency effects, local inflation and unemployment rates will actually have inverse relationships with electoral outcomes. That is, voters will actively vote against the party in power when these two economic indicators are rising. In contrast, per capita income should hold a positive relationship with incumbency-weighted electoral results. The rationale in literature will be presented in Chapter 2.

2. Literature Review – Kramer's Theory and Following Arguments

Since Kramer's influential work in the 1970s, a vast amount of research regarding the quantitative analysis of economics as related to political outcomes has been published into the public domain – literature ranging in topic from the appropriate level of voter analysis to even the validity of historical assumptions. However, this literature has proven insufficient in addressing the key theoretical components of this study – is there a *regional* variation in how voters respond to changing local economic conditions and does this relationship hold true on a local level. The following selection of literature pieces presents the existing lines of inquiry, as well as demonstrates the dearth of information regarding these overarching, fundamental questions. Indeed, after a complete analysis it is evident that – while there have been several variations to Kramer's method of aggregate-level, quantitative analysis over the past half century – there has been very little attention given to local and regional variations regarding changing economic indicators and electoral outcomes.

(2.1) A Pocketbook Versus Sociotropic Debate

Literature has been flooded with competing arguments regarding economic changes and aggregate voter behavior since Gerald H. Kramer's seminal work, "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964" in 1971. Works following Kramer have sought to identify links between economic changes and voter behavior, but have been varied in their data and scope. Central to identifying this relationship has been a contentious debate regarding the scope of study. Kramer has argued for a more aggregate-level approach – voters are treated as overarching groups whose behaviors can be modeled accurately and empirically (1971). Kramer further cements his stance on an aggregate-level analysis approach by effectively discounting virtually all individual-level analysis – that is, analysis performed at the individual voter level,

pursuing a more psychological approach to individual voter motivations. Such an analysis is inherently grounded in an individual, psychological level – or a "pocketbook voting" model, in which individuals vote based on their perception of their and immediate family's financial health. The majority of individual-level analysis is performed through panel data and surveys, rather than empirical analysis of economic and election data. Kramer fully concedes that at its core, individuals cast votes and aggregate-level analyses – or sociotropic voting patterns, where voters vote based on their evaluation of the entire country's economic situation – is still rooted in such a fact (1983). However, Kramer also readily acknowledges that a precise distinction between the pocketbook and sociotropic voting philosophies cannot be fully parsed (1983).

In sharp contrast, others – including Gregory B. Markus – have argued for a more individual, micro-level approach towards modeling a relationship between economic conditions and actual voter behavior (1988). Markus ascribes general voting patterns linked with economic conditions to the notion of personal choice and an individual's perception of economic stability. Indeed, the rationale behind Markus' argument is that individual-level voting patterns cannot be fully described through aggregate means – an aggregate result is not necessarily congruent with results found at the individual level. However, Markus also completely concedes severe issues with microdata-based models, as well – individuals are generally surveyed by cross-sectional means rather than longitudinally, making the study of long-term national economic changes especially difficult.

(2.2) Kramer's Quantitative, Aggregate Level Approach

Kramer's 1971 work truly proved to be integral to the study of voter behavior and motivation – a truly revolutionary piece that changed the paradigm of how political scientists viewed economic-motivated voting patterns. A portion of the transformative impact Kramer's

piece had on the field can be attributed to the existing paradigm of voter analysis in the late 1960s and early 1970s. Much of the literature debate from the 1920s to the 1940s was focused on empirical, aggregate level analysis of voter behavior – work that is very similar to Kramer's overarching arguments. However, this empirical analysis was not particularly sophisticated, having been based on very flawed assumptions and extreme limitations to practical use (Monroe 1979). After nearly two decades of such an analytical framework, quantitative, empirical studies were proven too flawed for reasonable, meaningful analysis. Indeed, the significant limitations of such narrow aggregate-level analysis effectively discounted such analysis in literature for an entire subsequent decade. Thus, a new framework of survey-based, individual-level analysis followed. While these methods were similarly not particularly accurate, they became the framework for voter analysis in response to economic fluctuations during the 1950s and 60s (Monroe 1979). One of Kramer's major transformative changes was to break such a crippling reliance on survey-based studies of voter behavior – effectively, bringing empirical analysis on an aggregate level back as a mainstream philosophy for voter behavior studies.

While Kramer's 1971 work had a revolutionary impact on the study of voter response to economic changes, there were significant issues with his analysis that were not fully addressed. Perhaps most paramount – and relevant to this study – was Kramer's failure to address regional differences. Indeed, Kramer and his numerous contemporaries have, for the most part, failed to ascribe any sense of regional nuance to their analyses. Kramer, specifically, focuses on the *national* economy in his analysis – local economic changes are patently discounted (1971). He found that there was a statistically significant correlation between several economic indicators and national electoral outcomes – however, he did not find unemployment indicators to be consistent with this relationship. His view on aggregate-level analysis is even more narrowed in

subsequent works, when he argues that analysis below a full, national-level aggregate analysis is effectively irrelevant (1983). While very likely furthering his argument against individual-level, panel studies, such an adamant opposition to aggregate analysis at a *more* micro-level discounts a substantial area of study.

Kramer's opposition stems largely from issue salience. Namely, he contends that local-level issues – or individual-level, as well – are not sufficiently salient for voters to absorb, process, and then act on in a rational manner. Because Kramer's entire aggregate model is based on the assumption that voters are effectively "rational economic-political" beings, there must be sufficient salience for voters to rationally act on issues – in this argument, Kramer essentially believes that voters rationally factor changing economic conditions into their voting behavior when casting their vote (Monroe 1979).

Methodologically, Kramer's 1971 analysis poses several problems as well. Kramer utilizes a fairly simple, yet robust model for his analysis – a model that incorporates incumbency, presidential coattail effects, and economic indicators.

$$y^0 = \beta_0 + \beta_1 T + \delta\left[\beta_2 + \beta_3\left(\frac{R_0 - R_{-1}}{R_{-1}}\right) + \beta_4\left(\frac{P_0 - P_{-1}}{P_0}\right) + \beta_5\left(U_0 - U_{-1}\right)\right] + u + \gamma^v$$

(Equation 2.1)

As illustrated in 2.1, Kramer's model is a regression, incorporating year-over-year economic changes in relative terms, as well as statistical controls for random disturbance – in this particular model, δ represents the incumbency index, *Y* represents monetary income, *P* for prices, *R* for real income, and *U* for unemployment as variables for measurement from election cycle to cycle (1971). Also, β and γ control for random disturbance parameters in this equation. However, Kramer's model relies on change in income only as that which is directly affected by the *government* (Yoon 2006). Such an assumption is arguably dangerous when evaluating voter

behavior, as income directly induced by the government theoretically represents only a small portion of the total relative change in income for any given voter during any given election cycle. Moreover, Kramer identifies voter behavior through party popularity at a national level for congressional races. While he is also incorporating coattail and incumbency effects per (2.1), he does not address anything below this level, suggesting that national economic trends dictate political changes at a more local level. The clear schism in Kramer's model is the use of *national* economic conditions and *national* party popularity to model relatively local, congressional races. Arguably, Kramer has not been able to robustly establish a clear, direct relationship between economic changes on a local level and similar-level outcomes.

(2.3) Stigler's Criticisms – Minor Statistical Changes and Philosophical Debates

Indeed, while Kramer's work certainly had a shaping role in the trajectory of the field of empirical voter behavior, his arguments do have several flaws that critics seize. George J. Stigler, one of Kramer's contemporaries, actually found that national economic effects do not exert a statistically significant effect on electoral outcomes. Analyzing data on roughly the same grounds as Kramer – namely, unemployment and real income as economic indicators – Stigler finds no statistically significant relationships when he alters Kramer's statistical methodologies (1973). By omitting the years of World War II – 1944 and 1946 – Stigler demonstrates a non-statistically significant relationship between economic conditions and elections; Kramer omitted 1942 and 1944 (Monroe 1979). Effectively, Stigler demonstrates how even the smallest changes could alter Kramer's findings and lead to a completely different conclusion.

Moreover, Stigler's criticisms are also qualitative rather than quantitative. Kramer and Stigler both touch on the more theoretical aspects of voter behavior – namely, the rationality of the average voter. Kramer argues that a rational economic-political voter can accurately perceive

changes in national economic conditions – such as rising unemployment rates – and process this information in a manner by which he or she votes differently (1971). Stigler absolutely dismisses such a notion, arguing that such minute economic changes have little salience with average voters. Furthermore, Stigler argues that voters change their voting associations so rapidly, utilizing the analogy, "it is foolish to sell one's stock in a corporation simply because that corporation has had recent reverses, and it is equally foolish to assume that the political fire is always more pleasant than the political frying pan." (Stigler 1973).

(2.4) Modeling the Effect of Economic Changes on State Elections

In response to criticisms to Kramer and Stigler's work – many of which this study also contends – Walker Pollard attempted to parse national-level analysis into a state-level model. While still focusing on electoral outcomes in national elections, Pollard studied state-level economic changes in per capita income and inflation between 1952 and 1972 (1978). However, Pollard did not keep his analysis strictly to economic causation on voting outcomes. Rather, he also incorporated the political variables of relative changes in party identification, partisanship, and presidential incumbency.

Pollard's findings revealed that political variables actually prove more demonstrative of electoral outcomes than economic variables. The most salient economic variable is inflation, in Pollard's view. However, this analysis still does not adequately address strictly *economic* changes on electoral dynamics over time – in his attempt to simultaneously analyze economic and political variables, Pollard allowed his analysis of the political factors overshadow economic indicators. This resulted in his dismissal of economic changes' relevancy, which is clearly not representative of the overall power of economics on presidential elections. Regardless of the shortfalls in Pollard's analysis, his study has become a proof-of-concept for utilizing more

narrow economic changes in national elections. Candidate saliency proves greater in presidential elections versus congressional or state races.

(2.5) Modeling Congressional Voting Patterns Related to District-Level Economic Conditions

As suggested, perhaps a more relevant analysis can be performed by a through comparison of local electoral outcomes against local economic changes. James E. Piereson performed such an analysis while studying voter motivations in gubernatorial elections – while his methodology trends towards a more accurate picture of aggregate, but still accurate, analyses, it still leaves several unanswered factors. Piereson, utilizing congressional-level economic and electoral data for the 1974 national election cycle – the relative change of economic and electoral data from 1972 to 1974 election cycles – actually finds a lack of relationship between district-level economic conditions and electoral votes (1977). Perhaps directly related to this finding can be the salience of congressional candidates and, therefore, the high incumbency rates found with congressional candidates. Piereson does not find a relationship, but it may be integral factors of his study – namely, studying congressional elections, rather than an election with lower incumbency rates and higher salience – that actually skew his results.

(2.6) Party Support Predicated on Economic Stability

A goal of this study has been to demonstrate the degree of correlation between economic conditions and voter behavior. Prior literature has indicated that there is some level of party bias during negative economic conditions. It has been hypothesized that poorer economic conditions tend to benefit Democrats in certain instances. A leading proponent of such a relationship has been Morris Fiorina, who found a marked difference in party support based on varied economic indicators. Importantly, Fiorina's analysis explicitly brings the idea of *retrospective voting* to light – the notion that voters cast votes based on their analysis of historical economic conditions

(1978). Specifically, voters in this model analyze the historical performance of the economy – local or national – compared to a historical benchmark and vote based on the analysis. This theory effectively casts voter support to the incumbent party if a given economy is performing well compared to its historical state or against the incumbent party if the opposite is true.

Fiorina's analysis dictates a slightly different approach, however. He asserts that voters also have some perceived identification between party representations – that is, that voters tend to favor Democrats on the issues of "employment, wage levels, and farm policies" and "government spending and taxation" for Republicans (Monroe 1979). While these conclusions are still significantly speculative, such a relationship is evident and supported by several other studies, including that of Wides and Klorman (Wides 1976, Klorman 1978).

However, for the purposes of this study, a more "incumbency-based" approach will be taken. Rather than attempting to confirm *a priori* expectations and generalizations regarding which party benefits from changing economic conditions, this study will weight incumbency into its model of analysis. In a manner trending after Kramer 1971, this will allow this study to answer a more useful question – *is there a strong correlation between local economic conditions and voter motivation*. That is, do individuals vote based on the positive or negative state of their local economic conditions.

(2.7) Conclusions from Literature

After a complete evaluation of the literature presented, several conclusions become evident. First, it is absolutely true that Kramer's 1971 analysis was truly in the field of quantitative social science research. In a stark departure from predecessors, Kramer demonstrated – quantitatively – that there was a casual relationship between national economic indicators and electoral outcomes in Presidential elections. However, a more precise question

was also answered. Kramer was able to demonstrate that *real personal income* held the strongest correlation between the percentage share of Republican to Democrat votes in Presidential elections (1971). He was also able to illustrate that unemployment levels – once thought to be perhaps one of the most significant drivers of electoral outcomes – to not have a statistically significant relationship to election returns (1971). Indeed, Kramer even noted that prior research failed to compare unemployment levels alongside *other economic indicators* – the relationship that was demonstrated between unemployment levels and election outcomes in previous research did not likely hold *as strong of a relationship as other economic indicators* (1971).

Subsequent research on this line of inquiry has also proven to be equally valuable. Barring the statistical arguments and critiques posed by Stigler, the work of Monroe, Piereson, and even Fiorina have attempted to refine Kramer's analysis – indeed, to make it more based on particulate arguments. Analysis has been performed down to the congressional and state level; however, there is truly a dearth of research linking local economic indicators to local results. When an election of high issue salience is chosen – namely, Presidential elections – the existing body of research still cannot give precise, quantitative predictive arguments regarding what type of *local* economic factors will drive voters. It is the hope of this research material to help in furthering the academic discussion regarding this fundamental question.

3. Data and Variables

(3.1) Kramer's Model: Data and Variables

This study aims to build upon the existing body of knowledge and line of inquiry related to Kramer's 1971 model demonstrating the correlation between economic indicators and congressional elections. As discussed in Chapter 2, subsequent studies have made significant alterations to Kramer's model, changing many aspects of his analysis in order to draw broader conclusions regarding voter preferences. Before the specific changes to metrics that have been performed in this study are presented, it is instructive to recapitulate the data and variables Kramer 1971 used.

Kramer's dependent variable was the Republican share of the two-party vote in congressional elections from 1896 to 1964. His independent variables were per capita personal income, prices (per the consumer cost-of-living index, a metric for inflation), real income (deflated monetary income), and unemployment as a fraction of the civilian labor force.

(3.2) Model Transformation for Local Analysis

It is particularly useful to initially summarize all variables and metrics used in this study and how Kramer's model has been transformed for the purpose of a county-level analysis of economic indicators related to electoral outcomes. Whereas Kramer used congressional elections by congressional district, this study will use county-level Presidential elections. For this study, the dependent variable is county-level election results and the independent variables are county-level average wage (a proxy for local inflation), county-level per capita income, and state-level unemployment rates. This allows for an analysis of how correlated these economic metrics (the independent variables) are to the electoral results (dependent variable) – that is, what effect do economic indicators have on electoral outcomes.

The justification for each of these variable choices is fully explained in Chapter 4.2; however, it is instructive to frame this subsequent discussion in this section. Narrowing the scope of study poses inherent problems and involves strategic decisions in order to preserve the integrity of analysis. There are fundamental issues in changing the scope of a relatively robust model such as Kramer 1971 – a model that incorporated *congressional* election results as the dependent variable, but *national* economic metrics as the independent variables. One of these issues is the availability of local data versus national-level data. National, time-series data is readily available; however, county-level data prior to 1990 is relatively difficult to find. While two of the indicators were maintained in a county-level and state-level manner – per capita income and seasonally unadjusted unemployment rate – Kramer's national inflation indicator was more problematic. In order to still incorporate local inflation into this study, county-level average wage was used as a proxy for national inflation rates. Finally, Kramer's weight for incumbency was also ported to this study.

The changes mentioned in this chapter are only provided as a short summary – a full explanation of the changes made to Kramer's model are presented in Chapter 4.2, along with justifications for the strategic decisions made in order to perform a local analysis while maintaining the rigor of Kramer's original model.

4. An Aggregate, County-Level Model Linking Economic Indicators to Electoral Outcomes

(4.1) Kramer's Model – A More Rigorous View

While Kramer's model was presented in Chapter 2 – particularly Model 2.1 – it is necessary to delve deeper into his methodologies on a more mathematical level. As stated in Chapter 2, Kramer's scope of study is a study of congressional election outcomes in the context of changing national economic conditions. The purpose of this study is to build upon Kramer's 1971 model by performing a local, county-level analysis of economic conditions as they correlate to electoral outcomes. The scope of this study is based on county-level results in Presidential elections, studying the relationship of these election results with county and state-level economic indicators. Because this study is essentially a variation of Kramer's model, certain aspects of his fundamental model are altered. These changes include altering the scope of his economic indicators and election results and changing an economic indicator in response to a lack of county-level availability of data. These changes are fully described in Chapter 4.2.

Effectively, Kramer created a spectrum of equations to model a national relationship between economic indicators – namely, per capita personal income in current dollars for the year of election being studied, the consumer cost-of-living index, unemployment as a percentage of the civilian labor force all as a function of time (or the year studied) – and one party's share of the two-party vote in congressional elections. This analysis was performed from the perspective of the Republican's share of the congressional party vote during on-year and midterm congressional elections. Kramer started his analysis only incorporating one variable – per capita personal income – and then modeled for linear coefficients. He then created a subsequent equation with another variable added, repeating this process with separate equations for each variable. This allowed Kramer to study the individual and combinative effects of all economic

indicators in relation to election results – Kramer was able to identify which statistically significant relationships existed between each dependent and independent variable sets.

Also of interest is Kramer's treatment of incumbency. Because he was working with national data – effectively only one universe of analysis – he could easily weight his equation with an incumbency factor based on the year of election which was positive if the incumbent President was a Republican and negative if Democrat. The ability to incorporate an incumbency bias is an important factor in creating a rigorous model; it is an issue that *will* be addressed in this study, but on a county-level basis. This issue will be further discussed in Chapter 4. However, Kramer's use of "coattail effects" will not be addressed as it was intended to directly affect *congressional* elections – inapplicable to this study, which is an analysis of county-level results in Presidential elections.

(4.2) Methodology

Kramer's model provided a quantitative framework from which to evaluate the relationship between economics and politics. For this study, Kramer's model will be modified to evaluate the claim of a regional bias between this economic and electoral relationship utilizing local, county-level data. This study will be performed per the following variables presented in Table 4.1:

Table 4.1: Regression Variable Definitions	
Variable	**Definition**
y_t	Democrat share of Presidential election results by county (percentage)
δ	incumbency index (+1 for if previous President is Democrat, -1 if Republican)
Uo, U_{-1}	seasonally unadjusted, annual unemployment rate by state, as fraction of labor force for current and preceding year (percentage)
Wo, W_{-1}	average wage by county for current and preceding year (percentage)
Io, I_{-1}	per capita income by county for current and preceding year (percentage)

A full equation of the novel model presented in this study is provided in Chapter 4.2.

However, such an evaluation does not lend itself to a direct comparison using Kramer's model – certain changes must be made. Kramer's scope of study is a *national* level of economic indicators. The major deviation from Kramer's model occurs in using average wage as a proxy for inflation – the justifications for this and other minor deviations are as follows.

Consistencies

While there are several inconsistencies between a national model such as that of Kramer and a local model that were accounted for, the overall methodology of Kramer's aggregate voting model was preserved. Three economic indicators – namely, per capita income, unemployment rates, and average wage – were found on a county-level basis from the 1980 Presidential election to the 2000 Presidential election. This data was then related to the ratio of Democratic percentage of the vote to Republican percentage of the vote during Presidential elections through ordinary least-squares multiple regression – effectively, multivariate regression. Also, Kramer's incumbency weight was also used – a +1 value was assigned if the previous President originated from the Democratic party and a -1 coefficient assigned if the

President in the previous term was a Republican. This weight will be addressed in greater detail – and in a mathematical form – in Chapter 5.

Choice of Elections

In his analysis, Kramer utilized the ratio of Republican share of the electoral vote in *congressional elections*. For the purpose of this study, Presidential elections were used – the percentage share ratio of Democrat to Republican electoral votes on a county-level basis was used. Kramer's reasoning for using congressional elections was due to a lack of candidate salience – he likened a lay-voter's voting rationale to be that of choosing between two relatively no-named candidates. The vote is then more centered on party affiliation rather than individual candidate qualities. Kramer also analyzes *both* congressional and Presidential elections in his study, incorporating "incumbency affects" and "coattail effects" affecting congressional elections.

An important consideration that should be noted is the introduction of significant third party candidates. During the 1980 and 1992 elections, there were significant third party candidates. The effects of these as related to the model presented – namely, the effect on the dependent variable, Democrat's share of the Presidential vote – should be considered. Kramer's model and this study are predicated on the analysis of a two-party vote. That is, the percentage of the county-level vote did one party receive. While this study has closely trended Kramer's analysis, the issue of independent voting poses a significant deviation. In a similar election – the 1912 Presidential election, in which there was a significant third party vote – Kramer actually omitted the entire year from analysis. This was possible as he was working with a dataset that extended from 1896 to 1964. Excluding both the 1980 and 1992 data points for this study is simply not possible due to the comparatively small number of Presidential elections that is being

analyzed. The years of analysis that were chosen for this study were based on the availability of county-level data – the trade-off between extending analysis prior to 1980 and being forced to use not use county-level data is one that was noted. In this study, omitting two Presidential elections out of six total elections studied is simply unacceptable.

Moreover, the effect of a third party candidate in 1980 and 1992 elections on this study are minimal. Analysis was performed using the methods discussed in this chapter and also substituting the Democrat share of the vote with a "Democrat to Republican election-over-election percentage change" metric. There was no significant difference between the results of both methods. The latter method allowed some mitigation of effects caused by the independent candidates in 1980 and 1992 – the results of that analysis did not significantly differ from the results presented in Chapter 5.

Incumbency

An important aspect of the creation of this regression model is the incorporation of incumbency effects. For example, it is hypothesized that if the President from 1976-1980 is a Democrat, the Democratic party will have a significant advantage when entering the 1980 election. Similarly, if a President is a Republican running for his second term, his incumbency must be incorporated into this model. This is also an effect that was correctly modeled by Kramer 1971.

For the purpose of this study, incumbency (δ) will be incorporated from the perspective of a Democrat-centric analysis – as the dependent variable is the share of the Presidential vote in a given year attained by the Democrat party, a +1 weight will be applied to all variables of that year if a Democrat is the "incumbent" per the preceding Presidential term. For example, from

1992-1996 a Democrat was President; therefore, during the 1996 election, all economic indicators will be weighted with a +1.

Analysis was, in fact, performed without such an incumbency factor initially. The results – both regression coefficients and the predictive capacity of the generated model – were completely unrepresentative of actual values and all proved to be statistically *insignificant*. Upon the addition of this incumbency weight, a major realignment of predicted election returns with actual historical election results occurred. The predictive capacity of this model increased dramatically, and several economic indicators became statistically significant, as will be presented in Chapters 5 and 7.

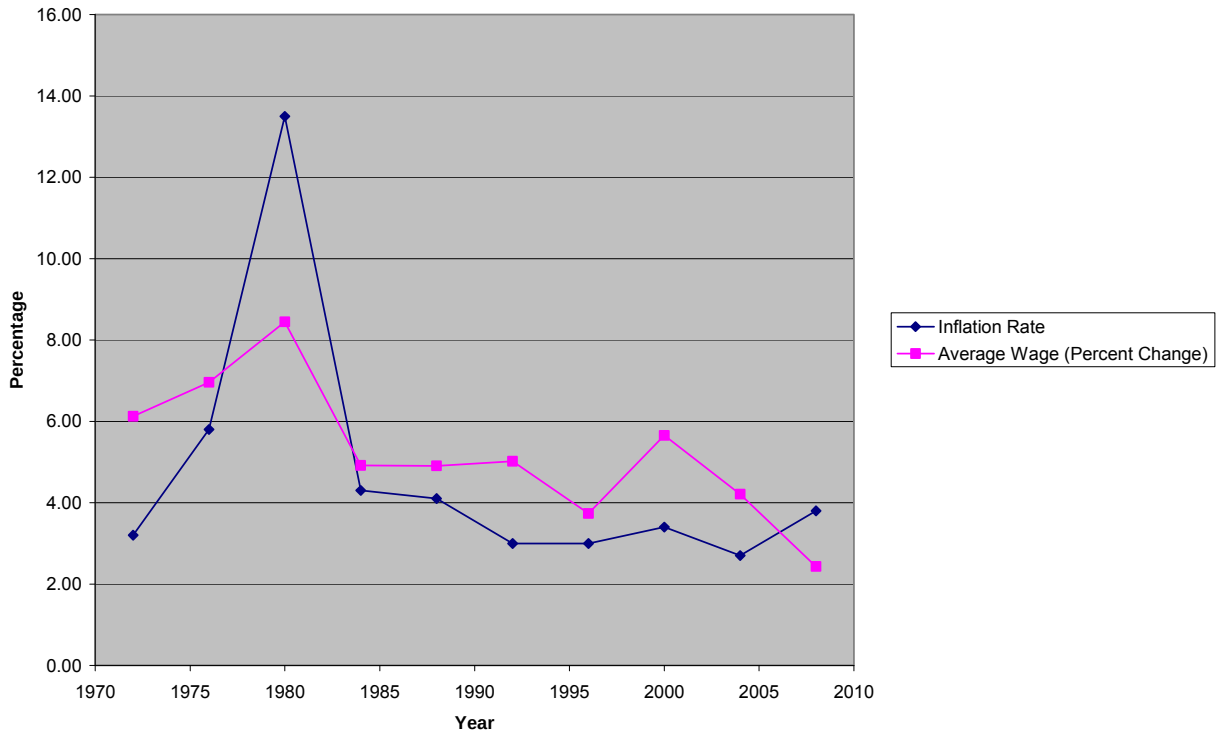
Inflation Metric

Upon even a cursory look at Kramer's 1971 model, it is quickly evident that several methods of his model would not be appropriate – or even possible, given reportable data that exists on this subject – on a local level. Perhaps most difficult of an indicator to port from a national level to a county level is the inflation variable. Federal agencies unfortunately do not calculate any version of local inflation metrics, explicitly. Metrics such as the Consumer Price Index and even the historical tally of the price of essential goods are calculated at a national and metropolitan-area level, respectively. Because local inflation metrics do not exist, one was created for the purpose of this study. Simply, average wage for each county in a time-series fashion was used as a metric of inflation.

Average wage was considered a good proxy for inflation – or the consumer cost-of-living index (prices), as used in Kramer 1971 – due to literature on the topic. A strong correlation between inflation and average wage has been established by Cukierman and Lippi (1999). While using average wage as a proxy for inflation is not entirely perfect, given that a consumer price

index does not exist on a county-level basis, this correlation serves as adequate for the purpose of this analysis.

Figure 4.1: Historical Inflation and Average Wage Rates in the United States



Indeed, as demonstrated in Figure 4.1, average wage has generally tracked inflation on a national level in the United States. As previously mentioned, average wage as a proxy for inflation is not a perfect measure. However, the availability of data, coupled with the general correlation between the two metrics, allows for a fairly representative approximation of inflation on a county-level. It should also be noted that the purpose of the inclusion of Figure 4.1 is simply to allow one to visualize this correlation – the analysis performed is on a *county-level measurement for average wage*, which will naturally provide a different correlation than what is presented in this figure.

Unemployment Metric

Kramer's model places a significant emphasis on the change in unemployment rate from each year of analysis compared to its immediately preceding year. Because Kramer's analysis is performed on a national level the *national* unemployment rate is used. However, such data was not available on a *county* level. The greatest depth of historical time-series, county-level unemployment data only went as far as 1990. Moreover, significant effort was placed in finding an adequate substitute that would accurately replace the general rationale for using unemployment data. Essentially, this entire line of inquiry proved to be fruitless, with limited data on local employment as a function of the total labor force available. Due to this consideration, state-level historical unemployment data was substituted. While it does not service the preferable universe of data, this substitute still maintains the integrity of one of the major goals of this study – to perform an analysis using the most finite form of data available to simulate a voter's cognizance of local economic conditions. It is particularly important to consider the results presented in this study in the context of one metric being of a larger universe than the otherwise highly granular data for the other two economic indicators.[1]

A Discussion on Statistics Methodology

An average of the given indicator in each county was then performed on a *regional* basis per the Census Bureau's definitions for regional divisions. For example, all economic indicators from all counties in the West North Central Midwest region, as defined by the United States Census Bureau, were averaged. This data was then transformed into a percentage change – all

[1] The robustness of substituting state-level unemployment data in lieu of county-level unemployment data is uncertain. When regression analysis was performed using the limited amount of county-level data – only 1990 through 2000 – the standard errors of coefficients was exceptionally high. This was likely because only two true data points were being used – the 1992 and 1996 Presidential elections. Such results made the decision to use a state-level unemployment data in order to gain a greater number of data points an obvious decision and resulted in a dataset that fits well to the model (see Chapter 7 for a more detailed discussion on the predictive capability of the models presented in the Results section and the "fitness" of the data to the various models).

data was calculated in percentage change from Presidential election cycle to the subsequent cycle. Finally, these averaged percentage changes for all economic indicators and election returns in each region of the United States was then regressed, using multivariate regression.

It should be recognized that this methodology poses several highly important considerations that will significantly affect the results presented in subsequent sections. Mostly importantly, this methodology treats all counties – or in the case of the unemployment rates, states as the geographical stratification – as equal. Effectively, a county with a large population is treated equally as a county with a small population. The rationale behind sacrificing significant context into the data being analyzed versus pursuing this methodology was simple – the large dataset that was created needed to be aggregated in some fashion. It was simply unreasonable to attempt to perform regressions on each county in a time-series manner. Moreover, a goal of this study is to analyze the differences – if any – that exist among various geographic regions of the United States in terms of their economic and political relationships. This pursuit would be impossible if no aggregation took place. Essentially, this method of aggregation is far from perfect. However, the benefits in analysis and logistics that were gained made such a methodology essential. Other differences in analysis were solely trivial, logistical issues.[2]

(4.3) The Step-Wise Assembly of a Model

Prior to the direct construction of the model that was primarily used to evaluate the relationship between economic indicators and election returns, it is particularly instructive to

[2] While trivial on a statistical level, it should be mentioned that the specific manner of calculation performed in this study differs slightly from that of Kramer 1971. Indeed, the net change of each economic indicator for each election cycle compared to the year immediately before it is performed. However, Kramer fails to convert indicators in his model – and even when evaluating the unemployment rate changes – into percentages. In this study, the percentage change specifically is used for these indicators. The effect of this change on regression analysis has been evaluated to be trivial, but does cause some concern when directly comparing Kramer's model – Equation 2.1 – to that of the finial model presented in Equation 4.7.

have a discussion regarding the individual aspects of this model compared *directly* to election returns. One of the key metrics as discussed in Chapter 4.2 for evaluating the affect of inflation on voting patterns is the indicator of average wage. As discussed, this indicator was used in lieu of a county-level inflation metric. However, the direct relationship between election returns and average wage should be studied. As defined in Equation 4.1, a simple linear relationship can be attributed to this indicator in relation to electoral returns.

$$y_t = \delta\left[\beta_1\left(\frac{W_o - W_{-1}}{W_{-1}}\right)\right] + \beta_o$$

(Equation 4.1)

In this equation, W_o represents the average wage on a county-level basis during the election year and W_{-1} represents the average wage in each county in the year immediately preceding the year of election. It should be noted that for *all* piece-meal assembled equations, an incumbency factor of δ was assigned based on the previous term President's party affiliation. A +1 was given to a Democrat and a -1 weight was given to a Republican Presidential predecessor.

Similar analysis can be performed independently relating the two other economic indicators – unemployment and per capita income – to electoral results. Using the following equations, it is possible to determine the independent effect of both unemployment rates and per capita income on the Democrat share of Presidential votes:

$$y_t = \delta\left[\beta_1\left(\frac{U_o - U_{-1}}{U_{-1}}\right)\right] + \beta_o$$

(Equation 4.2)

where U_o represents unemployment during the election year and U_{-1} represents unemployment rates in the year immediately preceding the year of election. This rate, as mentioned in Chapter 4.2, is an aggregate of *state* historical unemployment rates – this is not county-level data. A similar relationship is demonstrated in Equation 4.3:

$$y_t = \delta\left[\beta_1\left(\frac{I_o - I_{-1}}{I_{-1}}\right)\right] + \beta_o$$

(Equation 4.3)

with I_o representing pre capita income during the election year of study, and I_{-1} representing per capita income in the year that immediately precedes the election year. The per capita income was analyzed on a county level, and then aggregated as discussed in Chapter 4.2.

Finally, for ease of comparison in future regressor tables, multivariate, three variable equations have also been created as reference. Equations 4.4, 4.5, and 4.6 provide useful combinations of the economic indicators provided above and are as follow:

$$y_t = \delta\left[\beta_1\left(\frac{I_o - I_{-1}}{I_{-1}}\right) + \beta_2\left(\frac{W_o - W_{-1}}{W_{-1}}\right)\right] + \beta_o$$

(Equation 4.4)

$$y_t = \delta\left[\beta_1\left(\frac{I_o - I_{-1}}{I_{-1}}\right) + \beta_2\left(\frac{U_o - U_{-1}}{U_{-1}}\right)\right] + \beta_o$$

(Equation 4.5)

$$y_t = \delta\left[\beta_1\left(\frac{U_o - U_{-1}}{U_{-1}}\right) + \beta_2\left(\frac{W_o - W_{-1}}{W_{-1}}\right)\right] + \beta_o$$

(Equation 4.6)

(4.4) A Novel Model – An Aggregate, County-Level Model Linking Economic Indicators to Electoral Outcomes

The framework for this novel model can be best illustrated in Equation 4.7:

$$y_t = \delta\left[\beta_1\left(\frac{U_o - U_{-1}}{U_{-1}}\right) + \beta_2\left(\frac{I_o - I_{-1}}{I_{-1}}\right) + \beta_3\left(\frac{W_o - W_{-1}}{W_{-1}}\right)\right] + \beta_o$$

(Equation 4.7)

Equation 4.7 is modeled after the variables in Table 4.1.

The results presented in Chapter 5 are also consistent with the methodology used by Kramer 1971. Kramer actually performs a regression correlation *all variables* – and their combinative pairs – to election returns. Essentially, this method involves performing a

regression for each of the equations (Equations 4.1-4.7) presented in this section. Adding variables step-wise allows one to fully understand the relationships among all dependent and independent variables, including any synergistic effects or relationships that can be demonstrated. This study keeps Kramer's methodology consistent in this regard, as demonstrated through all data tables in Chapter 5.

5. Results

(5.1) Presentation of Data

One of the key issues of this study – indeed, one of the implicit goals of this investigation – has been to address the quantitative relationship between economic indicators and election results throughout broad swaths of the United States. Inherent in this mission is the processing of a large dataset. Therefore, it is particularly instructive to analyze the strength of these relationships in a region-by-region manner. By its very nature, this discussion will be lengthy and laden with several tables. All tables will reference equations presented throughout Chapter 4, as indicated in each table.

The initial assertion presented at the beginning of this study was that each region should have a variation in correlation between every economic indicator and electoral results. Every region of the United States has a different set of demographics and socioeconomic factors that contribute to its voting patterns. Therefore, *a priori* dictate that each region should react to economic changes in a slightly different manner with some overarching, national trends. As discussed in Chapter 1, this study is primarily focused on the relationship between economic indicators and political outcomes – other more qualitative aspects, such as cultural and social factors are not necessarily *explicitly* analyzed. However, it is reasonable to expect that the fundamentally different cultural, economic, and social demographics that comprise each region of the United States – and were evaluated in Chapter 1 of this study – would affect this economic-political relationship in *some regard*.

This is indeed the type of result set that this study has found. While it is dangerous to fully attribute these nuanced statistical differences among economic regression coefficients to demographic, economic, or even social justifications, it is important to fully realize the

differences among regions. The data presented in this section heeds these cautions, but also illustrates correlations that are stronger or weaker depending on the region of study – there *is* a regional variation among the results presented in this section. The exact causes of these differences cannot be fully elucidated at this time, but it is just simply recognized that the staggering differences between regions of the United States *can* and *have* produced some level of correlation variability on a regional basis.

An analysis of the broader data presented yields several broad conclusions across all regions. First, it appears that per capita income has the strongest correlation towards incumbency-weighted election results in Presidential elections. While in some instances this relationship appears with the wrong coefficient – that is, it appears as an inverse relationship in some regions – it is the metric with the lowest standard error, highest coefficients, and passes t-tests for statistical significance. In stark contrast, it appears that inflation metrics do not have a significant impact on election results. In virtually all regions, average wage has very small, statistically insignificant coefficient values. However, this metric does generally prove to have the correct sign. Finally, unemployment's correlation to election results varies the most out of the three metrics presented. It is the true variant among all Census Bureau regions presented in this study. This is notable because this is also the only metric that was attained on the *state* level and not county level due to a dearth of county-level data for historical unemployment. Also, such an observation is notable because it unemployment proved to be statistically *insignificant* during the Kramer 1971 study. While there have been numerous studies pointing to a statistically significant relationship between unemployment rates and electoral outcomes – including those cited by Kramer of Rees et al – in both this study and Kramer 1971, this relationship is inconsistent at best and statistically insignificant at worst.

While the abstracted conclusions presented above are useful, it is particularly interesting to analyze each region independently, looking for patterns and nuance in the data. A more fundamental comparison of regional differences will be presented in Chapter 5.2, but for now it is instructive to discuss each region independently.

West North Central Midwest

At the beginning of this study, several justifications were given as to why the West North Central Midwest should prove to have a significantly different relationship between economic and electoral indicators – and, perhaps more generally, why there should be a *regional bias* associated with this relationship. These justifications ranged from the demographic to even the structure of the regional economy. This region contains the states of Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota as previously discussed.

Indeed, after substantial analysis, it is possible to see a regional bias. However, the exceptionality that was expected in the West North Central Midwest is certainly debatable. Table 5.1 provides a complete review of the regression analysis performed on this region:

Table 5.1: Model Coefficients in West North Central Midwest United States Region

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)	(4.7)
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-1.263 (0.562)	-1.566 (0.529)	-1.195 (0.635)		-1.538 (0.684)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.347 (0.475)			0.199 (0.38)	0.337 (0.564)	0.048 (0.39)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	-0.113 (0.469)			-0.445 (0.296)		-0.040 (0.527)	-0.429 (0.385)
R^2	0.014	0.118	0.558	0.748	0.595	0.119	0.750

Standard errors are also presented in parentheses. Several observations are clearly evident in this table. First, it should be expected that the unemployment coefficients be negative as there is usually an inverse relationship between a positive – or, rising – unemployment rate and election

results, with voters shying away from incumbents during periods of high unemployment (Kramer 1971). However, in Table 5.1, this is certainly not the case. In fact, unemployment coefficients are *positive* for all equations. Also, the statistical significance of this metric should be questioned, as its standard error is very high compared to the actual coefficient. Moreover, there seems to be a robust correlation between per capita income and election results in the West North Central Midwest as demonstrated in Table 5.1. Standard error for this coefficient proves to be generally less than half of the actual coefficient value. Puzzlingly, however, the sign seems to be incorrect for this metric. It is expected that there is a *positive* relationship between per capita income and election results – that is, the incumbent benefits if per capita income is rising. In general such a metric is a measure of the effects of a growing economy, which are generally positive. Therefore, a negative coefficient for this metric is of concern.

Moreover, the inflation indicator used – average wage over time – actually proves results consistent with expectations. It is expected that this metric would have an *inverse* relationship, actually working to the detriment of an incumbent. The data presented in Table 5.1 is consistent with that – a series of negative coefficients holds true to this relationship. However, the strength of this correlation is very likely statistically insignificant for most models, but especially the Equation 4.7 model as standard error is problematically high. Finally, the dataset's fit to Equation 4.7 is also of importance. While contemporary research has, at times, placed an all too great emphasis on the R^2 coefficient of determination, it is instructive to incorporate this value into this discussion. The R^2 presented in Table 5.1 is 0.750, which is very solid in terms of the dataset's fit to Equation 4.7. However, as will be presented in later sections, this R^2 actually proves to be one of the worst regional fits to data. Again, an over-abstraction of conclusions

from this one simple value should be avoided. But such a finding is worth some level of discussion and note.

South Atlantic

The South Atlantic region includes the states of Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia. Similar conclusions can be drawn from Table 5.2. However, there are notable exceptions to this analysis. As presented below:

Table 5.2: Model Coefficients in South Atlantic United States Region

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)	(4.7)
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-1.634 (0.374)	-1.578 (0.402)	-1.664 (0.435)		-1.619 (0.464)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.138 (0.592)			-0.094 (0.288)	-0.001 (0.693)	-0.180 (0.323)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	0.246 (0.35)			0.128 (0.166)		0.246 (0.431)	0.164 (0.2)
R^2	0.110	0.013	0.826	0.855	0.832	0.110	0.875

Table 5.2 provides several stark contradictions to the data presented in Table 5.1. It appears that there is a largely *negative* correlation between incumbency-weighted unemployment and election results. The only exception to this trend appears in the coefficients presented for Equation 4.2, which happen to be positive. Moreover, the statistical significant of unemployment is questionable – it is statistically insignificant due to the vastly larger standard error as compared to the actual coefficient. Again, however, incumbency-weighted per capita income proves to have the *strongest* correlation to election results. While the correlation puzzlingly still remains negative, it proves to be a robust correlation between the two datasets. Finally, incumbency-weighted average wage actually proves to be positive in Table 5.2 – again, in stark contrast to the average wage presented in Table 5.1. This sign state is puzzling as it indicates that there is a positive relationship between an inflation indicator and pro-incumbent election results. Such an

observation is counter-intuitive. However, it should also be noted that the statistical significance of this relationship is small – in the coefficient data presented for Equation 4.7, the relationship appears statistically insignificant.

New England

Further continuing the trend of virtually ever region having differing regression coefficients, New England proves to be no exception. The states included in this region are Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. The data presented in Table 5.3 is as follows:

Table 5.3: Model Coefficients in New England United States Region

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)	(4.7)
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-1.107 (0.15)	-1.168 (0.192)	-1.071 (0.18)		-1.135 (0.224)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.323 (0.324)			0.058 (0.114)	0.238 (0.383)	0.076 (0.13)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	-0.219 (0.227)			0.051 (0.084)		-0.157 (0.266)	0.063 (0.098)
R^2	0.189	0.199	0.932	0.939	0.937	0.282	0.948

Particularly, in New England, the general trend of incumbency-weighted per capita income demonstrating the most robust correlation to election results continues. Indeed, t-tests for significance demonstrate a robust statistically significant relationship for per capita income as related to the Democrat's share of time-series election results in every regression involving this metric.

It seems that virtually every region has a unique set of regressor matrices. Indeed, it should be noted that the average wage coefficients in New England are not entirely positive nor are they negative. These coefficients' values alternate – in fact, this entire metric is actually statistically insignificant for New England. Table 5.3 does demonstrate, though, that the pattern

of average wage holding a relatively weak correlation to the Democrat share of election results continues. This pattern – when coupled with a relatively large standard error per the actual values of the regression coefficients – demonstrates a weak correlation for average wage with election results in New England.

Moreover, as in Table 5.1, it appears that the unemployment coefficients in Table 5.3 are also positive. The reasoning behind this phenomenon is unclear. It appears that standard error is quite high for this metric, as well, likely contributing to a puzzling result. Finally, some attention must be given to the strong dataset fit in this region – the coefficient of determination is 0.948, indicating a very tight fit to data. This strong fit will be further evaluated in Chapter 6, when the predictive capability of this particular model will be evaluated.

Middle Atlantic

The results for the Middle Atlantic region reflect one of the least accurate regressions performed in this study. It seems that virtually all metrics are significantly inaccurate, and standard error is quite high for all regression coefficients. This region includes New Jersey, New York, and Pennsylvania. This is best demonstrated in Table 5.4:

Table 5.4: Model Coefficients in Middle Atlantic United States Region

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)	(4.7)
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-1.271 (0.61)	-2.029 (1.008)	-1.127 (0.685)		-1.810 (1.332)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.465 (0.444)			0.272 (0.39)	0.452 (0.471)	0.173 (0.464)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	-0.251 (0.318)			0.372 (0.392)		-0.239 (0.322)	0.310 (0.494)
R^2	0.134	0.216	0.520	0.631	0.588	0.337	0.655

The regression for this region can be best described as a hybrid between the West North Central Midwest and New England regions. The average wage metric oscillates between positive and

negative, depending on the regression. Moreover, unemployment is puzzlingly positive. Also, per capita income – while still a very strong correlation – does not prove to have as strong of a correlation as in other regression datasets. The overarching conclusion after regression in this region is that standard error of coefficients is extremely high. It is difficult to identify any concrete statistically significant relationship in this matrix as standard error is often even larger than the coefficient values. This analysis is supported by t-tests for significance, which indicate that all three metrics actually have a statistically *insignificant* relationship to the Democratic share of votes in Presidential elections in this region. This lack of "good fit" of the data to the Equation 4.7 model is also supported by a comparatively low R^2 value of just 0.655. Other regions in this study have reported coefficients of determination upwards of 0.900. Regardless of the lack of data fit, the overarching trend of per capita income maintaining the strongest correlation among the three metrics continued in the Middle Atlantic region.

East North Central Midwest

Per the analysis performed in this study, the East North Central Midwest actually proves to have one of the best regressions for Equation 4.7. The East North Central Midwest contains the states of Indiana, Illinois, Michigan, Ohio, and Wisconsin. This region has the best fit of data to the regression and several of the strongest statistically significant relationships between economic indicators and the Democrat share of the election vote in Presidential elections.

Much of this analysis can be seen in Table 5.5 as follows:

Table 5.5: Model Coefficients in East North Central Midwest United States Region							
	(4.1)	**(4.2)**	**(4.3)**	**(4.4)**	**(4.5)**	**(4.6)**	**(4.7)**
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-1.380 (0.461)	-1.182 (0.238)	-1.429 (0.485)		-1.044 (0.198)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.080 (0.301)			0.146 (0.177)	-0.369 (0.278)	-0.169 (0.096)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	-0.473 (0.256)			-0.358 (0.1)		-0.760 (0.319)	-0.504 (0.112)
R^2	0.461	0.018	0.691	0.942	0.748	0.661	0.977

What makes this regression set so promising for the modeling done via Equation 4.7 is the strength of correlation and the significantly low standard coefficient error reported for each metric. Indeed, when t-tests for significance are performed, per capita income and average wage actually yield the strongest correlations – with per capita income holding true to the previously discussed patterns of high correlation to election returns. Indeed, as in virtually all other regions, per capita income actually reflects a negative coefficient sign, indicating an inverse relationship to election returns. Average wage also holds a negative sign, consistent with patterns seen in the West North Central Midwest, as well. However, a new pattern that emerges in this analysis is that of unemployment having negative values, which was only seen in the South Atlantic. Unemployment holds a statistically *insignificant* correlation in the East North Central Midwest and in this instance has a high standard error compared to the actual value of the unemployment coefficients.

Perhaps the most exciting aspect of this region, however, is the highly "fit" data to the model. Equation 4.7 holds a coefficient of determination of 0.977 – an *extremely* "fit" dataset to the model. This highly fit data is also presented in Chapter 6's predictive capability section with excellent results. Indeed, the East North Central Midwest region is perhaps the *best* representation of the relationships that this study seeks to elucidate. It accurately presents per

capita income as the strongest driver of electoral outcomes and also presents one of the other two metrics as having an effectively auxiliary affect on electoral outcomes.

East South Central

In general, the East South Central region maintains many of the same patterns seen in previous regions – namely, those seen in the West North Central Midwest and the South Atlantic. It includes the states of Alabama, Kentucky, Mississippi, and Tennessee. This regression data can be seen in Table 5.6:

Table 5.6: Model Coefficients in East North Central Midwest United States Region

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)	(4.7)
Per Capita Income $[I_o-I_{-1})/I_{-1}]$			-1.421 (0.467)	-1.505 (0.555)	-1.461 (0.494)		-1.927 (0.359)
Unemployment $[U_o-U_{-1})/U_{-1}]$		-0.148 (0.529)			-0.242 (0.31)	-0.068 (0.793)	-0.729 (0.277)
Average Wage $[W_o-W_{-1})/W_{-1}]$	0.111 (0.348)			-0.101 (0.23)		0.082 (0.524)	-0.470 (0.193)
R^2	0.025	0.019	0.699	0.717	0.750	0.027	0.937

One of the most notable aspects of the East South Central region is that virtually every coefficient is negative. The only exception to this is the average wage value for Equation 4.6. This is particularly striking as it goes against the expectations previously outlined – namely, that incumbency-weighted per capita income coefficients should be positive and incumbency-weighted unemployment and average wage coefficients should establish an inverse relationship with the Democrat share of election returns with negative regression coefficients.

Once again, per capita income provides a robust, statistically significant relationship between it and the Democrat share of election returns. This claim is further supported by statistically significant t-test for significance values. However, it appears that unemployment and do not provide as robust of an independent correlation to election returns. These two

metrics' standard error of coefficients still remains relatively high, although not as overbearing as in other regions. Finally, there appears to be a strong fit of this data to the model presented per Equation 4.7 – the coefficient of determination is above 0.90, indicating a good fit.

West South Central

The West South Central region is exceptional in that it provides one of the most persuasive arguments in support of per capita income as the leading economic indicator for electoral outcomes. This region includes the states of Arkansas, Louisiana, Oklahoma, and Texas. This region has a strong fit of data to the model and actually has given results that are generally consistent with expectations. Table 5.7 presents the regression data supporting such claims:

Table 5.7: Model Coefficients in West South Central United States Region

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)	(4.7)
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-2.590 (0.398)	-2.507 (0.394)	-2.496 (0.275)		-2.489 (0.337)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.583 (0.725)			0.374 (0.158)	0.415 (1.045)	0.350 (0.241)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	-0.459 (0.640)			-0.220 (0.198)		-0.238 (0.909)	-0.035 (0.211)
R^2	0.114	0.139	0.914	0.939	0.970	0.158	0.970

First, per capita income's linear coefficients in this model are incredibly robust. A coefficient of greater than 2.0 with such a low standard coefficient error is exceptional for this study. While the per capita income coefficients still remain negative – a common, yet puzzling, seen throughout the regressions performed in this study – it is this robust relationship that is especially promising. Statistical evidence also supports the robust relationship between per capita income and electoral outcomes, with a vast statistically significant relationship demonstrated through t-

tests for significance. Indeed, per capita income is the chief driver of the Democrat's share of votes for this region, in this study.

The other two metrics provide results similar to previously discussed regions. Again, unemployment coefficients are positive – an observation that is contrary to expectations. However, standard error for this metric is generally lower than in previous regions (barring the extraneous results seen for the unemployment coefficient for Equation 4.6 in this region). Also, the average wage coefficients appear to follow expectations – they have negative values and hold a relatively smaller level of correlation compared to per capita income.

These observations are very much consistent with results seen in previously discussed regions – that is, per capita income has the strongest correlation, average wage is a consistently weak force, and that unemployment coefficients have a variable effect. Moreover, what is particularly interesting and indicative of the strong level of "fit" for this data to the model is the extremely high coefficient of determination value. As seen in the East North Central Midwest region, the R^2 value is extremely high at 0.970 for the West South Central region. Not only is this promising relating to the predictive capability for this model in this particular region, but it is also doubly indicative of the earlier claims made in support of per capita income as a strong driving force for electoral results in this region.

Pacific

The Pacific region of the United States seems to provide data that runs slightly contrary to that of previously presented regions. This region includes the states of Alaska, California, Hawaii, Oregon, and Washington. Indeed, several maxims outlined in this chapter are still consistent – per capita income still maintains the greatest impact on electoral outcomes in this

region. However, the robustness of all economic metrics to electoral outcomes in this region is debatable. The regression analysis can be illustrated through Table 5.8:

Table 5.8: Model Coefficients in Pacific United States Region							
	(4.1)	**(4.2)**	**(4.3)**	**(4.4)**	**(4.5)**	**(4.6)**	**(4.7)**
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-1.185 (0.72)	-0.663 (0.928)	-1.106 (0.516)		-0.601 (0.539)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.762 (0.447)			0.714 (0.325)	0.720 (0.279)	0.707 (0.269)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	0.255 (0.139)			0.173 (0.187)		0.242 (0.090)	0.168 (0.109)
R^2	0.456	0.421	0.403	0.535	0.771	0.831	0.895

While the data in this region fits the regression well – the coefficient of determination is relatively high at 0.895 – standard error of coefficients for per capita income and average wage is very high relative to the coefficient values. For example, average wage has all positive values – again, contrary to the expected relationship that average wage, an inflation indicator, should hold to electoral outcomes – but has very high standard error. Its standard error is even higher than the coefficient values for several regressions. This relationship is not statistically significant.

Similarly, unlike other regions, the relationship between per capita income and Democrat's share of the election results is actually not statistically significant in this instance. Standard error is incredibly high – at levels not seen in any other region. Yet it is still negative, as seen in all other regions despite an apparent anomaly in all other relationships. What is particularly concerning is the low coefficient present for per capita income in Equation 4.7 – this study's chief model. Finally, it appears that unemployment actually holds the strongest correlation to electoral outcomes in the Pacific region. The relative ratio of the regression coefficients to their standard error is low, compared to the rest of the data presented in Table 5.8.

Indeed, it appears that while the data "fits" the models presented in Table 5.8 relatively well, the conclusions that can be drawn from the Pacific region are limited. All analysis points to these regressions for this region to be an anomaly, as the overarching conclusions that are very much applicable to other regions are simply not as robust – or not even statistically significant – in the Pacific region. The data's fit to this model is effectively irrelevant. The economic metrics' correlation to electoral outcomes is simply not statistically significant – and therefore, can be interpreted as ineffective drivers of electoral outcomes – in this region. Perhaps the most useful aspect of this analysis, however, is the clear evidence for a geographical bias in the effects of the set of correlations this study seeks to analyze. Regions such as the East North Central Midwest – as demonstrated in Table 5.5 – can have significant and robust correlations, while the Pacific region can have a dearth of virtually all identified relationships. The novelty of a regional analysis allows one to actually parse conclusions to a geographical level, rather than simply generalizing such an abstract relationship to the entire country – the latter was the nature of analysis performed by Kramer 1971, a generalized conclusion and model applied to the entire United States.

Mountain

While the Mountain region of the United States has data that very strongly fits to its Equation 4.7 model, there are a number of observations which are actually quite different in this region compared to other regions of the United States. The regression matrix that was used can provide significant context through Table 5.9:

Table 5.9: Model Coefficients in Mountain United States Region							
	(4.1)	**(4.2)**	**(4.3)**	**(4.4)**	**(4.5)**	**(4.6)**	**(4.7)**
Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$			-1.113 (0.97)	-1.076 (1.114)	-1.490 (0.345)		-1.480 (0.424)
Unemployment $\delta[U_o-U_{-1})/U_{-1}]$		0.746 (0.369)			0.885 (0.162)	0.732 (0.422)	0.881 (0.199)
Average Wage $\delta[W_o-W_{-1})/W_{-1}]$	0.083 (0.207)			0.059 (0.21)		0.053 (0.17)	0.013 (0.079)
R^2	0.039	0.505	0.248	0.267	0.931	0.521	0.932

Indeed, the per capita income figures maintain the same negative sign seen throughout this metric in previous analyses in this study. However, what is particularly interesting is the high rate of standard error in this metric. In fact, when all three economic metrics are regressed per Equation 4.7, per capita income actually has a *statistically insignificant* correlation to the Democratic share of election returns in Presidential elections. The metric that has been the strongest indicator of electoral outcomes for all other regions in this study – per capita income – actually proves to not be statistically significant factor for Equation 4.7. However, when average wage is removed as a factor, per Equation 4.5, per capita income's statistical significance returns. It appears that average wage is so grossly insignificant of a factor in this regression, that it actually distorts the regression to a very high degree such that per capita income becomes statistically insignificant per t-tests for significance.

Another notable deviation from the norms presented thus far in this study is that unemployment actually appears to have a strong statistical significance for Equation 4.7 in the Mountain region. During the 4.7 regression for the Mountain region, it is, in fact, the only statistically significant factor affecting the Democratic share of Presidential elections. Furthermore, a particularly puzzling deviation from other regions is that both average wage and unemployment coefficients are *positive* in this region's regressions – a state that is not seen in

any other region. Both factors should hold inverse relationships – that is, be *negative* – to election returns, as they are incumbency-weighted. Because this does not hold true, it is a particularly strange deviation.

The Mountain region actually demonstrates one of the best dataset "fit" to the models out of all regions studied. The coefficient of determination is 0.932, indicating an excellent fit of data. However, the several strange deviations from all other regions studied makes this regression set particularly interesting. Finally, while per capita income solely in this region does not prove to have the *most* significant impact on electoral outcomes, it is still a major factor affecting the Democrat share of election results. The statistically insignificant results found in the Equation 4.7 regression for per capita income is truly an anomaly caused by a major deviation in average wage. This level of data skewing is further discussed in Chapter 5, when increased measurement of data as a method for increasing accuracy is evaluated.

(5.2) Evidence for a Regional Difference

While the strength of correlation between the three economic indicators studied and actual election results is questionable – and will be discussed at length in the subsequent section – it is clear that there appears to be a strong difference in strength of correlation among various geographic regions of the United States. Effectively, it is clear that there *is a geographical and regional difference* in how economic metrics are related to election results.

The results given in various regions of the United States – for example, the East North Central Midwest compared to the Mountain region – is marked. As demonstrated in Tables 5.5 and 5.9, there is a clear difference in the coefficients found after regression.

Indeed, Table 5.5 can be compared to Table 5.9, which represents the Mountain region. Table 5.9 certainly has several striking differences. First, all of the average wage coefficients are

positive in Table 5.9 versus negative in the previous table. This is likely due to a number of reasons – including statistical externalities that were not accounted for in this relatively limited statistical model. However, such a marked difference in coefficients for particularly distinct regions is striking. It should be noted that similarities are also evident. It seems that per capita income serves as an inverse correlation for both regions – much more strongly related in the East North Central Midwest than in the Mountain region.

It should be absolutely recognized that there are still significant issues with the strength of correlation as presented in Tables 5.5 and 5.9. Namely, the standard error for each coefficient is arguably high in virtually every unit of data presented. However, even with this error the significant variation between these two regions does provide even evidence to support some *significant level* of a regional variation in this relationship. There are fundamental differences in the strength of correlation between these two regions

(5.3) Strength of Correlation

From the results presented in Chapter 4.1, it is absolutely clear that of the three economic indicators studied – average wage as an inflation indicator, per capita income, and unemployment rate – per capita income has the strongest correlation to the Democrat share of the general election results in Presidential elections. The other two indicators have demonstrated statistical significance in certain instances and particular regions. However, per capita income has been a robust indicator, demonstrating statistical significance in all regions except the Mountain region. (The latter case was also resolved by simply reevaluating statistical significance after removing the errant results of average wage from the regression. Effectively, this aberration was caused by another indicator distorting the results.)

The correlation of average wage and unemployment rate is very much debatable. Several t-tests for significance indicate that unemployment rate held very little statistical significance – or correlation to election results – for all regions except the Mountain region. These results are also reflective of the results presented in Kramer 1971 – unemployment simply was not a robust factor and was, in fact, statistically *insignificant*.

6. Data Verification

The integrity of data used for analysis is a problem that plagues virtually all forms of quantitative analysis. Departments across the bureaucratic spectrum – local, county, state, and federal – present a myriad of statistics that are calculated and measured inconsistently. The purpose of this chapter is to address such inconsistencies in analysis, demonstrate the integrity of the data presented, and illustrate steps taken to increase robustness of analysis.

(6.1) Increased Measureable Variables Increases Accuracy

A considerable amount of thought was given to the original model as presented in the fourth chapter of this analysis. While intentional consideration was given to remaining true to Kramer's original model for comparative consistency, several variations that were added actually increase this analysis' robustness. A greater number of measureable factors did marginally correlate to increased accuracy of analysis (Ansolabehere 2008). Because several additional factors – namely, the proxy indicator for local inflation given through average wage – have been added to this analysis, further accuracy is expected.

Indeed, it is possible to see that multiple survey items do, in fact, generally decrease measurement error – especially over the medium-term time period indicated in this study. However, the important caveat to this framework is that in many instances – especially in the models presented in this study – measurement error has actually *increased*. That is, a t-test of statistical significance has actually demonstrated a *worse* regression fit to the data with an increasing number of measurement variables. For example, the t-values are actually negatively affected in Table 6.1 actually decreases observed t-values when adding average wage:

Table 6.1: Regression Coefficients and T-Values in Mountain Region			
	Per Capita Income $\delta[I_o-I_{-1})/I_{-1}]$	**Unemployment** $\delta[U_o-U_{-1})/U_{-1}]$	**Average Wage** $\delta[W_o-W_{-1})/W_{-1}]$
Equation (4.6)	-1.490 (0.345)	0.885 (0.162)	
t_{obs}-t_{crit} **(4.6)**	1.133	2.284	
Equation (4.7)	-1.480 (0.424)	0.881 (0.199)	0.013 (0.079)
t_{obs}-t_{crit} **(4.7)**	-0.815	0.132	-4.132

Certainly, Table 6.1 illustrates that adding average wage in this specific example actually decreases the t_{obs}-t_{crit} (when this value is positive, the correlation of economic indicator to Democrat share of the election results is *statistically significant*; when negative, this relationship is *statistically insignificant*) of per capita income such that it becomes negative. Therefore, per capita income actually becomes statistically insignificant due to the extraneous effects of adding average wage as an indicator. The major conclusion that is drawn from Table 6.1 is that adding multiple variables of measurement does not necessarily increase the accuracy of results. In this instance, adding the relatively erroneous average wage factor – it should be noted that it is erroneous only for this region largely due to the very high relative standard coefficient error of 0.079, which is larger than the coefficient value of 0.013 – actually makes per capita income insignificant. Without a step-wise analysis comparing the correlation and statistical significance of all variables to each other and to the dependent variable, this relationship – and the negative externality that average wage creates, by making per capita income insignificant in Equation 4.7 – would not have been isolated. Not only is Table 6.1 indicative of the caution that should be taken in including extra variables of measurement, but it also instructive regarding statistical analysis.

It should be noted that this phenomenon is generally the exception in this study, but is mentioned due to its overall importance for quantitative analysis – usually a greater number of measurement variables can actually *increase* the reliability of a model, barring extraneous effects. For this study, the situation illustrated in Table 6.1 only occurred once – in the Mountain region.

(6.2) Fundamental Problems in Statistical Research – Competing Levels of Data

While political scientists will often be among the first to argue that government agency data can be misleading at times, it is often the only data available for analysis. This data is measured through a number of possible scenarios – calculated, survey, or a hybrid. Generally, survey data is perhaps the most robust form of material that can be analyzed. However, often state agencies will simply calculate statistics on demographics, economic indicators, and other reportable statistics. Such competing measurements directly force researchers to question whether their data is valid. The difficulty in assessing local data's authenticity lies in whether it is calculated or surveyed. The purpose of this section is to compare and contrast the integrity of the various sources of data presented in Section 3 utilizing variance analysis to demonstrate the wide range of conclusions and extrapolations that would be possible given competing forms of data.

(6.3) How to Analyze Particulate Data – A Problem of Parsing Data

One of the difficulties – and really a major weakness – of this study has been the transformation of particulate, county-level data into broader regional abstractions. As outlined in Chapter 4, the methodology for this study was to utilize county-level data for average wage and per capita income, coupled with state-level unemployment rates – all to provide regional generalizations about how these economic indicators relate to the Democrat share of Presidential

elections on a county-level basis. It is evident that some level of abstraction must be performed in order to transform this data into broader regional conclusions. The methodology chosen for this goal was to effectively average each county-level economic indicator by region. The inherent problem with this methodology is that all counties are effectively treated as the same, regardless of size. When federal agencies calculate average wage, per capita income, and unemployment rates, it is done by considering entire population totals – this was not performed in this study, and is a concern that is well taken.

However, this preceding discussion is one that is central to any researcher wanting to utilize local-level data in order to make broader assertions. Is the method previously discussed effective? The answer to this question is debatable. Per the discussion in Chapter 6, the models that were created in this study were fairly robust. That is not to say that they cannot be improved, but it is clear that on some fundamental level, the methodology performed in this study did perform well.

In social science research, there are those investigators who are effectively *parsers* and those who are *aggregators* – that is, those who believe that local, particulate data is a more effective level of analysis and those that believe that aggregating data proves to be much more effective. For the course of this study, had the particulate data been kept in county-level form, it would have been virtually impossible to perform enough regressions to accurately obtain any evidence of geographical biases or differences among various regions of the United States.

The alternative to this study's chosen method of analysis is a simple analysis of broader regional statistics. There is an availability of Census-defined regional data regarding average wage, election returns, per capita income, and even unemployment data. Such a method of analysis is perfectly valid and may even prove to be more accurate in some respects. However, it

was not chosen as the universe of data for this study as a more local-centric approach was desired.

(6.4) Data Availability – Historical, Time-Series Local Data is Absent

One of the greatest challenges of this study has been simply obtaining accurate, local data. For recent analysis – 1990s and onwards – obtaining parsed, local historical data is relatively easy. Several United States federal agencies make such data readily available. However, considering time-series data prior to 1990 on a county level becomes more problematic. County-level data from the 1970s is simply not available in certain cases – hence the major structural changes that this study performed to Kramer's model. Local metrics and robust substitutes for inflation, for example, were not recorded on a county-level basis prior to approximately 1990. Therefore, this study substituted one of the few inflation-resembling metrics that have been available on a county-level, time-series basis – average wage.

One of the more surprising aspects of the dearth of local, county-level data was the lack of reliable county-level unemployment rates prior to 1990. Agencies such as the Bureau of Labor Statistics and the Bureau of Economic Analysis simply did not go below the universe of state and major metropolitan area data during the 1970s and 1980s for employment (and unemployment figures). Therefore, this study attempted to use the next best thing – state-level data.

This discussion – and in a way, this study overall – illustrates an issue that many social science researchers are facing: a problem of data tradeoffs. It would be perfectly logical for this study to simply use data from 1990-2000 in its regression analysis, as this was the local data available for all three economic metrics. However, such an analysis represented a tradeoff that seemed unreasonable to make – a sacrifice of the robustness of the generated regressions and

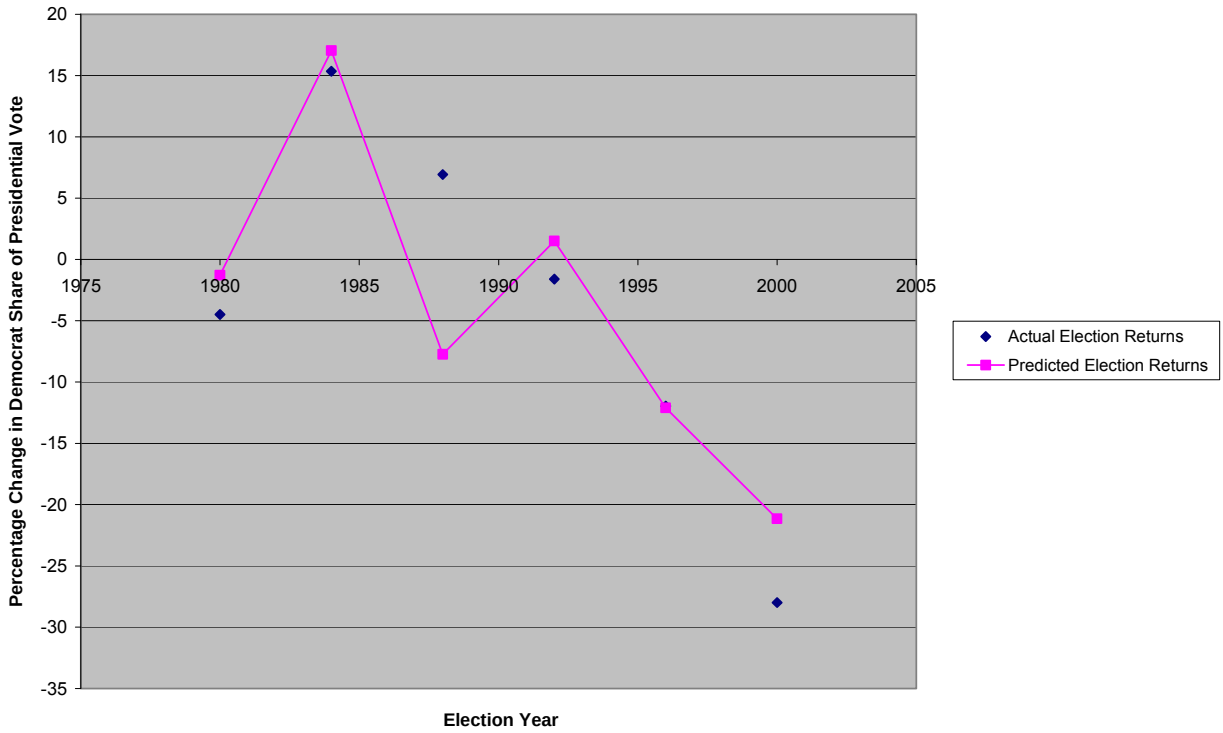
models for the sake of simply preserving an entirely "local" dataset of economic indicators. As demonstrated in Chapter 7, the predictive capability of the generated models is actually quite robust – it is because a large swath of 20 years was able to be evaluated, rather than only two election cycles as otherwise would have been possible.

7. Predictive Capability

(7.1) Historical Versus Predicted Election Results – Model Efficacy

Given the regional differences discussed in Chapter 5 and the meaningful conclusions that were able to be drawn from the regressions performed per Equation 4.7, it is imperative to test the robustness of this model's predictive capability. Because significantly different coefficients were found for each region, predictive scores have been performed on a regional basis compared to the actual election data that helped construct the models. The predictive capability of Equation 4.7 using the coefficients and economic data found in the West North Central Midwest is illustrated in Figure 7.1:

Figure 7.1: Predictive Capability in the West North Central Midwest Region

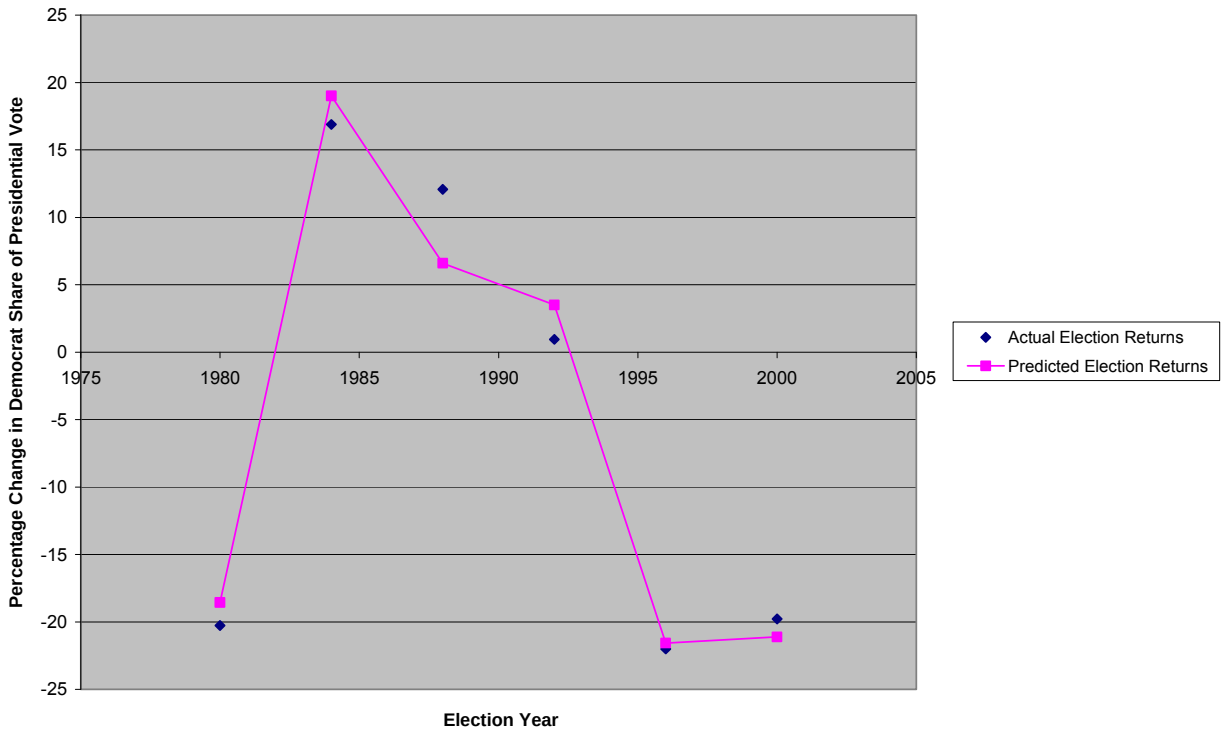


On a simple visual analysis, it is clear that actual Presidential Democratic vote shares are generally in line with this model. The two major exceptions are the 1984 election – in which

Democrats outperformed considerably from 1980 – and the 1992 election. The most egregious error in this figure comes during the 1988 election, when the model vastly overestimates the downward trend in the Democratic share of the vote.

However, again a regional difference among these models emerges. As demonstrated in Figure 7.2, there is a much tighter correlation between predicted election returns and actual historical values for the West South Central region as compared to the West North Central Midwest region:

Figure 7.2: Predictive Capability in the West South Central Region



It is clear that this region has a *significantly* better representation of model fit than that of Figure 7.1. Once again, the strength of correlation – and therefore, these models' predictive power – seems to have a strong geographic bias or difference.

Indeed, the strength of the predictive capability of this model can also be demonstrated through Figures 7.3 and 7.4, which illustrate the very tight linear fit of predicted annual election results to actual historical values in the New England and Mountain regions, respectively:

Figure 7.3: Predictive Capability in the New England Region

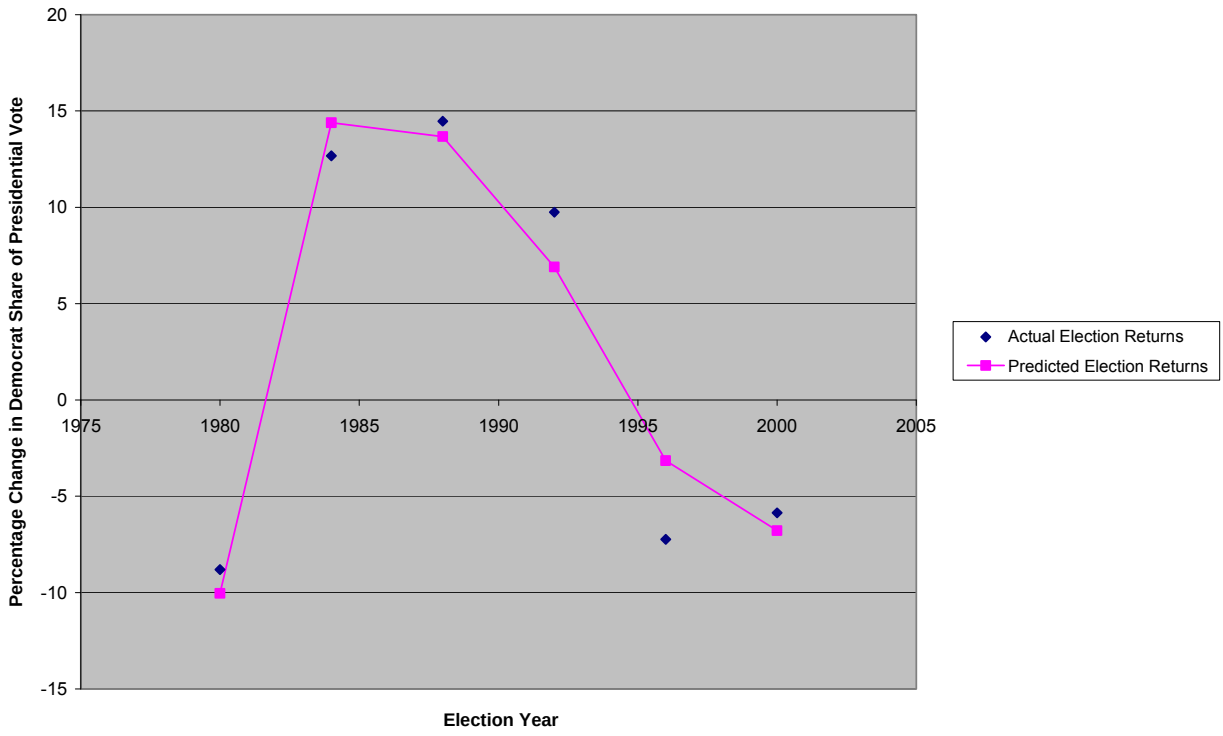
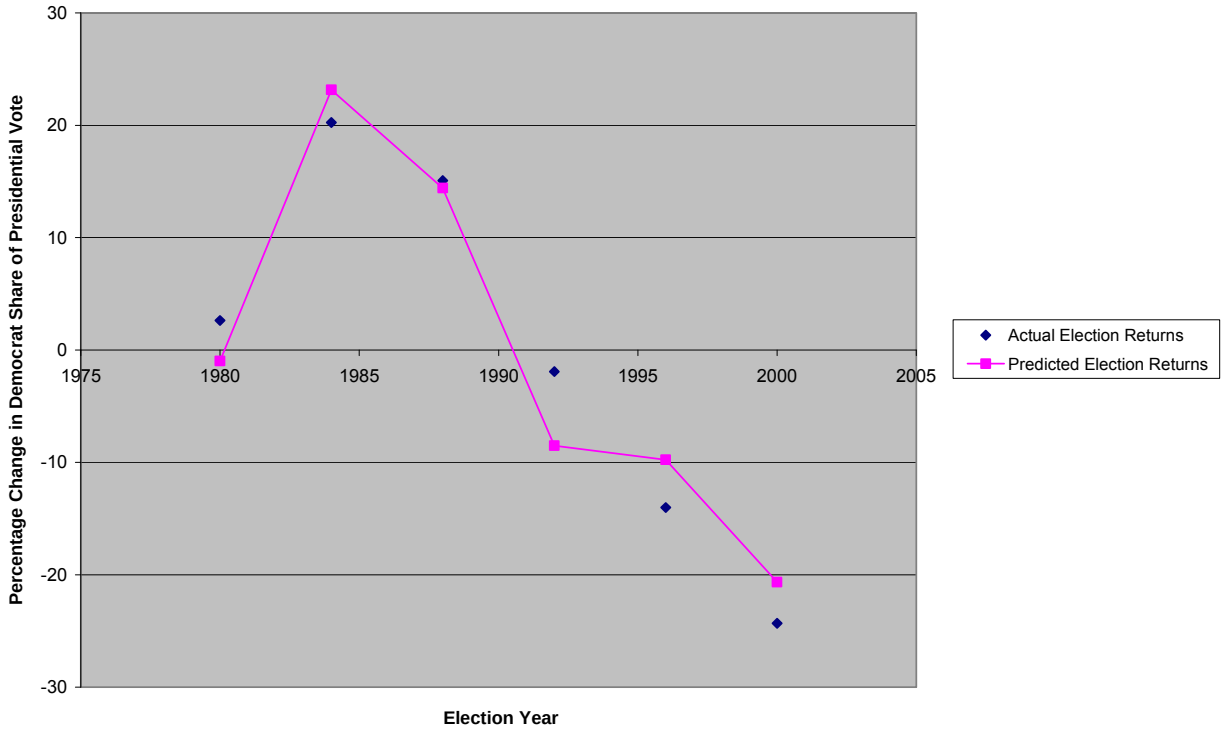


Figure 7.4: Predictive Capability in the Mountain Region



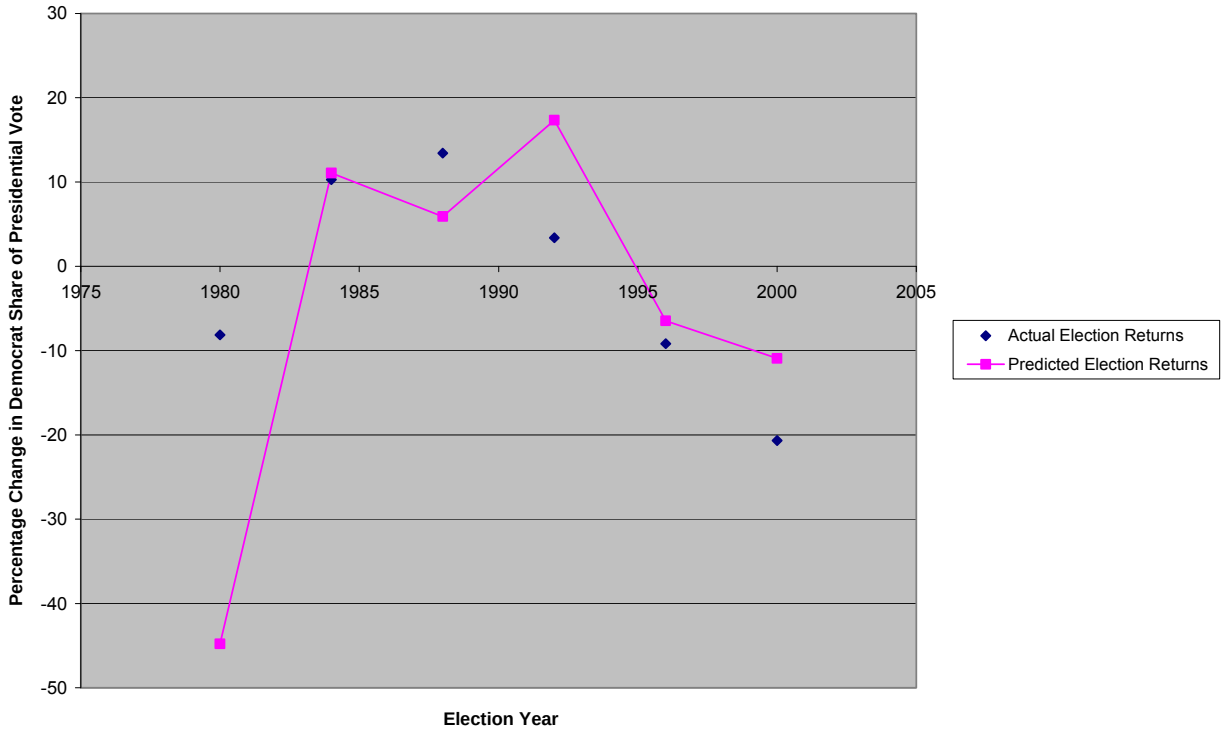
(7.2) A Discussion of Dataset Best Fit

Interestingly, it appears that several southern regions data actually fits its regression exceptionally well. In general, the two southern Census regions have data that fits their regressions better than other regions of the United States per the coefficient of determination. Two of the three best "fitting" regions are the East South Central and West South Central regions. The other regions with the best coefficients of determination are New England, East North Central Midwest, and the Mountain regions.

However, dataset best fit – simply the coefficient of determination being as close to 1.000 as possible – is not a fully accurate indicator of how well the regressions presented in this study actually establish correlation between economic indicators and election results. While this is a point that was mentioned in passing throughout Chapter 5.1, it is a point that deserves to be

underscored. Several of the regressions presented in this study have had superb R^2 values – 0.970, and 0.948 for the West South Central and New England regions, respectively. As demonstrated in Chapter 7.1, several of these regions do hold a tremendously close relationship in their predicted values to actual election results. However, regions such as the East North Central Midwest region – with the best R^2 of all regions of 0.977, still indicating excellent fit – fails to accurately predict election results.

Figure 7.5: Predictive Capability in the East North Central Midwest Region



As demonstrated in Figure 7.5, it has a relatively poor predictive capability compared to Figures 7.1-7.4 – this is due to the lack of statistical significance in correlation between the economic indicators and electoral results. This case is a prime example of how an over-reliance on coefficients of determination should not be used in analysis. Data simply fitting a model well is an exceptionally poor indicator of predictive capacity.

Indeed, Figure 7.2 is an excellent example of how even a relatively poorer "fitting" dataset can actually produce a significantly more robust model. The West North Central Midwest region, as modeled by Figure 7.2, has an R^2 of only 0.750 as compared to the high value of the Mountain region at 0.977. However, Figure 7.2 much more accurately models the correlations between per capita income, unemployment, and average wage to electoral outcomes than the relationship demonstrated in Figure 7.5.

8. Summary of Major Findings

Several important conclusions can be drawn from the data and analysis presented in this study. Not only did this study present a novel method of analysis – the introduction of a local universe of analysis and regional variation in Kramer's 1971 model – but it also has worked to establish a rigorous framework for analysis of local economic and electoral data.

The first topic that this study sought to analyze was the relationship between economic indicators and electoral outcomes. The three independent variables studied were the economic indicators of county-level average wage, county-level per capita income, and state-level unemployment rate. Of these three economic indicators, per capita income held the strongest correlation to the county-level share of the Presidential vote that incumbency-weighted Democrat candidates received. Per capita income proved to have a robust correlation to incumbency-weighted electoral results in virtually all regions.

The second topic of study was the degree of regional difference between economic indicators and electoral outcomes in the United States. Per the results presented in Chapter 5, it appears that there is a significant regional variation among the various Census Bureau defined regions in the United States. While per capita income had a statistically significant relationship to electoral outcomes in virtually all regions, average wage and unemployment's significance to election returns varied widely across regions. Moreover, the predictive capacity – a reflection of the robustness of regression, and therefore a reflection of the relationship between the chosen economic indicators and election returns – varied significantly depending on region. There, in fact, was a *regional difference* in this analysis. Figures 7.2 through 7.4 – all with excellent predictive capacities – can be compared to Figure 7.5 to clearly demonstrate this difference in correlation across regions.

However, it is important to not take these findings out of context. This study studied the relationship between local economic conditions and election results almost in isolation. With the exception of an incumbency weight, demographics and party affiliation were not really considered. These factors were considered as *implicit* in the regional differences among the various regions of the United States as discussed in Chapters 1 and 4. Some of the non-measured factors were also slightly mitigated by using a year-over-year percentage change methodology for all indicators. Therefore, factors that were changing each year were studied. Regardless, this model should not be considered a definitive study of the relationship between economics and politics. It can certainly be expanded upon – several suggestions are provided in Chapter 9.

9. Recommendations for Future Research

As partially addressed in previous chapters, there are several aspects of this study that can – and should – be further expanded upon or corrected by future researchers. First, the statistical methodology of this study particularly lends itself to improvement. This study uses relatively straightforward regression statistics in order to establish correlations between economic indicators and electoral outcomes. This is a process that can be *vastly* improved with a greater use of more sophisticated statistical models. The model presented in this study could be significantly improved by simply adding more indicators or performing regression in a different manner. Correlation in this study was tested using t-tests for significance. Various other statistical measures for testing correlation exist and could possibly improve the results obtained in this study.

On a more fundamental level, future methods for abstracting particulate data should also be developed. One of the difficulties of this study, as discussed in Chapter 6, was abstracting local data onto a regional level while minimizing the introduction of major systemic error. Essentially, when indicators are not measured well, attenuation in correlation occurs. While several sets of correlation have been established in this study – the results in Chapter 5 clearly demonstrate this – perhaps these correlations can prove to be more robust if more effective methods of abstraction are used. Certainly, error was introduced in this study, as is the case in virtually all studies – it is impossible to abstract such a large dataset onto a regional level and perform regression analysis without doing introducing some error. However, great care and many trials with varying universe of data were performed in order to maintain the integrity of this study. Regardless, the future development of methods to successfully abstract local data

while minimizing systemic error would greatly enhance the context and validity of correlations presented.

Moreover, one of the most enjoyable aspects of performing such analysis was modeling this method of analysis from Kramer 1971. Several formative political scientists have furthered his analysis – each investigator adding his or her unique twist to the method of analysis – since the early 1970s. In fact, the latest study involving this line of research actually occurred as late as 2006. The fundamental refinement of this almost four-decade old model represents an evolution of knowledge – an evolution of how far political science has come from the simple level of qualitative analysis to something more rigorous. Kramer's work has inspired many political science researchers in subsequent years since the publication of his 1971 work. However, the pursuit has remained the same – an attempt to apply a quantitative, rigorous framework to the relationship between economics and politics. Future studies with a greater number of economic indicators, regional, social, and demographical trends *must* be performed in order to maintain progress within this line of study. Not only is this information useful on a practical level – understanding electoral outcomes given temporal changes in the economy is absolutely useful for those attempting to project the effects of economic changes – but it is also imperative on a essential level of why voters vote for the leaders that they do. Our society is predicated upon the promise that individual voters can exercise their rights to create fundamental change every four years – if anything, further research is required to simply expand our knowledge on this basic fact.

10. Conclusion

Over the course of this study, a repeated emphasis on methodology, regional biases, and the importance of quantitative analysis has been stated. While certainly there are several significant externalities that must be accounted for, it is absolutely clear that there is a local correlation between economic indicators and electoral outcomes and there is also significant regional variation among various areas of the United States when evaluating this correlation. Such an assertion is significant due to the quantitative research performed on this subject to date – simply, prior to this study there has been a dearth of research on regional variability and local analysis when evaluating economic and political relationships.

It is also imperative to note that there are significant correlation differences present among the three economic indicators evaluated – the data analysis performed in this study indicates that per capita income has the most robust correlation to electoral outcomes. Much as Kramer was able to demonstrate in his 1971 study, per capita income proves to have the strongest driver of electoral outcomes in this model – especially on a regional basis. In eight of the nine evaluated regions, per capita income had the strongest statistically significant correlation to the Democratic share of the Presidential vote in the time-series data used. Per this same analysis, the unemployment rate held the least amount of statistically correlative significance, and the average wage indicator studied varied from region to region.

One of the strengths that regression analysis provides is an ability to have some sense of prospective, predictive capability. The analysis presented in this study clearly demonstrates that the models developed do, in fact, prove to have some level of accurate predictive power – the comparison of historical election results to predicted results outlines this relationship clearly. However, as with any quantitative study, the assumptions underlying this model must be clearly

recognized. Moreover, in order for any concrete predictive power to hold true, advanced statistics must be used to strengthen the models presented in this study. Clearly, several other indicators, measurements, and statistics can at one day be applied to this model to increase the accuracy of predictions.

More generally, a major implicit conclusion of this study relates to the overarching initial sentiment addressed in the introduction of this presentation – broad voter choice. A driving principle of this study – as many in the political science research realm – was to better understand broad voter choice in elections. Essentially, this study hoped to shed some light on the question of why voters vote the way the do. What is absolutely clear after the analysis performed is that there is *some level* of rationale behind how voters make choices for an economically motivated voter. The introduction of incumbency – and the strong predictive power of voters voting *against* incumbents when economic downturns occur – robustly supports this point. This sentiment, coupled with the more specific regional conclusions that were identified, actually provide a significant framework for future analysis.

Literature Cited

Ansolabehere, S., Rodden, J., Snyder, J. M. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *The American Political Science Review* 102(2): 215-232.

Bloom, Howard S., and H. Douglas Price. 1975. "Voter Response to Short-Run Economic Conditions: The Asymmetric Effect of Prosperity and Recession." *The American Political Science Review* 69(4): 1240-1254.

Cukierman, Alex and Francesco Lippi. 1999. " Central bank independence, centralization of wage bargaining, inflation and unemployment: Theory and some evidence." *European Economic Review* 43: 1395-1434.

Kiewiet, D.R., and D. Rivers. 1984. "A retrospective on retrospective voting." *Political Behavior* 6(4): 369–393.

Kiewiet, Roderick D., and Michael Udell. 1998. "Twenty-Five Years After Kramer: An Assessment of Economic Retrospective Voting Based Upon Improved Estimates of Income and Unemployment." *Economics and Politics* 10(3): 219-248.

Klorman, R. 1978. "Trend in personal finances and the vote." *Public Opinion Quarterly* 42(1): 31-48.

Kramer, G.H. 1971. "Short-term fluctuations in US voting behavior, 1896-1964." *The American Political Science Review* 65(1): 131–143.

Kramer, Gerald H., and Alvin K Klevorick. 1974. "Existence of a "Local" Co-operative Equilibrium in a Class of Voting Games." *Review Literature And Arts Of The Americas* 41(4): 539-547.

Markus, Gregory B. 1988. "The Impact of Personal and National Economic Conditions on the Presidential Vote: A Pooled Cross-Sectional Analysis." *American Journal of Political Science* 32(1): 137-154.

Monroe, K. R. 1979. "Econometric Analyses of Electoral Behavior: A Critical Review." *Political Behavior* 1(2): 137-173.

Piereson, J. E. 1977." District economic conditions and congressional elections." In *Paper Presented at Annual Meeting of the Midwest Political Science Association*, Chicago.

Pollard, W. 1978." Effects of economic conditions on elections a study controlling for political variables." In *Paper presented at the Meetings of the Public Choice Society*, New Orleans.

Radcliff, Benjamin. 1988. "Solving a Puzzle: Aggregate Analysis and Economic Voting Revisited." *The Journal of Politics* 50(2): 440-455.

Stigler, George J. 1973. "General Economic Conditions and National Elections." *The American Economic Review* 63(2): 160-167.

Wides, J. W. 1976. "Self-perceived economic change and political orientations: a preliminary exploration." *American Politics Quarterly* 4(October): 395-411.

Yoon, David Hyung-Jin. 2006. "Political Self-Interest and the Economic Vote."

Appendix A

The following sources were used for primary data as presented in this study:

Average wage: Bureau of Economic Analysis, United States Department of Commerce, Local Area Personal Income.

Per capita income: Bureau of Economic Analysis, United States Department of Commerce, Local Area Personal Income.

Unemployment rate: Bureau of Labor Statistics, United States Department of Labor, Local Area Unemployment Statistics.

Election results: CenStats – USA Counties, United States Census Bureau