Public Finance Review
Volume 35 Number 4
July 2007 480–499

10.1177/1091142106295768
http://pfr.sagepub.com
hosted at
http://online.sagepub.com

Accounting for Spatial Error Correlation in the 2004 Presidential Popular Vote

Donald J. Lacombe
Ohio University

Timothy M. Shaughnessy
LSU in Shreveport

One problem with describing election vote shares using ordinary least squares (OLS) is that it ignores the possible presence of spatial error correlation, whereby the errors are correlated in a systematic manner over space. This omission can bias OLS standard errors. We examine the 2004 presidential county vote outcome using OLS and a spatial error model (SEM) that accounts for spatial autocorrelation in the error structure. We find that spatial error correlation is present, that the SEM is superior to OLS for making inferences, and that several factors deemed important to the 2004 election outcome are not significant once the spatial error autocorrelation is taken into account.

Keywords: *spatial econometrics; spatial error model; presidential election*

1. Introduction

A November 5, 2004 online column at Salon.com asks "Did this man cost the Democrats the election?" below a picture of San Francisco mayor Gavin Newsom, who allowed gay marriages in City Hall prior to the presidential election.[1] Guest columnist for the *Seattle-Post Intelligencer* Anthony B. Robinson wrote on November 14, 2004, that "there was a kind of 'in your face' quality to Newsom's opening City Hall for gay marriages that ran ahead of where many people were. The resulting furor was clearly a factor in the 11 states that passed anti-gay marriage bills."[2] However, Slate.com's Paul Freedman posted a column on November 5, 2004, that claims "Gay marriage and values didn't decide this election. Terrorism

Authors' Note: The authors gratefully acknowledge helpful comments and suggestions from Roy Boyd, Tony Caporale, Randall Holcombe, and two anonymous referees.

did."[3] CNN's national election exit poll reports that, of seven possible issues influencing people's presidential vote, moral values was cited as the most important.[4]

As in any election, there were a number of hot-button issues that political pundits cited as determining or influencing the outcome of the 2004 presidential race. To the extent that these issues can be quantified, researchers can implement these policy measures into standard political econometric models to assess their significance. What is frequently omitted, however, is the importance that geographic space and location dependency may or may not have on election outcomes or voting patterns. Unobserved effects that influence two adjacent or nearby voting districts may not be captured in standard econometric specifications, leading to incorrect inferences being drawn from the results.

The problem is that unobserved effects are usually difficult to quantify, though they nonetheless may be quite important in influencing the general political leanings of a neighborhood, city, or county. To the extent possible, most voting studies capture economic, education, demographic, and possibly religious variables believed important in determining an area's presidential choice, but other recognizable factors remain unmeasured. Unobservable location-specific factors will generally reflect a common set of values held by the community, which may be a result of a shared background, access to only a few sources of news media, the prominence of one or two industries in which a large portion of the community is employed, or other similar influences. The benefit of a spatial econometric analysis is that it allows us to quantify the aggregate effect of these location-specific factors, and thus to recognize and disentangle formerly unobservable effects that may corrupt standard ordinary least squares (OLS) results.

This paper has two aims: to assess the influence that spatial effects have on the voting patterns of counties in the 2004 presidential election, and to compare the factors popularly believed to have been important in that election to the results of an econometric specification that accurately accounts for these spatial effects. In particular, we will see that a typical OLS model describing the percentage of a county voting for George W. Bush yields quite different results from a model describing the county voting that explicitly tests and accounts for spatial error autocorrelation.

The rest of the paper proceeds as follows: Section 2 reviews a sample of recent literature on voting models that are relevant to our discussion of geographic effects. Section 3 describes the empirical model, with hypotheses about the included variables and the data used in our estimations. Section 4 describes the baseline results from a voting model using OLS that does not account for spatial effects, and section 5 reviews the results from the same model using a spatial error model (SEM) to account for spatial autocorrelation. Section 6 concludes.

2. Literature Review

Several papers have examined elections with cursory attempts to account for the spatial location of voters. Admittedly, the theses being tested in these papers is generally not the effect of location per se, but a review of them provides adequate direction on model specification.

The work by Fair has been influential for modeling the economic factors affecting presidential elections. Fair (1996) provides a good review of his previous work. His estimated equation describes the percentage of the national popular vote for the Democratic candidate as a function of a time trend, incumbency variables, growth rate of real per capita GDP, and the absolute value of the inflation rate. Given that a new data point only becomes available every four years, he is limited to a small number of observations. Fair (1996) summarizes results using data from elections beginning in 1916, and the 1992 update only yields 20 observations. Prediction errors from estimated equations were quite small. Fair is able to combine a relatively small number of explanatory variables into a good predictive and forecasting model of national presidential elections, and we use his insights in adapting a model to describe presidential election results at smaller levels of aggregation, namely county votes. The obvious benefit of using county vote results, rather than national vote results, is the much greater number of observations used.

Blackley and Shepard (1994) and Abrams and Butkiewicz (1995) both examine the importance of state-level economic conditions in determining state presidential vote outcomes. Blackley and Shepard use weighted least squares to predict the state vote percentages for both George H. W. Bush and Bill Clinton in the 1992 election, using state unemployment, real state per capita income, variables representing Bush's state vote share in the 1988 election, the change in the number of voters, dummy variables for president and vice president home states, and dummy variables for three regions in the United States. The state-level unemployment and per capita income variables are significant in all their models. Even with a relatively small set of control variables for demographic or economic characteristics, they appear to obtain significant results. Abrams and Butkiewicz also examine the 1992 election for evidence of state-level economic effects. They use both OLS and WLS and obtain similar results for each method. Their state-level variables include real state per capita income over the four years prior to the election, unexpected growth in real per capita income over the prior four years, growth in real per capita income one year prior to the election, and two measures of state unemployment. They also include the state percentage vote for Bush in the 1988 election and a dummy variable for Clinton's home state of Arkansas. Again, a small set of control variables seem to result in models of good fit with expected signs, and support the hypothesis that state-level economic conditions significantly affect state presidential voting outcomes.

Other articles focus more specifically on the influences on presidential elections while not necessarily including geographic controls. Kan and Yang (2001) measure the significance of expressive versus instrumental voting by examining the American National Election Study (ANES) of the 1988 election. In describing the factors influencing both individual voter turnout and votes for Bush, the authors include race, income, sex, age, education, and marriage variables as well as variables measuring political interest or beliefs.

Eisenberg and Ketcham (2004) examine presidential elections using county-level data and the influence of macroeconomic conditions on voting. One conclusion at which they arrive is that county-level economic conditions are significant but less important in terms of a party's vote shares than state-level conditions, and both state and county are much less important than national conditions. They find that voter support for a candidate or party is influenced by local and national economic conditions, but that voter perceptions of national conditions are based on a shorter period of time versus local conditions. They also find that Blacks, the elderly, and men are more, less, and less influenced, respectively, by local economic conditions in deciding for whom to vote. Their results indicate that county-level conditions do not appear to be very important in influencing votes, results that will be interesting to reconsider when we explicitly test for spatial autocorrelation while also using county-level data.

Alvarez and Nagler (1997) assess the relative importance of economic conditions and candidate stances on political and social issues (basically, noneconomic factors) in determining voter choice. They use the ANES 1992 survey data, so are again testing the effects of individual-level measures on votes. Their results support those of Eisenberg and Ketcham

(2004) that economic conditions tend to outweigh other political or ideological considerations in determining a voter's choice.

One important point to make about the existing literature is that, while there has been important progress in identifying characteristics that influence election outcomes, there is no attempt to account for omitted factors that vary geographically over space, namely spatial error correlation. The effect of omitting spatial factors can be incorrect or misleading inferences from the results of standard OLS specifications.

3. Model and Data

We describe an individual voter's choice for president in 2004 as being influenced by demographic, economic, and political factors. Voters for Bush should have identifiable demographic, economic, and political characteristics that distinguish them from voters for John Kerry. To the extent that these individual preferences manifest themselves in observable data, the popular vote percentages for each candidate in each county should be influenced by the same factors. Many of the control variables we include and discuss below are used frequently in studies of voting outcomes. Our research will also examine the extent to which these variables may influence the vote percentages of nearby counties as well.

The demographic characteristics we included represent the county's racial composition, percent urbanization, education level, and religious climate. The economic environment is represented by including the county's per capita income and unemployment rate.[5] There were a couple of identifiably important political issues during the 2004 presidential election. The first was the controversy over the Iraq War, and we include a proxy that we believe is consistent in measuring county sentiment supporting the Iraq War. A second issue was the appearance of gay marriage amendments on many state ballots, and we include dummies for those states in which such an amendment was voted upon during the 2004 general election. We also presume that the two candidates should enjoy unusual support in their home states, and therefore we include a dummy variable for both Texas and Massachusetts.

The election under consideration is the 2004 presidential election, and we focus on the individual county popular votes. The often-shown "red/blue" county map is usually used to demonstrate strong county-level support for Bush over Kerry despite the much closer national popular vote or Electoral College vote. The red/blue map also motivates interest in the spatial component of county voting, since the map appears to demonstrate a nonrandom distribution of election outcomes. Red counties are likely to be surrounded by red counties and vice versa. This nonrandom grouping of red/blue counties indicates that there is possible spatial error correlation in that unobserved factors that vary over space need to be accounted for in a systematic manner. There may be local customs or historical reasons that certain areas of the country vote in a certain manner that is ignored in the existing literature. For example, the so-called "Rust Belt" that extends from roughly Chicago to New York City in the upper portion of the central United States has traditionally voted for the Democratic candidate based on that party's stance on such issues as trade and labor policies, such as the minimum wage. This geographic issue has normally been dealt with by using dummy variables for certain regions (e.g., Blackley and Shepard 1994) but the dummy variable approach does not properly model the geographic interaction among the geographic entities. The spatial econometric techniques advocated here have the potential to properly model these interactions relative to an ad hoc dummy variable specification.

The dependent variable represents the percent of the county popular vote going to George Bush, and the data come from POLIDATA Demographic & Political Guides.[6] The sample includes 3,065 counties and excludes Alaska, Hawaii, Washington, D.C., and all cities in Virginia.[7] Because of third-party candidates, one minus the dependent variable does not necessarily represent the percentage of the popular vote for Kerry, and we ignore the effects of third parties in our subsequent analysis.

The proxy that we use to express support for the Iraq War is the percentage of the civilian population age 18 or over who are civilian veterans. These data come from Census 2000 information via the Department of Veteran Affairs, which collected data on the veteran population in the United States and Puerto Rico sorted by county and by period of service.[8] Veterans on average tend to support the Republican party on issues of national security and the military, especially the recent Iraq War. Therefore, our hypothesis is that a higher percentage veteran population should be positively correlated with President Bush's popular vote total, and therefore, we expect a positive sign for this coefficient estimate.

White, Black, and Hispanic indicate the percentage of a county's population of the respective racial category. These data came from the Census 2000 Summary File (SF) 1 database. Hypotheses regarding which candidate or party-specific racial groups will support are difficult to elucidate so we therefore simply point out some plausible, but not definitive, reasons why certain racial groups may vote for a particular candidate in the presidential election.

Black voters have generally supported the goals of the Democratic party; CNN's national exit poll of 13,660 respondents indicates that 88 percent of Black respondents supported Kerry compared to 11 percent for Bush.[9] However, polls are an imperfect indicator of individual voter preference; further, there was much debate during this election about the share of the Black and Hispanic votes that the two candidates would achieve. The Republican party was perceived to have devoted considerably more of its efforts to courting these two groups than it had in the past. Nonetheless, given their historical support, we hypothesize that Black voters will tend to favor Kerry over Bush and that its coefficient will be negative. Hispanic voters can be hypothesized to vote either for or against Bush. Hispanic voters tend to be religious (specifically Catholic) and may favor the social policies of the Republican party and therefore support Bush. Additionally, both President Bush and his brother Jeb Bush, the Governor of Florida, speak Spanish, which may have some cultural appeal to Hispanic voters. However, Hispanics as a group are, on average, at a lower socio-economic scale relative to White voters and may therefore support the economic policies of the Democratic party. We therefore will let the econometric results speak for themselves regarding the coefficient estimate (CNN's exit poll reports 44 percent of Latino respondents voted for Bush and 53 percent for Kerry). White voters are a politically diverse group, comprised of Republicans and Democrats, as well as third-party (e.g., Green Party) supporters. Given the diversity of these voters, we do not make any a priori hypotheses about their support (or lack thereof) for Bush (CNN's exit poll reports 58 percent for Bush and 41 percent for Kerry).[10]

One of the conclusions from Abrams and Butkiewicz (1995) and Blackley and Shepard (1994) is that local economic conditions are important determinants in presidential elections. We include two variables to capture the economic environment. Per capita income represents the log of county per capita personal income in 2004, which comes from the Bureau of Economic Analysis' Regional Economic Accounts. Higher-income individuals are popularly believed to be supporters of the Republican party, since Republicans reputedly back policies that favor higher-income individuals. CNN's exit poll on the vote by income supports this view. On the other hand, the stereotypical "limousine liberal" effect may be relevant when higher income individuals vote for the party that will serve the interests of poorer voters, typically the Democratic party. We once again do not have any a priori hypothesis as to the effect of income on the popular vote share for Bush.

A second economic variable is the county's 2004 unemployment rate. These data come from the Bureau of Labor Statistics Local Area Unemployment database. Given that Bush was the incumbent in 2004 and that voters typically credit or blame the president for business conditions, a bad job market will likely lead voters to the alternative. Thus, we expect that a higher county unemployment rate is associated with a lower vote share for Bush.

Also included in our regression model is the percentage of the county population living in urban areas, obtained from the 2000 Census SF1 database. Urban residents are believed to favor the Democratic party and we therefore expect that the urban coefficient will have a negative sign. Again, CNN's exit poll of the vote by size of community supports this view.

We constructed two variables on likely voters in each county, the percentage male and the percentage female holding bachelor's degrees, to test for a gender difference in voting outcomes. One variable is the percentage of a county's total male 25-and-over population who hold a bachelor's degree. We constructed a variable similarly for females in the county. These data also come from the 2000 Census SF3 database. Men tend to support the policies of the Republican party on issues such as national security and economic policy, so we hypothesize that this variable will be positively correlated with support for President Bush. On the other hand, women tend to identify with the policies of the Democratic candidate, namely education and social issues. We therefore hypothesize that this variable will be negatively associated with the popular vote total for Bush. The CNN national exit poll from 2000 shows stronger party division among males (55 percent for Bush versus 44 percent for Kerry) than among females (48 percent for Bush versus 51 percent for Kerry).

It is popularly believed that religious voters played an important role in the 2004 election, so we collected two variables measuring religiosity in counties. The religion variables come from the Glenmary Research Center of the Glenmary Home Missioners.[11] The Research Center reports data from a study by the Association of Statisticians of American Religious Bodies, who surveyed 149 religious groups on different aspects of their membership counts. One variable is the percentage of the total population in the county who are considered adherents of one of the religious groups. An adherent is a member of the religious body, their children, or a participant who is not a member. The other variable is the number of churches per 10,000 people in the total population in 2000. The Republican party is generally believed to cater to the "Religious Right," and we hypothesize that increases in the number of religious adherents as well as the number of churches will be positively correlated with the popular vote share for Bush.

The influence of religious voters is also believed to have been augmented due to the number of gay marriage constitutional amendments on state ballots that year. We use a dummy variable that equals one if, in the 2004 election, the state in which that county resides had a popular vote on a state constitutional amendment to ban gay marriage. The data came from CNN's 2004 Election website on state ballot measures.[12] The gay marriage issue tends to elicit strong conservative opinion, so we hypothesize that this variable will be positively correlated with the popular vote share for Bush.

Finally, we also include in our estimating equations dummy variables for counties that reside in the states of Texas and Massachusetts to capture any "home state" effects. Several of the papers we reviewed included similar controls.[13] We expect the sign of the Texas dummy variable to be positive and the Massachusetts dummy variable to be negative. Descriptive statistics of the data used and their sources are in table 1.

It is possible that some omitted unobservable factors that vary systematically over geographic space will result in residual spatial autocorrelation. If this spatial autocorrelation is ignored, the inferences drawn from the OLS results may be misleading, due to the fact that the OLS standard errors will have downward bias (Barry, Pace, and Sirmans 1998). In light of this possible econometric problem, we use spatial econometric techniques to correct for the residual spatial autocorrelation.

The conditions under which spatial residual autocorrelation arise are nicely illustrated in a housing context by Dubin (1998): "Housing prices are a prime example: clearly the location of the house will have an effect on its selling price. If the location of the house influences its price, then the possibility arises that nearby houses will be affected by the same location factors. Any error in measuring these factors will cause their error terms to be correlated." A similar phenomenon can occur in the context of a voting model: it may be that our unobserved factors used to explain the vote shares across counties are the factors mentioned earlier (e.g., access to media markets) and that these unobserved factors are geographically correlated, resulting in residual spatial autocorrelation.

In order to account for the possibility of residual spatial autocorrelation, we use a spatial error model (SEM). A justification for using the SEM is provided by Anselin (1988), who states that "The linear regression model with spatially autoregressive errors is by far the most relevant spatial specification for applied empirical work on cross-sectional data."

Table 1
Summary Statistics

	Mean	Standard Deviation	Source
% vote for Bush	0.605	0.124	Polidata.org
% Black	8.622	14.370	Census 2000 SF1
% White	85.001	15.891	Census 2000 SF1
% Hispanic	6.223	12.112	Census 2000 SF1
% urban population	0.393	0.304	Census 2000 SF1
Female population with bachelor's	0.109	0.049	Census 2000 SF3
Male population with bachelor's	0.109	0.051	Census 2000 SF3
Gay marriage dummy	0.290	0.449	CNN 2004 Election website
% veterans	0.140	0.029	Census, Dept. of Veteran Affairs
% religious adherents	0.531	0.183	Glenmary Research Center
Churches per 10,000 people	22.143	13.145	Glenmary Research Center
Log per-capita personal income	10.152	0.004	Bureau of Economic Analysis
Unemployment rate	5.664	0.032	Bureau of Labor Statistics

Other researchers have successfully used spatial econometric techniques to address spatial autocorrelation. For example, Brasington (2004) uses spatial econometric techniques to examine a hedonic house price equation to determine the effect of school district consolidation. Brasington and Hite (2005) use spatial econometric techniques to estimate the effect of point-source pollutants on housing prices. Cohen (2002) uses spatial econometric techniques similar to the ones used in this paper to examine how states and localities react to federal airport grant cuts. Finally, Garrett and Marsh (2002) use spatial econometric techniques to analyze cross-border lottery shopping.[14]

Formally, the SEM model can be expressed as follows:

$$y = X\beta + u \tag{1}$$

$$u = \lambda W u + \varepsilon \tag{2}$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

where y is the dependent variable, X is a vector of independent variables, and ε is the error term. The scalar term λ is the spatial error parameter to be estimated. It measures the degree of spatial dependence among the residuals. The W term is an n by n first-order spatial contiguity weight matrix "which expresses for each observation (row) those locations (columns) that belong to its neighborhood set as nonzero elements" (Anselin and

Bera 1998, p. 243).[15] Normally, a row stochastic weight matrix is used in a regression modeling context, which means that the rows of the spatial weight matrix sum to unity. This transformation of the spatial weight matrix provides for an intuitive explanation for the λW term. The λW term can be thought of as a weighted average of the surrounding error terms. The use of OLS to estimate the above model results in a multitude of econometric problems and therefore maximum likelihood methods are used.[16]

4. Baseline OLS Results

Most studies that attempt to explain the outcomes of presidential elections use ordinary least squares (OLS) regression techniques (or variants thereof, such as weighted least squares) that regress a matrix of independent variables against a dependent variable that represents the popular vote percentage, or some other election outcome metric. Our initial estimate uses standard OLS techniques, and constitutes our baseline regression model with which our spatial econometric model will be compared. The estimating equation that we use is represented by the following:

$$\begin{aligned} Bush = {} & constant + \beta1^{*} Black + \beta2^{*} White + \beta3^{*} Hispanic \\ & + \beta4^{*} UrbanPop + \beta5^{*} Female\ w/Bachelor's \\ & + \beta6^{*} Male\ w/Bachelor's + \beta7^{*} GayMarriage \\ & + \beta8^{*} Veterans + \beta9^{*} Adherents + \beta10^{*} Churches \\ & + \beta11^{*} Texas + \beta12^{*} Massachusetts \\ & + \beta13^{*} Per\ Capita\ Income + \beta14^{*} Unemployment + \varepsilon, \end{aligned} \tag{3}$$

where Bush is the percentage of the county popular vote received by George W. Bush in the 2004 presidential election, $\beta1 - \beta14$ represent the coefficients for the independent variables explained in the previous section, and ε represents an i.i.d. normal error term.

The results from the OLS specification can be divided into the three different categories of variables used: demographic, political, and economic.[17] In the demographic category (i.e., Black, White, Hispanic, Urban, Females with Bachelor's, and Males with Bachelor's), all the variables have the expected signs with some noteworthy exceptions. The White variable is statistically significant at the 1 percent level and has a positive coefficient estimate, indicating that White voters were more likely to vote for Bush. The percent Hispanic did not seem to affect the vote total for Bush, as this variable is insignificant at conventional levels. The surprising result, however, is that the percent Black also seems to not have any statistically significant effect on the popular vote share for Bush, despite the earlier reference to their overwhelming support for Kerry.

For the other demographic variables, results for the two categories of educated individuals offer no surprises in terms of the signs and significance of the coefficient estimates: men with bachelor's degrees favor Bush and their respective female counterparts do not favor Bush. Both variables are statistically significant at the 1 percent level and the magnitudes of the coefficient estimates demonstrate that females vote more consistently against Bush than males vote for him. Finally, the coefficient estimate for urban voters is negative, but is not statistically significant.

The next category of independent variables is the political variables, which include the gay marriage, veterans, religious adherents, number of churches, and the Texas and Massachusetts dummy variables. As mentioned earlier, one of the more contentious aspects of the 2004 presidential election was the effect of the so-called "gay marriage" initiatives on state ballots. The presence of these state ballot initiatives is highly statistically significant at the 1 percent level and has a positive impact on the vote share for Bush, albeit small, with a coefficient estimate of 0.02. As expected, veterans supported Bush, with the coefficient estimate positive and statistically significant at the 1 percent level. The effects of religious sentiment are captured by the adherents and churches variables; both coefficients are positive and churches is statistically significant at the 1 percent level, though adherents is not significant. Not surprisingly, the state dummy variables predict that Texans support Bush and Massachusetts residents do not support Bush; both variables have the expected signs and are statistically significant at the 1 percent level. The Massachusetts effect is slightly stronger than the Texas effect.

Our final category of independent variables is the economic variables, which include the log of per capita personal income and the unemployment rate. Higher per capita personal income is associated with lower vote totals for Bush and is highly significant, which supports the hypothesis that higher income individuals did not support Bush. Further, the unemployment rate also has a negative coefficient estimate and is highly significant.

Our baseline OLS specification indicates that most of our hypotheses are reasonable. The only surprise is the insignificant Black variable, given the Black population's propensity to overwhelmingly support the Democratic party candidate. The model fit is also fairly good, where this specific set of independent variables explains approximately 42 percent of the variation in the popular vote for Bush.

5. Spatial Econometric Model Estimates

We now employ the spatial error model (SEM) to determine if correcting for spatial residual autocorrelation changes any of the inferences regarding our independent variables. As noted by Dubin (1998), "The consequences of spatial autocorrelation are the same as those of time series autocorrelation: the OLS estimators are unbiased but inefficient, and the estimates of the variance of the estimators are biased." Specifically, as Pace et al. (Barry, Pace, and Sirmans 1998) point out, "In the case of positive spatial autocorrelation, the OLS standard errors have a downward bias." Therefore, it is possible that independent variables that appear significant under the assumption of no spatial residual autocorrelation (by using standard OLS techniques) will not be significant once the spatial autocorrelation in the residuals are properly modeled.

Table 3 contains estimates for the following model, reproduced here for convenience:

$$\begin{aligned} Bush = {} & constant + \beta 1^* Black + \beta 2^* White + \beta 3^* Hispanic \\ & + \beta 4^* UrbanPop + \beta 5^* Female\ w/Bachelor's \\ & + \beta 6^* Male\ w/Bachelor's + \beta 7^* Gay\ Marriage \\ & + \beta 8^* Veterans + \beta 9^* Adherents + \beta 10^* Churches \\ & + \beta 11^* Texas + \beta 12^* Massachusetts \\ & + \beta 13^* Per\ Capita\ Income + \beta 14^* Unemployment + u \qquad (4) \end{aligned}$$

$$u = \lambda Wu + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n), \qquad (2)$$

where all variable definitions are identical to the OLS counterpart and we use maximum likelihood methods to estimate the SEM. The highly significant coefficient for λ demonstrates that the errors are spatially correlated, so it just remains to formally determine the appropriate spatial model to estimate.

Our choice of the SEM model is motivated by the use of several Lagrange Multiplier (LM) tests that have been developed by Anselin et al. (1996). Florax, Folmer, and Rey (2003) outline a procedure for testing the residuals of an OLS model using the following steps. First, the OLS model is tested for spatial dependence due to an omitted spatial lag or to spatially autoregressive errors using the nonrobust test statistics, LM Lag and LM

Error. Table 2 indicates that both of these tests are highly statistically significant. In this case, we next examine the robust varieties of the LM tests, which test for a spatial error process when the specification potentially contains a spatially lagged dependent variable (LM Error Robust) and alternatively, a test for a spatially lagged dependent variable in the potential presence of a spatial error process (LM Lag Robust). The results in table 2 indicate that both of the robust varieties are statistically significant, and in this case, Florax, Folmer, and Rey suggest estimating the specification indicated by the more significant of the two robust tests (i.e., whichever test statistic has greater value). Since our results indicate that the robust LM Error test has greater significance than the robust LM Lag test (635.10 vs. 4.86), we estimate the spatial error model (SEM).[18] One important difference to note is the large increase in explanatory power, as the SEM explains about 78 percent of the variation in Bush while the OLS model only explains about 42 percent.

We next examine our three different groups of independent variables and how these estimates compare to the model estimated via OLS. Results for the variables representing the percent White and percent Hispanic do not change from the OLS estimations: percent White is highly significant with a positive coefficient estimate and percent Hispanic is insignificant. The principal difference among the demographic variables between the OLS and SEM models is that the percent Black variable is now highly statistically significant at the 1 percent level and the coefficient estimate is negative, indicating that as the number of Black voters increases, support for Bush decreases, which is consistent with our hypothesized effect.

The SEM is consistent with our OLS specification indicating that urban population is insignificant. Our final group of demographic variables is the male and female measures of educational attainment. Under the OLS model, both variables are highly significant and conform to our hypothesized effects: females with bachelor's degrees do not favor Bush while their male counterparts do. In the SEM, however, the male measure of educational attainment drops in statistical significance from 1 percent to 10 percent. The female measure of educational attainment, however, remains highly significant at the 1 percent level. We conclude that for this particular election, women's preferences against Bush were much stronger than men's preferences for him (despite CNN's polling results discussed earlier).

The political variables used in our regression models tell different stories depending on which model is examined. The Texas state dummy and percent veterans variables have the same signs and significances in the

Table 2
OLS Results for Election Outcome

Independent Variables	Dependent Variable: Popular Vote Percentage for Bush by County	t-statistics
Constant	1.4827	8.86*
% Black	−0.0006	−1.47
% White	0.0016	4.30*
% Hispanic	−0.0002	−0.57
% urban population	−0.0025	−0.26
Female population with bachelor's	−1.1312	−10.13*
Male population with bachelor's	0.8016	7.09*
Gay marriage dummy	0.0221	5.61*
% veterans	0.2432	3.42*
% religious adherents	0.0150	1.07
Churches per 10,000 people	0.0017	7.06*
Texas state dummy	0.1171	15.22*
Massachusetts state dummy	−0.1690	−8.23*
Log per-capita personal income	−0.0901	−5.34*
Unemployment rate	−0.0262	−18.13*
F statistic	156.2463*	
Adjusted R^2	0.4150	
LM Lag	1897.96*	
LM Error	2528.20*	
LM Lag Robust	4.86**	
LM Error Robust	635.10*	

Note: The specification tested positive for heteroskedasticity and therefore White's standard errors are used.
*Significant at the 1% level. **Significant at the 5% level.

OLS and SEM models. The Massachusetts state dummy drops in significance from 1 percent to 5 percent.

The results for the religion variables differ between the OLS and SEM models. OLS gave the somewhat unexpected result that the percentage of a county's population who consider themselves religious adherents was insignificant, while the number of churches per 10,000 people was significant. Both were of expected signs, but the results seemed to indicate that what influenced support for Bush was the number of physical church structures instead of the proportion of a county's population who considered themselves religious. The SEM results are more plausible and consistent with our theoretical expectations. Again, both coefficient estimates

Table 3
Spatial Error Model (SEM) Results for Election Outcome

Independent Variables	Dependent Variable: Popular Vote Percentage for Bush by County	Asymptotic t-statistics
Constant	0.5823	5.27*
% Black	−0.0022	−8.92*
% White	0.0037	18.64*
% Hispanic	−0.0002	−0.95
% urban population	−0.0073	−1.26
Female population with bachelor's	−0.4436	−6.26*
Male population with bachelor's	0.1262	1.88***
Gay marriage dummy	0.0102	1.59
% veterans	0.2056	3.82*
% religious adherents	0.0267	2.71*
Churches per 10,000 people	0.0003	1.56
Texas state dummy	0.0610	3.79*
Massachusetts state dummy	−0.0564	−2.20**
Log per-capita personal income	−0.0235	−2.10**
Unemployment rate	−0.0096	−9.11*
Lambda	0.8810	84.29*
Log-likelihood	5098.9076	
Adjusted R^2	0.7775	

*Significant at the 1% level. **Significant at the 5% level. ***Significant at the 10% level.

are positive, but the percent adherents is now highly significant while the number of churches is insignificant. Our results indicate that it is more important how many people identify themselves as religious as opposed to the number of churches that are available for worship.

A second significant difference among the political variables between the OLS and SEM specifications is the gay marriage dummy variable. It is now statistically insignificant, which can be caused by two factors: (1) the standard errors of the OLS specification are biased or (2) it may be that accounting for the spatial nature of the data is picking up some unobserved ideological preferences and that the gay marriage dummy variable is a crude measure of this unobserved effect.

The final group of variables is the economic variables. Per capita income is negative and significant at the 1 percent level in the OLS model, but drops in significance in the SEM (this is consistent with Fair's national vote model, where Democratic votes were positively associated with the growth rate of per capita income). The unemployment variable remains negative and highly significant between both models, indicating that a worse labor market is associated with fewer votes for the incumbent Bush.

The results from the SEM specification lend support to the idea that county-level data are appropriate when estimating national voting models, a conclusion that contradicts the findings of Eisenberg and Ketcham (2004).

6. Conclusions

The purpose of this article is two-fold: to assess the level of spatial autocorrelation in a typical political model describing county voting patterns, and to test hypotheses about particular variables thought influential in the 2004 presidential election. As demonstrated, the analysis of election outcomes, or any econometric model that may contain geographic dependency among observations, may be misleading if spatial error autocorrelation is unaccounted for. The presence of spatial error autocorrelation in an OLS model yields standard errors that are biased downward, and the researcher may be led to overestimate the significance of particular variables.

In our OLS analysis of county presidential votes, we obtained two unexpected results: First, that the percent of a county's population that is Black was insignificant or only marginally significant in predicting the vote for Bush; and, second, that the number of churches and not the percent of a county's population that considered itself religious was significant in predicting votes for Bush. The OLS model also yielded results that were more consistent with our expectations: the percent urban, the percents male and female, and economic climate all were significant. These models also had decent predictive power, explaining over 40 percent of the variation in Bush.

However, the results from various Lagrange Multiplier tests indicated that the SEM is the appropriate model to use. Accounting for the spatial effects drastically improved the predictive power of our model, and also seemed to correct some of the downward bias of the standard errors expected in OLS estimation. Percent Black became much more significant and consistent with expectations, and a few other variables thought important in the 2004 election were judged to be relatively weak. Indeed, as mentioned in the introduction, numerous pundits were steadfast in their belief that the "Religious Right" were turned out in droves to support Bush as they voted for state gay marriage bans. Our results, however, lend only limited support to this idea; the presence of gay marriage amendments and the number of churches in a county are not significantly associated with Bush's vote totals, though the number of religious adherents did tend to boost Bush's vote.

As Anselin (1988) noted, cross-sectional data over geographic space can be susceptible to spatial error autocorrelation, and proper estimation procedures are necessary if accurate inferences from results are to be made. Researchers interested in obtaining accurate estimates may wish to explore the possibilities offered by spatial econometric techniques.

Notes

1. http://www.salon.com/opinion/feature/2004/11/05/gay_marriage/index_np.html.
2. http://seattlepi.nwsource.com/opinion/198889_focussub14.html.
3. http://slate.msn.com/id/2109275/.
4. http://www.cnn.com/ELECTION/2004/pages/results/states/US/P/00/epolls.0.html.
5. All variables are measured at the county level for the contiguous 48 states.
6. http://www.polidata.org/.
7. Technically, cities in Virginia do not reside in counties, so the city data were excluded. Only county-level data for Virginia are included. Alaska and Hawaii are excluded because of their lack of geographical proximity to the contiguous United States. This is standard practice in spatial econometric studies.
8. http://www1.va.gov/vetdata/page.cfm?pg = 1.
9. http://www.cnn.com/ELECTION/2004/pages/results/states/US/P/00/epolls.0.html.
10. There may appear to be a specification issue when we include controls for Black, White, and Hispanic, since the remainder is likely a very small percentage. However, we estimated models with all three racial categories and models with only Black and White, and the results were not demonstrably different from what we report here.
11. http://www.glenmary.org/grc/RCMS_2000/rankings.htm.
12. http://www.cnn.com/ELECTION/2004/pages/results/ballot.measures/.
13. For example, Abrams and Butkiewicz (1995) and Blackley and Shepard (1994).
14. For an excellent general introduction to spatial econometric techniques, see LeSage (1997).
15. An anonymous reviewer suggested that we use a spatial panel model with county fixed effects. However, we are unable to do so because county definitions have changed over time and the spatial panel methods require that the weight matrix be constant over time. One example is the creation of Cibola County, NM, which was created in part by taking land from Valencia County, NM on June 19, 1981. See http://www.census.gov/geo/www/tiger/ctychng.html for further examples.
16. A detailed proof of how spatial statistics achieve consistent and unbiased parameter estimates, unbiased estimates of the standard errors, and efficient parameter estimates where least squares may not, is available in Griffith (1988, 94-107).
17. Our OLS specification indicated the presence of heteroskedasticity. We therefore estimated the OLS model with White's standard errors. Technically, when OLS is corrected for heteroskedasticity, it is referred to as a GLS specification, but we use the term OLS for convenience throughout the article.
18. MATLAB code used to calculate all of the various Lagrange Multiplier tests used in this section is available from the authors upon request.

References

Abrams, Burton A., and James L. Butkiewicz. 1995. The influence of state-level economic conditions on the 1992 U.S. presidential election. *Public Choice* 85(1-2): 1-10.

Alvarez, Michael R., and Jonathan Nagler. 1997. Economics, entitlements, and social issues: Voter choice in the 1996 presidential election. California Institute of Technology, Social Science Working Paper 1021 (October), Pasadena, CA.

Anselin, Luc. 1988. *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Anselin, Luc., and Anil K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of applied economic statistics,* edited by Aman Ullah and David E. A. Giles, 237-89. New York: Marcel Dekker.

Anselin, Luc, Anil K. Bera, Raymond J. Florax, and Mann J. Yoon. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26(1): 77-104.

Barry, Ronald, R. Kelley Pace, and C. F. Sirmans. 1998. Spatial statistics and real estate. *Journal of Real Estate Finance and Economics* 17(1): 5-13.

Blackley, Paul R., and Edward M. Shepard, III. 1994. A statistical analysis of the effect of state-level economic conditions on the 1992 presidential election. *Public Finance Quarterly* 22(3): 366-82.

Brasington, David M. 2004. House prices and the structure of local government: An application of spatial statistics. *Journal of Real Estate Finance and Economics* 29(2): 211-31.

Brasington, David M., and Diane Hite. 2005. Demand for environmental quality: A spatial hedonic analysis. *Regional Science and Urban Economics* 35(1): 57-82.

Cohen, Jeffrey P. 2002. Reciprocal state and local airport spending spillovers and symmetric responses to cuts and increases in federal airport grants. *Public Finance Review* 30(1): 41-55.

Dubin, Robin A. 1998. Spatial autocorrelation: A primer. *Journal of Housing Economics* 7(4): 304-27.

Eisenberg, Daniel, and Jonathan Ketcham. 2004. Economic voting in U.S. presidential elections: Who blames whom for what. *Topics in Economic Analysis & Policy* 4(1), article 19.

Fair, Ray C. 1996. Econometrics and presidential elections. *Journal of Economic Perspectives* 10(3): 89-102.

Florax, Raymond, Hendrik Folmer, and Sergio Rey. 2003. Specification searches in spatial econometrics: The relevance of Hendry's methodology. *Regional Science and Urban Economics* 33(5): 557-79.

Garrett, Thomas A., and Thomas L. Marsh. 2002. The revenue impacts of cross-border lottery shopping in the presence of spatial autocorrelation. *Regional Science and Urban Economics* 32(4): 501-19.

Griffith, Daniel A. 1988. *Advanced spatial statistics: Special topics in the exploration of quantitative spatial data series*. Dordrecht: Kluwer Academic Publishers.

Kan, Kamhon, and C. C. Yang. 2001. On expressive voting: evidence from the 1988 U.S. presidential election. *Public Choice* 108(3-4): 295-312.

LeSage, James P. 1997. Regression analysis of spatial data. *The Journal of Regional Analysis and Policy* 27(2): 83-94.

Donald J. Lacombe is an assistant professor of economics at Ohio University in Athens, Ohio. He received his doctorate in economics from Florida State University. His research focuses on public finance and the application of spatial econometric techniques. His work has appeared in the *Journal of Economic History*, the *Southern Economic Journal*, and *Public Choice*.

Timothy M. Shaughnessy is an assistant professor in the Department of Economics and Finance at Louisiana State University in Shreveport. He received his doctorate in economics from Florida State University. His research focuses on applied microeconomic analysis of public choice and public finance issues. His work has appeared in *Regional Science and Urban Economics* and the *Journal of Labor Research*.