Bayesian estimation of the spatial Durbin error model with an application to voter turnout in the 2004 presidential election

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1177/0160017612452133

Publisher: Sage

www.reading.ac.uk/centaur

Bayesian Estimation of the Spatial Durbin Error Model with an Application to Voter Turnout in the 2004 Presidential Election

Donald J. Lacombe[1], Garth J. Holloway, and Timothy M. Shaughnessy

Abstract

The potential for spatial dependence in models of voter turnout, although plausible from a theoretical perspective, has not been adequately addressed in the literature. Using recent advances in Bayesian computation, we formulate and estimate the previously unutilized spatial Durbin error model and apply this model to the question of whether spillovers and unobserved spatial dependence in voter turnout matters from an empirical perspective. Formal Bayesian model comparison techniques are employed to compare the normal linear model, the spatially lagged X model (SLX), the spatial Durbin model, and the spatial Durbin error model. The results overwhelmingly support the spatial Durbin error model as the appropriate empirical model.

JEL Codes: C11, C21

Keywords: spatial econometrics, spatial Durbin Error model, Bayesian inference

I. INTRODUCTION

Elections have proven to be fruitful areas of research for empirical economists studying public choice issues. It is interesting to observe that the amount of research devoted to uncovering the factors influencing *how* a person votes is dwarfed by the volume devoted to uncovering the factors influencing whether a person votes at all. The sizable amount of literature attempting to explain voter turnout is perhaps due to the fact that most studies do not yield consistent results (Tollison and Willett, 1973; Matsusaka, 1995; Matsusaka and Palda, 1999; Geys, 2006). Attempts have been made to capture the effects of formerly omitted variables in an effort to drive up the predictive power of models, but most have been underwhelming.

This paper examines whether there is a spatial component to the as-yet unobserved influences on voter turnout. Factors that influence the costs and benefits of voting that are unobservable or difficult to measure may yet retain identifiable spatial effects. As a simple

---

[1] Donald J. Lacombe is Associate Professor of Agricultural and Resource Economics and Economics and Research Associate Professor, Regional Research Institute, West Virginia University, 886 Chestnut Ridge Road, Morgantown, WV, 26508-6825. Phone: 304-293-8543. Fax: 304-293-6699. email: donald.lacombe@mail.wvu.edu. Garth J. Holloway is Reader, Department of Agricultural and Food Economics, University of Reading, United Kingdom. Phone +44 118 378 6775. Fax +44 118 975 64 67. email: garth.holloway@reading.ac.uk. Timothy M. Shaughnessy is Associate Professor of Economics, Department of Economics and Finance, LSU in Shreveport, 1 University Place, Shreveport, LA, 71115. Phone: 318-797-5334. Fax: 318-797-5253. email: tshaughn@lsus.edu

example, good weather has been known to spur more turnout, yet quantifiable definitions of "good weather" may be hard to determine.[2] Two inches of rain during commuting times may affect turnout differently than two inches at 5am or 10pm. However, while it is likely that good weather in county A will affect turnout positively, it is equally likely that this same good weather effect will spill over to adjacent county B. Though it is difficult to observe the effect of good weather itself, we can estimate the aggregate effect of all unobservable factors that have a geographic or spatial component. For a given geographic area, these may include a common set of values or political beliefs, local partisan competition, access to only a few sources of news media, a predominant employer or industry in the area, or other similar factors. Shachar and Nalebuff (1999), Feddersen (2004), and others describe the importance of groups motivated by a leader, and it certainly stands to reason that group membership and influence is highly spatially related.

Similarly, political scientists and geographers have noted the importance of space for political issues, and have demonstrated the influence of local context and social interaction through spatial econometric techniques (e.g., O'Loughlin, et al., 1994) or more traditional non-spatial estimation methods (Pattie and Johnston, 2000; Gerber, et al. 2008, which analyzes voter turnout). Key (1984) presents evidence (prior to its original publication in 1949) from Alabama on localism or the "friends and neighbors" effect, where candidates win significant support from their home counties even though statewide support is minimal. He attributes this to several factors, including the absence of strongly-established, well-organized political parties; lack of a traditional recruitment and advancement system for candidates; the relatively high importance that voters place on local versus state issues; and lack of party loyalty. Similar patterns were seen in Georgia, Florida, South Carolina, and other Southern states. The reasons for the strength of localism differed among the states, but its empirical observability has generated academic research explicitly accounting for spatial effects.

Reliance on traditional non-spatial estimation techniques for data with known spatial effects is problematic. The presence of spatial error dependence can lead to biased standard errors much like other forms of autocorrelation. On the other hand, spatial dependence that is exhibited in the dependent variable (as a spatial lag) can lead to biased and inconsistent estimates of the parameters of interest. The benefit of utilizing spatial econometric techniques is that they can take into account these econometric issues that may unknowingly bias standard normal linear model results.

II. Model and Data

There are a few seminal economics articles in the voter turnout literature (Downs, 1957; Riker and Ordeshook, 1968; Barzel and Silberberg, 1973; Ashenfelter and Kelly, 1975; Filer and Kenney, 1980) whose theories motivate our models, but are generally well-known and need no review here. (A good review and meta-analysis of past studies is found in Geys, 2006). Voters make rational decisions on the costs and benefits of voting and behave accordingly. Typically, the benefits and costs are so small that slight disturbances in one or the other can tip turnout in unpredictable ways (see Aldrich, 1993). Despite this, there are characteristics of voters (or groups of voters) that most researchers generally agree will usually affect turnout. The job of

---

[2] Though Knack (1994) finds only weak support for the idea that rain hampers turnout.

empirical economists has been to quantify these characteristics and measure their effects while trying to shore up the unpredictability of turnout through adding more control variables. Typically overlooked in the existing literature, though, is an explicit accounting for geographically-correlated factors that affect turnout.[3]

Shachar and Nalebuff (1999) provides two important contributions to the study of voter turnout. First, it critiques the typical use of an *ex post* measure of election closeness as a control variable to describe the *ex ante* decision whether to vote. The potential value an individual vote has will depend on how close the election outcome is, since in a slim margin an individual vote has more weight. While most past research uses the actual outcome of the election as the measure of closeness, Shachar and Nalebuff recognize that this information is not available to voters heading to the polls.[4] A second contribution, mentioned earlier, is the recognition of the influence that political leaders have when they expend effort to boost turnout. Their estimations include controls for US regions, but it is this concept of an influential political leader that will be more relevant for our discussion of spatial dependence in county voter turnout.

The political science literature has considered the effect of geography on turnout and other forms of political participation. A recent paper by Cho and Rudolph (2008) seeks to uncover, via a spatial lag model, whether the observed spatial dependence in political participation is attributable to individual-level or macro-level characteristics (e.g. race, income), to diffusion through formal or informal social networks, or to subtle environmental cues taken through casual observation (e.g. yard signs, well-kept gardens, bumper stickers). After controlling for a typical set of individual, aggregate, and social network variables, the model's spatial lag parameter is positive and significant, and a Lagrange Multiplier test for residual autocorrelation is insignificant. The authors maintain that this provides evidence that the spatial effect of participation is a result of casual observation. Concern remains, however, about their definition of their dependent variable. The data come from the Social Capital Benchmark Survey where respondents were asked if, in the last year, they 1) signed a petition, 2) attended a political meeting/rally, 3) worked on a community project,[5] or 4) went to a protest/march/etc. The dependent variable was thus an additive index ranging from 0-4. Quantifying political participation in this way seems somewhat arbitrary; is a "2" respondent 100% more participatory than a "1" respondent? What if the "1" respondent attended 20 political meetings of different parties while the "2" respondent signed a single petition and worked on a single community project? The construction of the dependent variable may call into question the results, although the premise is interesting.

A second paper in political science that measures spatial dependence is Cutts and Webber (2010). In explaining political party vote shares, they find evidence that campaigning exhibits

[3] Attempts are occasionally made at capturing broad geographic effects, as in Ashenfelter and Kelly (1975) including a dummy variable for turnout in the South. It is doubtful that a convincing spatial story can be told when the control used is so broad.

[4] Contrary to this, Matsusaka (1993) finds that in ballot proposition elections, closeness does not affect turnout.

[5] No definition was given for "community project," and it is not clear how involvement in this activity demonstrates political participation. The demonstration/protest activity suffers from the same critique. A "2" respondent may have volunteered at a town fair or charity carwash as their "community project," and sat in for a few minutes at a protest at a company headquarters on their way to work, and be deemed more politically involved than a "1" respondent who attended all major campaign appearances by every mayoral candidate of every political party in their city. This may be a flaw in the survey instrument itself and outside of the authors' control, but the authors do not seem to consider it an issue worth addressing.

spatial dependency; i.e. local campaign spending by a political party not only positively affects that party's vote share in the constituency in which the spending is done, but also positively affects that party's vote share in neighboring constituencies. They also estimate spatial error models. A possible flaw is the limited set of control variables (percent with higher educational degrees; percent of the population employed in manufacturing, agriculture, or education; percent who are pensioners, students, and Muslims; and the percent living in owner-occupied housing). There are also controls for campaign spending by the political parties, but other variables identified as important in the extant literature are not included in their analysis. One shortcoming of the study is that the authors do not calculate the proper marginal effects. The coefficients from a standard SAR model do not represent the true marginal effect of a change in an explanatory variable on the dependent variable. LeSage and Pace (2009) provide details regarding the calculation and interpretation of the direct, indirect, and total effects.

Kim, et al. (2003) use a spatial lag model in a Bayesian framework to uncover spatial dependence in vote shares for the Democrats and Republicans in Presidential elections from 1988 to 2000, as well as empirically testing whether voters decide current votes on past candidate behavior or on their expectation of the success of a particular party at fixing current economic problems. They construct two spatial weight matrices, first-order contiguity and a commuting matrix to capture the number of commuters from one county to another ("geographic" neighbors and "economic" neighbors). Their model of vote share is estimated using per capita income and the unemployment rate as controls. Even with this small set of explanatory variables (which, in some models, appear insignificant), they find significant spatial dependence. One shortcoming of this study is that the authors use maximum likelihood based Lagrange Multiplier test statistics to determine the most appropriate model and then subsequently use Bayesian techniques to estimate the model. In the Bayesian paradigm posterior model probabilities should be calculated for each model and then compared to determine the most appropriate model[6].

Darmofal (2006)[7] says "citizens' costs and benefits of voting and, by extension, their turnout, are recognized as depending not just on who citizens are [e.g., capturing their demographic or economic characteristics], but also on where they live." (p.127). He examines county-level turnout for each presidential election from 1828 through 2000 and computes a global Moran's *I* to detect spatial autocorrelation. After a series of diagnostic tests, Darmofal (2006) then estimates several spatial error models (SEM) to correct for the resulting spatial error dependence. Interestingly, Darmofal (2006, p. 145) finds "no clear evidence of contagion, or spatial lag dependence." Another paper in the political science literature that indirectly addresses the geographic component of voter turnout is Nagel and McNulty (1996). The model is somewhat reversed from ours as they attempt to explain the vote share for Democrats as a function of turnout, but they examine elections from 1928 through 1994 and separate Southern and non-Southern states, and then look at elections in each state. Their primary aim is to assess the prevailing notion that turnout helps Democrats, so they are not addressing our topic directly but at least recognize geographic differences in turnout (albeit at a more aggregated level than our county-level data). They do not quantify, however, any spatial spillover effects.

---

[6] The authors also note that "the Bayesian approach allows us to estimate the model without normality assumption..." when in reality the likelihood for these models is based on the assumption of normality.

[7] Cox (2003) discusses the differences in turnout among EU countries, but the analysis is merely descriptive.

Nickerson (2008) examines the spillover effect on turnout after one member of a two-person (two-voter) household has a direct face-to-face contact with a volunteer get-out-the-vote (GOTV) representative. In two cities, three groups of households were selected: one group was contacted with GOTV, a second with a message to recycle, and a third was not contacted at all. The spillover effect was determined by comparing the turnout rates of non-contacted persons (secondary message recipients). Directly-contacted GOTV voters were about ten percent more likely to vote, and indirect voters were six percent more likely to vote.

Some issues with Nickerson's results help to motivate our study. First, the estimated effect on secondary voters is not significant at 5% in a one-tailed test (though it is for direct voters) calling into question the spillover effect of voter turnout. Second, the effect on the secondary voter decays pretty quickly (and could be irrelevant if the effect was insignificant) and the implication is that the effect on tertiary voters (voting-age children, friends, coworkers) would be minimal, implying a spillover effect that would be quite small or negligible at the household (rather than individual voter) level, which again would indicate a lack of spatial autocorrelation at the county level in voter turnout. Third, the estimated spillover effect of the GOTV effort is determined by the percentage of the effect on the direct voter that induces the secondary voter to turn out. The author argues this effect is quite large (the direct voter passed on about 60% of the propensity to vote), but the individual effects on the direct and secondary voters was quite small (10% and 6%, respectively). It is not clear that a sizable spillover effect has been discovered when its impact is only a four-percentage point increase in turnout.

The existing economics and political science literature provides the foundation for our estimated models. The decision to vote should be influenced by demographic, economic, and political factors that can be distinguished from those likely associated with nonvoters. To the extent that these individual factors are observable at the county level, hypotheses about county turnout can be made.[8] The control variables we include are standard controls seen in the voter turnout literature. The contribution of this paper is in formulating and estimating several spatial econometric models designed to measure the extent to which factors influencing turnout at the county level are spatially correlated and to draw inferences from the model that best fits the data. As we will see in the model comparison exercise, the heretofore unused spatial Durbin error model "wins" our model comparison race[9] and we draw inferences from this model..

We chose to analyze the 2004 Presidential election turnout at the county level, given that the popular "red/blue" county map appears to indicate significant spatial clustering. 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 influencing turnout that vary over space need to be accounted for in a systematic manner.[10] There may be local customs or historical reasons that certain areas of the country

---

[8] It is important to avoid the ecological inference problem in interpreting the results; our data are county-level and thus inferences drawn therefrom should not be assumed consistent at the individual level. See Robinson (1950): "the only reasonable assumption is that an ecological correlation is almost certainly not equal to its corresponding individual correlation."

[9] Elhorst (2010, p. 14) states the following regarding the spatial Durbin error model: "A model with spatially lagged independent variables (WX) and a spatially autocorrelated error term has hardly been used in the literature." In fact, the authors are unaware of its use in any previous empirical study.

[10] It is also possible that spatial dependence may be exhibited in the dependent variable, which would imply a spatially lagged y (i.e. rWy) model. However, theoretical reasons exist that cast doubt on this possibility and our model comparison exercise overwhelmingly does not favor this type of model.

vote in higher numbers that is ignored in the existing literature. This issue has normally been dealt with by using dummy variables for those regions but this approach does not properly model the interaction among the geographic entities. The spatial econometric techniques that we use better model these interactions. The sample includes 3,061 counties and excludes Alaska, Hawaii, Washington, D.C., and all cities in Virginia.[11]

The dependent variables that we use represent the turnout of the county voting-age population (VAP) and the turnout of the citizen voting-age population (CVAP). Turnout is defined as the number of total votes cast for Bush and Kerry (from POLIDATA Demographic & Political Guides[12]) in the county divided by the relevant population. The voting-age population is the 2004 resident population estimate of those 18 and older from the Census Bureau's Annual County Population Estimates,[13] and the citizen voting-age population multiplies this by the Census' Current Population Survey report of the percentage of the state 18 and older population that is a citizen.[14]

McDonald (2002) uses a similar voting-eligible population (VEP) measure, where he excludes not only non-citizens as we do but ineligible felons, as an alternative to the traditional VAP. Typically, felons include prisoners, parolees, and half of probationers, but state laws vary as to which of these groups are allowed to vote (e.g., in McDonald's data, Virginia denied the vote to all of these groups, whereas Vermont denied it to none). McDonald calculated the number of "ineligible felons" in each state (according to the state laws existing at the time) to arrive at his VEP.

For comparison purposes, we constructed a VEP turnout rate using 2004 data on parolees and probationers,[15] and prisoners.[16] The VEP is calculated similar to the CVAP where the county population was multiplied by the state citizen percentage, but the VEP further adjusts this for the state "ineligible felon" percentage.[17] The ineligible felon population is relatively small; the means of the VAP, CVAP, and VEP turnout rates, respectively, are 58.1%, 61.6%, and 62.6%. We estimated models using the VEP, but the results (available upon request) were

---

[11] 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. Four counties were dropped for turnout rates exceeding one (due in two cases to the state citizen percentage being lower than the likely actual county citizen percentage, causing citizen turnout to exceed one, and in two cases likely to data reporting error).

[12] http://www.polidata.org

[13] http://www.census.gov/popest/counties/asrh/CC-EST2005-agesex.html

[14] http://www.census.gov/population/www/socdemo/voting/cps2004.html, and http://www.census.gov/population/socdemo/voting/cps2004/tab04b.xls

[15] Bureau of Justice Statistics, *Probation and Parole in the United States*, 2004. Tables 2 and 4, using 12/31/04 populations. http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=1108

[16] U.S. Census Bureau, Statistical Abstract 2012, Table 347, which references Bureau of Justice Statistics *Prisoners in 2009* and earlier report.

[17] Though our data is from 2004, we identified ineligible felons according to McDonald's Table 1 which lists the voting eligibility of prisoners, parolees, and probationers by state. It is possible that state laws have changed between the time of McDonald's article and 2004.

not noticeably different from the VAP;[18] further, our model comparison exercise still demonstrates preference of the VAP model over the VEP.

Our use of two measures of turnout rather than one has support in the literature, especially as it applies to aggregate-level studies. Geys (2006) notes that of the 83 studies reviewed, about 60% of them used turnout measures similar to ours. We also transformed the dependent variables using the logit transformation given that our dependent variable is measured as a proportion.[19]

Demographic variables

Several papers argue that religious sentiment is associated with a stronger sense of civic duty,[20] and it was also 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.[21] 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. % religious adherents is the percentage of the total population in the county who are considered adherents of one of the religious groups. The churches variable is the number of churches per 10,000 people in the total population in 2000. A higher percentage of adherents or number of churches at the county level is expected to positively affect turnout.

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. The typical belief in most studies of voter turnout associates white voters with higher turnout and black and Hispanic voters with lower turnout, so our hypotheses are similar.[22] A different opinion is offered by Oberholzer-Gee and Waldfogel (2001), which examines the phenomenon of how black turnout is increased when the proportion of blacks in a jurisdiction rises. This finding, while throwing some doubt on the expectation of lower turnout for blacks generally, also supports the group theory of voting of Shachar and Nalebuff (1999) and others.

Most articles include a measure of the age of the electorate,[23] since older voters are expected to turn out in higher numbers than younger voters. Thus, we include the median age

---

[18] For example, duplicating Table 5 (Total Effects) using the VEP yields ten significant variables compared to the eleven using the VAP. Four coefficients that were significant in the VAP (white, Hispanic, same house, and male with ba) become insignificant in the VEP, and three that were insignificant in the VAP (median age, rGDP growth, and the Texas dummy) become significant in the VEP. Three coefficients switch sign in the VEP model but two (churches and gay marriage vote) are insignificant in both models while the Texas dummy becomes positive and significant. The magnitudes of the coefficients change somewhat but for the significant variables the change is relatively minor.

[19] Specifically, we use the following transformation: $y = \log(y/(1-y))$.

[20] Timpone (1998) finds that church attendance is positive and significantly related to turnout. This is true even of those who had recently moved in the past two years, demonstrating that church attendance is closely associated with social involvement.

[21] http://www.glenmary.org/grc/RCMS_2000/rankings.htm

[22] Geys (2006) finds support for this hypothesis as well: two-thirds of the studies find lower turnout where minority share of the population is higher.

[23] See Merrifield (1993) for an example.

of the resident population for each county from the Census Bureau's Annual County Population Estimates, and expect it to be positively associated with turnout.

A higher cost of voting should be associated with lower turnout, and some articles include a measure of the number of single-parent families.[24] Such a family would presumably have less time available for voting, and so we include the percentage of family households that were single parent in each county from the 2000 Census.[25] The percentage of single-family households is expected to be negatively associated with county turnout.

We also hypothesize that voters with stronger ties to their communities will be more likely to vote in elections. Stigler (1975) suggests the importance of length of residence in a community, and Geys (2006) finds strong support among existing turnout studies of a positive relationship between "population stability" (including homeownership) and voter turnout. To proxy this, we calculated the percentage of the county population who resided in the same house between the years 1995 and 2000.[26] The higher the proportion of long-term county residents, the higher we expect county turnout.

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 turnout. These variables are defined as the percentage of a county's total male 25-and-over population who hold a bachelor's degree, and a similar definition for females. These data come from the 2000 Census SF3 database. Lacombe and Shaughnessy (2007) found that women's preferences for Kerry were stronger than men's preferences for Bush though in some papers (e.g. Ashenfelter and Kelley, 1975) men have a higher turnout; we therefore expect a difference in turnout between the sexes with perhaps a slight edge toward males.

Also included in our regression model is the percentage of the county population living in urban areas, obtained from the 2000 Census SF1 database. Hackey (1992) finds a positive association between turnout of black voters in the 1976, 1980, and 1984 presidential elections and whether the voter lived in an urban area, but one might posit that voters living in higher density areas perceive more competition, and thus a lower weight attached to their vote, than do rural voters. This negative relationship between urbanization and turnout is weakly supported in Geys (2006) meta-analysis of previous voter turnout studies, where urbanization and turnout are usually negatively but insignificantly related.

Economic variables

A persistent finding in the literature on turnout is that economic conditions of voters influence turnout. First, higher income voters tend to vote more often than lower income voters,[27] so we include a measure of the natural log of per capita personal income in 2004 for the county. The data come from the Bureau of Economic Analysis' Regional Economic Accounts.

---

[24] Wittman (1975) suggests the importance of this variable in influencing voting costs.
[25] Summary File 1, Table P19. (Male householder, no wife present + Female householder, no husband present) / Family households.
[26] Census 2000 Summary File 3, Table P24.
[27] Wittman (1975) hypothesizes that this does not have to be true; higher income people may vote less and lower may vote more, though the existing evidence does not support his point.

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 voice their displeasure.[28] Thus, we expect that a higher county unemployment rate is associated with higher turnout.

Similar to unemployment, voters would be more inclined to be satisfied with the incumbent if output growth was good. Shachar and Nalebuff (1999) include a measure of GNP growth that is positively associated with turnout and significant. We thus include the growth rate of real state GDP, and expect that a higher GDP growth rate is associated with higher turnout.

Given the prominence that labor unions place on political efforts and voter turnout drives,[29] the amount of unionization is expected to influence turnout as well. We collected 2004 data from the Bureau of Labor Statistics' Labor Force Statistics from the Current Population Survey on union affiliation of employed wage and salary workers by state.[30] Higher unionization is expected to be positively associated with turnout.

Political variables

An important issue in the 2004 election was controversy over Bush's handling of the War in Iraq. A proxy that we use to express strong sentiment about the Iraq War is the percentage of the civilian population age 18 and 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.[31] Veterans on average would have strong opinions about the continuation of the Iraq War, and thus we hypothesize that a higher veteran percentage in the county is associated with higher county turnout.

Another important political issue in the 2004 election was the presence of constitutional amendments to define marriage as "one man-one woman" on state ballots; voters (particularly religious voters) were believed to have turned out at much higher rates to voice their opinion. 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.[32] Even though popular opinion was that the presence of a gay marriage amendment would boost turnout, Lacombe and Shaughnessy (2007) found that the presence of these amendments did not significantly affect the percentage of Bush votes, so we expect at best a weak positive association between these amendments and turnout.

An important hypothesis in the voter turnout literature is that the value of a vote drives turnout; if voters perceive that their vote contributes more to the outcome of the election, then they will vote in higher numbers. The typical way this is measured is by including a variable on population, but since we already utilized population in constructing our turnout rate variable

---

[28] Merrifield (1993) found a positive relationship between unemployment and turnout.
[29] See Delaney, Masters, and Schwochau (1988) and Zullo (2004).
[30] Table 5, http://www.bls.gov/webapps/legacy/cpslutab5.htm
[31] http://www1.va.gov/vetdata/page.cfm?pg=1
[32] http://www.cnn.com/ELECTION/2004/pages/results/ballot.measures/

we rejected that approach.[33] We proxied "vote value" in two ways. First, it is reasonable to assume that voters assign a weight to their Presidential vote in proportion to the number of electoral votes assigned to their state since electoral votes will mirror population; thus, to measure the county's share of the state's electoral vote we multiplied the number of state electoral votes[34] by the percentage of the state population that resides in the particular county. We expect a negative association between this county electoral vote share and county turnout.

"Vote value" is also likely affected by the growth rate of the surrounding population. Faster growth of the population of competing voters will more quickly dilute the importance of a particular vote.[35] To measure this, we collected data on the county age 18 and over population in 2003 and 2004 from the Census' Annual Estimates of County Population and calculated the growth rate between them.[36] As the voting-age population grows in the county, we expect that county turnout should fall.

A debate in the turnout literature exists over how to control for the perceived closeness of the election. Obviously, the closer the election is perceived to be the greater value an individual will place on his vote, and thus the higher the expected turnout. Many times the measure for expected closeness is the actual election outcome itself,[37] but as detailed in Shachar and Nalebuff (1999) and Geys (2006) this raises the problem of trying to measure an *ex ante* expectation with an *ex post* outcome. In support of *ex ante* measures, Geys' (2006) meta-analysis finds that *ex ante* measures of closeness have a 74% "success rate" (achieving an expected result) versus a 51% "success rate" for *ex post* measures. One proxy is to include pre-election opinion polls or professional analysis or opinion. Carter (1984) tests both *ex post* and *ex ante* measures, where the *ex ante* are measured by *Time* and *Newsweek* magazines considering the state to favor one candidate strongly, to simply favor one candidate, or to be too close to call. The *ex post* measures of actual closeness (percent of vote going to winning candidate) were unexpectedly positively associated with turnout, while the *ex ante* results showed that stronger leaning for a candidate reduced turnout as expected. We adopt the same approach here, and use the October 29, 2004 Cook Electoral Rating from the Cook Political Report.[38] A state that was "solid" for Bush or Kerry was coded as 0; "likely" Bush or Kerry was coded as 1; "leans" Bush or Kerry as 2; and a "toss up" as 3. Given our coding, we expect a positive association between the Cook variable and county turnout.

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. We expect that voters in the candidates' home states will be more likely than others to turn out, *ceteris paribus*, since they

---

[33] Geys (2006) meta-analysis finds only a third of the studies test for the effect of population on turnout, but of those, 56% find the expected negative relationship between population size and turnout. Quantitatively, a one-standard deviation change in population reduces turnout by 0.48 to 0.65 standard deviations, so that Geys concludes "Population size has a statistically significant negative effect on turnout."

[34] From the Federal Elections Commission: http://www.fec.gov/pages/elecvote.htm

[35] Merrifield (1993) includes a measure of population growth and finds it is highly significant and negatively related to turnout.

[36] http://www.census.gov/popest/counties/asrh/CC-EST2005-agesex.html

[37] For example, Matsusaka (1993). Geys (2006) finds strong support for this position (one standard deviation increase in closeness leads to a statistically significant 0.58-0.69 standard deviation increase in turnout), but also points out that several studies take an indirect approach, since close elections drive more voter mobilization efforts by political parties, which thus drives higher turnout.

[38] As reported on CNN's 2004 Election page:
http://www.cnn.com/ELECTION/2004/special/president/showdown/cook.html

presumably have more, and more reliable, information about the home candidate than "foreign" states. Descriptive statistics of the data used and their sources are in Table 1.

III. Econometric Models

While we included what we consider a relatively rich set of controls, it is possible that there are unobservable factors such as political sentiment, sense of civic duty, or group identities that influence voter turnout that retain a strong geographical or spatial component. A strong sense of civic pride and duty to vote that prevails in one county likely will affect turnout in neighboring counties. Thus, these omitted unobservable factors that vary systematically over geographic space may result in residual spatial autocorrelation. There is also the possibility that our dependent variable could exhibit spatial autocorrelation if we believe that voter turnout in one county affects voter turnout in neighboring counties. In either case, the inferences drawn from standard econometric techniques (such as the normal linear model) may be misleading.

In order to overcome these deficiencies, we estimate voter turnout using four separate models: the normal linear model, the spatially lagged X model (SLX), the spatial Durbin model, and the spatial Durbin error model. We then employ Bayesian model comparison techniques to choose the most appropriate model.

The first model that we estimate, and the most common one found in the literature, is the normal linear model. The estimating equation for this model is as follows:

$$y_{VAP} = a + b_D X_D + b_E X_E + b_P X_P + e$$
$$y_{CVAP} = a + b_D X_D + b_E X_E + b_P X_P + e$$
$$e : MVN\left(0, s^2 I_n\right)$$

where y_{VAP} is the (logit transformed) proportion of the voting-age population who turned out to vote; y_{CVAP} is the (logit transformed) proportion of the citizen voting age population who turned out to vote; X_D is the vector of demographic variables (% religious adherents, churches, % white, % black, % Hispanic, median age, % single-parent, % in same house, % male with bachelor's, % female with bachelor's, and % urban population); X_E is the vector of economic variables (log of per capita income, unemployment rate, real GDP growth from 2003-04, and state union member rate); X_P is the vector of political variables (% veterans, gay marriage amendment dummy, county share of state electoral votes, population growth of 18 and older from 2003-04, Cook closeness rating, and Texas and Massachusetts dummies); b_D, b_E, and b_P are the vectors of coefficients to be estimated, and e is an i.i.d. normal error term.

Our first spatially explicit model, the spatially lagged X (SLX) model is identical to the normal linear model specification above with the exception that spatially weighted independent

variables are added to the specification. The spatially weighted independent variables take the general form of $W´X$, where W is an $N´N$ row-standardized contiguity spatial weight matrix and X is an independent variable. The newly formed WX variables represent a weighted average of surrounding values of the independent variables and are designed to pick up any influence of surrounding values of the independent variables in question on the dependent variable. It may be that independent factors outside of a voter's home county may influence their decision to go to the polls and vote. For example, if the unemployment rate in an adjoining county is high, voters may perceive that the economy is worse off than if they just observed their own county's unemployment rate. A spatially lagged independent variable is designed to pick up such factors in a voter's decision making process.

The spatially lagged X (SLX) estimating equations take the following form:

$$y_{VAP} = a + b_D X_D + b_E X_E + b_P X_P + q_D W X_D + q_E W X_E + q_P W X_P + e$$
$$y_{CVAP} = a + b_D X_D + b_E X_E + b_P X_P + q_D W X_D + q_E W X_E + q_P W X_P + e$$
$$e : \ MVN\left(0, s^2 I_n\right)$$

where q_D, q_E and q_P represent the coefficients on the spatially weighted demographic, economic, and political variables, respectively.

The third model that we estimate and test is the spatial Durbin model. The spatial Durbin model allows for spatial spillovers in the dependent variable (through a spatially lagged dependent variable) as well as through spatially lagged independent variables. LeSage and Pace (2009) note that the spatial Durbin model is appropriate if two separate conditions hold. First, there must be omitted variables from the model that are spatially correlated. Second, these spatially correlated omitted variables must be correlated with an included explanatory variable in the model. In our particular case, it may be that a variable that measures some membership in a civic group is omitted from our model and that this variable is spatially correlated, simply because memberships in these groups tend to be geographically based. Additionally, this spatially correlated omitted variable that measures group membership could be correlated with another included explanatory variable, such as per-capita income. If our model were to exhibit such characteristics, then the spatial Durbin model would be most appropriate.

The spatial Durbin model takes the following form:

$$y_{VAP} = a + rW y_{VAP} + b_D X_D + b_E X_E + b_P X_P + q_D W X_D + q_E W X_E + q_P W X_P + e$$
$$y_{CVAP} = a + rW y_{CVAP} + b_D X_D + b_E X_E + b_P X_P + q_D W X_D + q_E W X_E + q_P W X_P + e$$
$$e : \ MVN\left(0, s^2 I_n\right)$$

where r is the spatial autocorrelation parameter, W is an $N´N$ spatial weight matrix, and q_D, q_E and q_P represent the coefficients on the spatially weighted demographic, economic, and

political variables, respectively and the b's represent the coefficients on the non-spatially weighted explanatory variables.

As mentioned before, although we have included a number of independent variables that are believed to influence voter behavior, there is the possibility that there are unobserved factors that may vary systematically over space, resulting in residual spatial error correlation. Since neither the normal linear model nor the spatially lagged X (SLX) model formally take this residual spatial correlation into account, we estimate what LeSage and Pace (2009, pp. 41–42) refer to as the *spatial Durbin error model*. The spatial Durbin error model combines the spatially lagged X (SLX) model with an error process that accounts for residual spatial autocorrelation:

$$y_{VAP} = a + b_D X_D + b_E X_E + b_P X_P + q_D WX_D + q_E WX_E + q_P WX_P + u$$
$$y_{CVAP} = a + b_D X_D + b_E X_E + b_P X_P + q_D WX_D + q_E WX_E + q_P WX_P + u$$
$$u = lWu + e$$
$$e : \ MVN\left(0, s^2 I_n\right)$$

where the dependent and independent variables are as before, W is an $N \times N$ row-stochastic spatial weight matrix, l is the spatial autocorrelation coefficient, and e is an i.i.d. error term. The spatial Durbin error model is an augmented spatially lagged X (SLX) model that allows for a spatially autocorrelated error structure. We note that this model does not contain a spatially lagged dependent variable, which is consistent with Nickerson's (2008) observations regarding spillover effects in voter turnout at the household level, i.e. that spillovers are most likely highly localized and are probably non-existent at the county level.

Even though there may be theoretical or empirical reason to doubt the validity of including a spatially lagged dependent variable in our specification, we estimate and test the spatial Durbin model as part of the modeling exercise. Given that we have four possible models that can describe the data generating environment, we now turn to the statistical model and a description of the Bayesian techniques that are utilized to make inferences.

IV. Statistical Model

By way of notation, let q denote a vector of parameters of interest, $p\left(q\right)$ the prior probability density function (pdf) for q, and let $f\left(y\,|\,q\right)$ represent the likelihood function. The posterior distribution of the parameters, namely $p\left(q\,|\,y\right)$, is derived via Bayes' Rule

$$p\left(q|y\right)= \frac{p\left(y|q\right)p\left(q\right)}{p\left(y\right)}$$

where $p\left(y\right)$ is the integrating constant that ensures that the posterior probability density integrates to unity.[39] Given that $p\left(y\right)$ does not involve the parameter vector q, we can ignore this constant in subsequent analyses and write Bayes' Theorem in a familiar form

$$p\left(q|y\right) \mu \ p\left(y|q\right)´ \ p\left(q\right)$$

thus resulting in the familiar Bayesian phrase, "the posterior is proportional to the likelihood times the prior." Ideally, we would like to draw inferences regarding the parameters of the model by analytically integrating the joint posterior distribution for each of the model's parameters, resulting in a marginal distribution for each parameter. However, the analytical solution to this integration problem is available only in a few select cases. In deriving the marginal distributions these complications force us to draw inferences using iterative procedures referred to, generically, as *Markov Chain Monte Carlo* methods (MCMC). Specifically, we will make use of the Gibbs sampler and the Metropolis-Hasting algorithm to provide robust inferences regarding the model parameters.

The Gibbs sampler is an algorithm to generate a sequence of samples from the joint posterior distribution of the parameters when an analytical solution is unavailable. Gibbs sampling is applicable when the joint posterior is intractable, but the full conditional distributions of each parameter is known. In the case of the normal linear model as well as the spatially lagged X (SLX) model, the full conditional distributions follow standard forms for the b and s^2 terms, namely the multivariate normal distribution for b and an inverted Gamma distribution for the s^2 term[40]. In the case of the spatial Durbin model and spatial Durbin error model, the full conditional distributions for the b and s^2 terms retain the same distributional form as in the normal linear model case[41], however, there is the added complication regarding the full conditional distribution for the l term (or r term in the spatial Durbin model). The full conditional distribution for the l (or r) term does not fall into any recognizable distributional form, and therefore the Metropolis-Hastings algorithm must be used to draw inferences regarding this parameter. Details regarding MCMC estimation of spatial econometric models is covered in LeSage and Pace (2009, Chapter 5) and Lacombe (2008).

---

[39] The $\pi(y)$ quantity is also referred to as the marginal likelihood and plays a vital role in model comparison exercises.

[40] Details regarding the Gibbs sampler as applied to the normal linear model can be found in Koop (2003, Chapter 4).

[41] The full conditional distributions for the b and s^2 terms will take the same form, however, the arguments in each of the distributions will differ from the normal linear and spatially lagged X (SLX) models.

The formula for Bayes' Rule explicitly allows for prior information to be included in the statistical analysis. In each of our models, we use proper prior distributions, but with relatively non-informative values. Specifically, we set the prior for the b's to come from a multivariate normal distribution with mean $\hat{b} = 0_K$ and covariance $C_{\hat{b}} = 10,000 \times I_K$. The prior values for the s^2 parameter, which comes from the inverted Gamma distribution, are $v_0 = 1$ and $s_0^2 = 1$ and the prior values for the r term in the spatial Durbin model are mean $\hat{r} = 0$ and covariance $C_{\hat{r}} = 10,000$, which is distributed univariate normal. These same values for the prior mean and prior covariance are used for the l parameter in the spatial Durbin error model.

Another appealing aspect of Bayesian analysis is the formal statistical derivation of model comparison techniques. The essential inputs in Bayesian model comparisons are the marginal likelihoods of competing models. The marginal likelihood, denoted $p(y)$, is the integrating constant that ensures that the posterior distribution integrates to unity. Until recently, the computation of the marginal likelihood has proved to be extremely burdensome for all but the simplest models. The innovation that paves the way for general application is an algorithm developed by Chib (1995). The marginal likelihood

$$p(y) = \int f(y \mid q) p(q) \, dq$$

can alternatively be written as the product of the sampling density as the prior for q divided by the posterior for q

$$p(y) = \frac{f(y \mid q) p(q)}{p(q \mid y)}$$

The above quantity is what Chib (1995) refers to as the basic marginal likelihood identity and robust estimates of $p(y)$ are available. Once converted to the computationally convenient log scale, an estimate of the log marginal likelihood $\ln p(y)$ is available as

$$\ln p(y) = \ln f(y \mid q^*) + \ln p(q^*) - \ln p(q^* \mid y)$$

where $f(y \mid q^*)$, $p(q^*)$ and $p(q^* \mid y)$ denote estimate of the sampling density, the prior ordinate, and the posterior ordinate, respectively at the point $q = q^*$. In the case that one of the full conditional distributions is derived via a Metropolis-Hastings step, the basic algorithm of Chib

(1995) can be amended to include the parameter of interest in the calculation of the log marginal likelihood (Chib and Jeliazkov, 2001) .[42]

V. Results

We begin our discussion of the results by noting that across all four model specifications (i.e. the normal linear model, the spatially lagged X (SLX) model, the spatial Durbin model, and the spatial Durbin error model) the model exhibiting the largest value of the log-marginal likelihood was associated with the dependent variable that measured county turnout as a proportion of the county voting age population. [43]

Table 2 contains the result of our model choice exercise. Looking at the second column, we see that the model with the highest value of the log-marginal likelihood is the spatial Durbin error model. Alternatively, we may calculate posterior model probabilities according to the following formula:

$$p\left(M_i \big| y\right) = \frac{p\left(M_i\right) p\left(y \big| M_i\right)}{\overset{J}{\underset{j=1}{\textrm{å}}}\, p\left(M_j\right) p\left(y \big| M_j\right)}$$

where $p\left(M_i \big| y\right)$ is the posterior probability of model i, $p\left(M_i\right)$ is the prior probability of model i, and $p\left(y \big| M_i\right)$ is the marginal likelihood for model i, where $\left\{M_1, \textrm{K}, M_J\right\}$ denotes each of the J models. We place equal prior probability on each of the three models in question, so that $p\left(M_i\right)$ is equal to $1/4$. The second column of Table 2 contains the results of our calculation and indicates that the spatial Durbin error model is almost certainly the preferred model, with a posterior model probability » 1. Given the overwhelming evidence that the most preferred model is the spatial Durbin error model, we restrict our interpretation of the results to this particular model. However, we must first make a few notes regarding the interpretation of such models.

Spatial econometric models that contain a spatially lagged y variable (i.e. rWy) need to be interpreted with care in that the coefficients from these models need to be transformed before interpretation can take place. LeSage and Pace (2009, Chapter 2) develop a method that allows one to calculate the direct, indirect, and total effects estimates that allows one to properly interpret the regression results. The direct effect is the effect of a change in an

---

[42] Both methods allow for the calculation of a numerical standard error for the marginal likelihood. See Chib (1995) for details.
[43] The log-marginal likelihood calculation also indicates the VAP model is preferred to the VEP (McDonald, 2002).

explanatory variable at location i on the dependent variable at location i (plus feedback effects), the indirect effect is how a change in an explanatory variable at location j ¹ i affects the dependent variable at location i, and the total effect is the sum of the two effects estimates.

Spatial regression models such as the spatial error model (SEM) that do not involve spatial lags of the dependent variable produce coefficient estimates that are interpreted in the standard fashion, i.e. effect of a change in an explanatory variable (regardless of location) on the dependent variable is simply equal to the coefficient of the explanatory variable. However, standard SEM models do not allow for indirect impacts to arise from changes in explanatory variables, i.e. a change in an explanatory variable at location j cannot affect the dependent variable at location i, which could be considered a shortcoming of the standard SEM model.

The spatial Durbin error model that we use to draw inferences, as explained by LeSage and Pace (2009, p. 42), does allow for spatially lagged independent variables, in the form of our WX variables. In terms of the direct, indirect, and total effects estimates, these are easily derived from the underlying econometric model. The direct effects correspond to the model parameters b, the indirect effects correspond to the model parameters q, and the total effects are the sum of the two.[44] We perform inference by calculating Bayesian 95% credible intervals using the set of 10,000 retained draws from the MCMC estimation routine. Credible intervals that do not contain the value zero point to a marginal posterior distribution for the parameter that is far enough away from zero that it gives credence to an important role played by that particular explanatory variable in explaining voter turnout at the county level.

In sum, our spatial Durbin error model allows for a richer interpretation compared to a standard spatial error model (SEM) and has the advantage that calculation of the direct, indirect, and total effects estimates is straightforward.

Table 3 contains information regarding the direct effects in our spatial Durbin error model. Recall that the direct effect is the effect of a change in an explanatory variable at location i on the dependent variable at location i. We can think of the direct effects as the "own effects" of a change in an explanatory variable. Entries in bold represent independent variables that are associated with explaining variation in the dependent variable, i.e. that do not contain zero within their credible intervals.

For our demographic variables, we find that the signs of the coefficient estimates are in accordance with our a priori expectations for those variables associated with turnout. Religious sentiment may have played a role in the 2004 Presidential turnout as Bush self-identified as a Christian, and to capture the possible relationship we utilized two different measures of county religiosity. The results indicate that the number of churches per 10,000 population is positively associated with voter turnout at the county level, while the number of religious adherents is not. The adherents variable captures anyone who self-identifies with one of over 149 religious

---

[44] Gelfand et al. (1990) show that one can perform valid inference regarding non-linear parameter relationships using non-linear combinations of the MCMC draws.

groups, which obviously includes some non-Christian denominations or others who might be wary of voting for someone who self-identifies at Christian. The churches variable is more likely to capture voter sentiment regarding religious preferences and this may be why we obtain these results.

As expected, our results indicate that white voters are more likely to have higher turnout at the county level and that Hispanic voters are less likely to turn out to vote at the county level, while the black variable is not associated with explaining variation in the dependent variable. As mentioned earlier, Geys (2006) shows mixed evidence on black voting as well: aggregate-level studies indicate lower turnout in areas with a higher proportion of blacks, but this may be due to the fact that as the black population increases, black turnout actually rises while white turnout (and thus overall turnout in the population) falls.

In terms of the aggregate level of education and how this affects voter turnout at the county level, we note that variables measuring educational attainment at the county level for both males and females have positive coefficient estimates. The point estimate for males is slightly higher than the estimate for females, but the credible interval for the difference between the male and female bachelor's degree holder variables contains zero, so that we cannot say that these coefficients are statistically different from each other.[45]

Of the remaining demographic variables, we note that the median age of residents, single-parent family, same house, and the urban variables are all associated with explaining variation in the dependent variable and all have the expected coefficient sign. Counties with older populations are more likely to turn out to vote, counties with more single-parent families have lower turnout, counties where the population has lived in the same house for five years are also more likely to turn out to vote, and the population as a whole in urban counties is less likely to turn out to vote.

Our economic variables are designed to capture any sentiment at the county level related to economic events, either good or bad. Of the four economic variables included in the estimating equation, the per-capita income at the county level and the real GDP growth rate from 2003 to 2004 are associated with explaining variation in the dependent variable. The unemployment rate at the county level and the state union membership, though they have the expected signs, do not appear to have any effect on voter turnout at the county level. These results seem to indicate that overall economic conditions are relevant for turnout at the county level whereas unemployment and union membership appear to play no role.

Our final category of independent variables, the political variables, has some surprising results. Of the variables that are included in the regression model, only the veterans, electoral votes, and the Cook political rating variables are associated with explaining variation in the

---

[45] In order to test if the point estimate for the male with a bachelor's degree variable was different from the point estimate for the female with a bachelor's degree variable, we took the Gibbs draws for each of these variables, subtracted them, and calculated a 95% credible interval, which equaled [-0.0842,1.7187]. Since the credible interval contains the value zero, we cannot say that there is any difference between the point estimates.

dependent variable. However, each of these variables has a sign that is consistent with our a priori expectations. The veterans variable has a positive coefficient estimate, indicating that as the percentage of veterans in a county increases, voter turnout in the county increases. Considering that each of the two candidates running in this particular election had some military experience, this result makes intuitive sense since veterans are more likely to turn out to vote if the candidates share a common characteristic with the voter. The electoral vote share variable, which is designed to capture "vote value," has a negative coefficient indicating that if the aggregate value of the vote is diluted, then it can suppress voter turnout. Finally, the Cook political rating, which measures how close the election is, has a positive coefficient estimate, which shows that as the race becomes more of a "toss up" between the candidates, county turnout is higher. Interestingly, we find no evidence that the various gay marriage ballot initiatives had an effect on voter turnout at the county level. Also, there appears to be no "home state" advantage for either candidate as the Texas and Massachusetts state dummy variables were not associated with explaining variation in the dependent variable. Overall, these results accord with our ex ante expectations.

Table 4 contains the results for the indirect effects for the spatial Durbin error model. The indirect effects measure how a change in an explanatory variable at location j ^{1} i affects the dependent variable at location i. These spatially weighted X variables are designed to capture any spillover effects from our set of explanatory variables. Out of our full set of 22 explanatory variables, only six of them have 95% credible intervals that do not span zero: religious adherents, Hispanic population, median age of the county population, female bachelor's degree holders, the urban population, and veterans.

The percentage of county population who are religious adherents has a positive coefficient estimate, indicating that as the population that self-identifies as a religious adherent in surrounding counties increases, an increase in voter turnout is seen in the home county. Interestingly, the direct effect for churches was a factor that affected voter turnout while the direct effect for adherents did not. We have the exact opposite result in terms of the indirect effects: the adherents variable is a factor that affects voter turnout while churches is not. This may reflect the fact that although churches are located in a single county, the county level percentage of the population that self-identify as religious adherents are not necessarily confined to a single county. People are mobile and may actually travel to another county to attend services, broadly defined. If they then associate with other like-minded congregants who live in neighboring counties, the county spillover effect on turnout may be more pronounced.

The percentage of the county population that is Hispanic has a positive spillover effect on county level voter turnout. Recall that the direct effect of this variable was negative, indicating a negative relationship between county level population of this demographic group and voter turnout, which was expected. Our indirect results indicate that as the Hispanic county level population increases in surrounding counties, the percentage of the own county population that turns out to vote decreases. At first, this observation may seem contradictory but may simply reflect demographic changes at the county level. If surrounding counties'

Hispanic populations are increasing, then it may be the case that the own county's population is becoming less Hispanic and this demographic change may increase voter turnout if other demographic groups are "replacing" the Hispanic population in the own county.

Of the two variables related to education, only the percentage of the county population who are female with a bachelor's degree variable is associated with explaining variation in the dependent variable. If the county population who are female with a bachelor's degree variable in surrounding counties increases, then the own county's voter turnout increases. One possible explanation for this result is that better educated females may be more politically active and that this activity spills over across county borders, due to media exposure or other publicity (akin to Cho and Rudolph's (2008) "casual observation" hypothesis) and this publicity may inspire or otherwise motivate individuals to head to the polls.

The positive coefficient estimate for veterans indicates that as the population of veterans in surrounding counties increases, we see an increase in county level voter turnout in the own county. One possible explanation for this result is that veterans may be more politically active than other demographic groups and that this political activity may be more readily noticed in surrounding counties and hence may boost turnout. Veterans, relative to other identifiable characteristics, may be more likely to know of and associate with other nearby veterans so the county spillover effect on turnout is more pronounced, similar to the religious adherents variable above.

Our last two indirect effects, the percentage of the county population living in urban areas and the median age of the resident county population, both have negative spillover effects. The spillover effect for the urban variable reinforces the direct effect due to the same sign however the median age variable has a negative indirect effect, which dilutes the direct effect somewhat.

The final set of results for the total effects is contained in Table 5. The total effects measure the sum of both the direct effects and the indirect effects.[46] The total effects measure how a change in an explanatory variable affects voter turnout inclusive of the own county and surrounding county spillover effects. The results for the total effects estimates can be broken down by examining the different groups of variables that we use in our analysis in a manner similar to our previous results. The total effects estimates as a whole have signs on the coefficient estimates that are in accordance with our a priori expectations for those variables that have credible intervals that do not contain zero.

Of the two variables that measure religious sentiment, only the percentage of the county population that are self-identified religious adherents matters in terms of county turnout. This result seems to reinforce the idea that religious sentiment is more important in explaining turnout than the actual number of churches per 10,000 county residents. Adherents to one of

---

[46] The total effects were calculated by summing the Gibbs draws for each explanatory variable for their respective b and q coefficient draws and then a 95% credible interval was calculated for each variable.

many religious groups as defined in the Glenmary survey may have a more intense interest in public affairs or have a more nuanced sense of civic duty than non-religious types and hence appear to have higher turnout, all else equal.

As was the case for the direct estimates, the white and Hispanic variables are associated with explaining variation in the dependent variable, while the variable that measures the county population that is black does not. Additionally, the signs are in accordance with our ex ante expectations, i.e. an increase in the white population increases voter turnout while an increase in the Hispanic population decreases turnout.

The total effects also indicate that when the percentage of the county population that has a bachelor's degree increases (regardless of gender), there is an increase in voter turnout. It appears that education plays a major role in explaining voter turnout at the county level.

Of the remaining demographic variables, the county urban population and the percentage of the county population living in the same house are associated with voter turnout, but in opposite directions. The urban population has a negative coefficient estimate which may indicate that counties with large urban areas experience a dilution in the weight attached to their vote. Even though the value of an individual vote is quite low, non-urban voters on election day may not have many social or recreational activities in which to participate, so the opportunity cost of voting is quite low. Urban voters, however, consider the opportunity cost of voting as quite high in terms of forgone opportunities for entertainment, shopping, dining out, etc., relative to their vote value. This result is in contrast to the result found by Geys (2006) which established that urbanization and voter turnout were weakly associated.

We turn our attention now to the economic variables, and the total effects results indicate that the only economic variable that is associated with the dependent variable is the state unionization variable, which has a positive coefficient estimate. The per-capita income, unemployment, and real GDP growth rate appear to have no effect on voter turnout at the county level in our model. Given the prominence that labor unions place on political efforts and voter turnout drives,[47] this result should not be too surprising. However, it is surprising that none of the other economic variables in terms of total effects are associated with explaining voter turnout at the county level.

Our final group of variables that we examine are the political variables. Of the seven political variables that we utilize in our model specification, only the veterans, electoral votes in a county, and Cook closeness rating are associated with our dependent variable.

The percentage of the county population that are veterans has a positive association with voter turnout. In the 2004 election, a group known at the Swift Boat Veterans for Truth ran several television advertisements that were highly critical of the candidacy of John Kerry, who made his military service a major part of his election campaign. Regardless of the veracity

---

[47] See Delaney, Masters, and Schwochau (1988) and Zullo (2004).

of the claims made by the group, this kind of grass-roots organizing may indicate the power that veterans groups (broadly defined) have in terms of stimulating voter turnout, much like the influence that labor unions have in organizing voter turnout.

As we mentioned in the section explaining our choice of political variables, we proxied "vote value" in two ways, one of which was that voters assign a weight to their Presidential vote in proportion to the number of electoral votes assigned to their state since electoral votes will mirror population. Our results indicate that our a priori expectation of a negative relationship between this county electoral vote share and county turnout is confirmed.

Finally, we attempted to proxy for the closeness of the election using the Cook Electoral Rating from the Cook Political Report. Recall that a state that was "solid" for Bush or Kerry was coded as 0; "likely" Bush or Kerry was coded as 1; "leans" Bush or Kerry as 2; and a "toss up" as 3. Given our coding, we expected a positive association between this explanatory variable and county voter turnout. The coefficient estimate for this variable indicates that there is indeed a positive relationship between the Cook report rating and voter turnout at the county level. It appears that states that were deemed "toss up" states had higher turnout at the county level showing that voter turnout can be a function of how close the race is perceived *ex ante*.

Conclusion

The amount of literature on voter turnout is extensive mostly because the presumed benefits and costs of voting are so small that slight disturbances can alter turnout in seemingly unpredictable ways. This paper adds to the existing literature by comparing four different possible models of voter turnout at the county level: the normal linear model, the spatially lagged X (SLX) model, the spatial Durbin model, and the heretofore unused spatial Durbin error model. Using modern Bayesian econometric estimation and model comparison techniques, our results overwhelmingly provide support for the spatial Durbin error model, which provides inferences regarding direct, indirect, and total effects of the set of explanatory variables on the dependent variable. Quantifying these various effects may be important for policy reason as well. For example, the direct effects show how a change in an explanatory variable in one county affects voter turnout in that same county, which may be of interest to certain stakeholders, such as voter turnout advocates. However, our preferred model also shows that there may be indirect effects also associated with changes in explanatory variables that affect voter turnout. For example, the percentage of the county population who are veterans had a positive direct effect which shows that as the county level veteran population

increases, we see an increase in voter turnout in neighboring counties. This result indicates that there may be additional benefits to voter turnout drives that extend beyond a county's border.

The results from the spatial Durbin error model also shows that many independent variables commonly thought to influence voter turnout actually are not associated with voter turnout at the county level at all, including some rather surprising ones such as per-capita income and the county unemployment rate. Additionally, some commonly-held opinions such as the effect of a gay marriage ballot initiative to spur voter turnout prove not to have any empirical support.

Spatial econometric techniques are seeing increasing use in various fields such as political science and economics, to name but two. These models in general, and the spatial Durbin error model we use in this paper in particular, can provide much more detailed empirical findings for researchers who are interested in quantifying spillover effects and how these effects can guide policy for various stakeholders.

Table 1: Summary Statistics

	Mean	Stan. Dev.	Source
Turnout, Voting-Age Pop.	0.5804	0.0904	Polidata, Census Population Estimates
Turnout, Citizen Voting-Age Pop.	0.6152	0.0940	Polidata, Census Population Estimates
% religious adherents	0.5316	0.1828	Glenmary Research Center
churches per 10,000 people	22.0993	13.0318	Glenmary Research Center
% white	84.9912	15.8983	Census 2000 SF1
% black	8.6330	14.3757	Census 2000 SF1
% Hispanic	6.2098	12.0989	Census 2000 SF1
median age of resident	38.4284	4.2668	Census Population Estimates
% single-parent family 2000	0.2743	0.0763	Census 2000 SF1
% in same house, 1995-2000	0.5905	0.0734	Census 2000 SF3
% male pop. with bachelor's	0.1089	0.0510	Census 2000 SF3
% female pop. with bachelor's	0.1093	0.0483	Census 2000 SF3
% urban population	0.3940	0.3035	Census 2000 SF1
ln (per capita income)	10.1514	0.2085	Bureau of Economic Analysis

unemployment rate	5.6647	1.7966	Bureau of Labor Statistics
real GDP growth '03-'04	0.0354	0.0176	Bureau of Economic Analysis
state unionization	0.0999	0.0520	Bureau of Labor Statistics
% veterans	0.1397	0.0290	Census, Dept. of Veterans Affairs
gay marriage dummy	0.2803	0.4492	CNN 2004 Election website
county share of state electoral votes	0.1725	0.5006	Federal Election Commission, Census Population Estimates
age 18 and over pop. growth, 2003-2004	0.0102	0.0165	Census 2000 SF1, Census Population Estimates
Cook closeness rating	0.8971	1.1741	CNN 2004 Election website
Texas state dummy	0.0827	0.2754	
Massachusetts state dummy	0.0046	0.0675	

N = 3,061

Table 2: Log-Marginal Likelihood Values and Posterior Model Probabilities

Model	Log-Marginal Likelihood Value	Posterior Model Probability
Normal Linear	-102.84 (2.2138e-5)	≈0
Spatially Lagged X	-136.00 (4.1693e-5)	≈0
Spatial Durbin	379.26 (.1964)	≈0
Spatial Durbin Error	391.59 (.0052)	≈1

Log-marginal likelihood values are calculated via the method of Chib and Jeliazkov (2001). Numbers in parentheses represent numerical standard errors for the log-marginal likelihood calculation. Calculations are based on 10,000 sampled values.

Table 3: Direct Effects from the Spatial Durbin Error Model

Direct Effect	Mean	Lower 95%	Upper 95%
adherents	0.0499	–0.0125	0.1121
churches	**0.0026**	**0.0016**	**0.0037**
white	**0.0025**	**0.0009**	**0.004**
black	0.0016	–0.0001	0.0032
Hispanic	**–0.0072**	**–0.0087**	**–0.0057**
median age	**0.0139**	**0.0102**	**0.0176**
single-parent	–0.1854	–0.4092	0.0367
same house	**1.4944**	**1.2868**	**1.6999**
male with ba	**2.3636**	**1.886**	**2.834**
female with ba	**1.5803**	**1.0921**	**2.0743**
urban	**–0.1833**	**–0.2285**	**–0.1384**
ln (per cap income)	**0.1879**	**0.1193**	**0.2581**
unemploy rate	0.0027	–0.0039	0.0095
rGDP growth	**2.1355**	**0.95**	**3.3374**
unionization	0.2325	–0.3318	0.7996
% veterans	**0.5161**	**0.1153**	**0.9191**
gay marriage vote	0.0057	–0.0364	0.0476
electoral vote share	**–0.0313**	**–0.0494**	**–0.0129**
population growth	–0.036	–0.5795	0.5223
Cook closeness	**0.0509**	**0.0327**	**0.0686**
Texas dummy	–0.0661	–0.1789	0.0491
Mass. dummy	0.0179	–0.1388	0.1792

Parameter inferences are based on 10,000 sampled values with 10,000 sampled values used as burn-in. Lower 95% and Upper 95% represent the lower and upper 95% credible interval bounds, respectively. Entries in bold represent credible intervals that do not contain zero.

Table 4: Indirect Effects from the Spatial Durbin Error Model

Indirect Effect	Mean	Lower 95%	Upper 95%
adherents	**0.1858**	**0.0436**	**0.3258**
churches	-0.0021	-0.0048	0.0005
white	0.0026	-0.0008	0.006
black	0.0001	-0.0033	0.0034
Hispanic	**0.0036**	**0.001**	**0.0061**
median age	**-0.0096**	**-0.0185**	**-0.0007**
single-parent	0.4919	-0.0412	1.0429
same house	0.1127	-0.3593	0.5831
male with ba	-0.3816	-1.6744	0.8866
female with ba	**2.3506**	**1.1397**	**3.5607**
urban	**-0.1725**	**-0.2931**	**-0.0523**
ln (per cap income)	-0.079	-0.2533	0.0964
unemploy rate	0.0131	-0.0021	0.0282
rGDP growth	-1.1671	-2.7782	0.4634
unionization	0.676	-0.0489	1.4055
% veterans	**1.4349**	**0.4317**	**2.419**
gay marriage vote	0.0225	-0.0382	0.0831
electoral vote share	-0.046	-0.0963	0.006
population growth	0.6666	-0.8289	2.1523
Cook closeness	0.0002	-0.0243	0.0251
Texas dummy	0.0034	-0.1409	0.144
Mass. dummy	-0.1214	-0.4155	0.1734

Parameter inferences are based on 10,000 sampled values with 10,000 sampled values used as burn-in. Lower 95% and Upper 95% represent the lower and upper 95% credible interval bounds, respectively. Entries in bold represent credible intervals that do not contain zero.

Table 5: Total Effects from the Spatial Durbin Error Model

Total Effect	Mean	Lower 95%	Upper 95%
adherents	**0.2357**	**0.0745**	**0.3928**
churches	0.0005	-0.0025	0.0036
white	**0.005**	**0.0012**	**0.0088**
black	0.0016	-0.0019	0.0052
Hispanic	**-0.0036**	**-0.0062**	**-0.001**
median age	0.0043	-0.0059	0.0143
single-parent	0.3065	-0.3225	0.9442
same house	**1.607**	**1.0701**	**2.1427**
male with ba	**1.9819**	**0.438**	**3.5041**
female with ba	**3.9308**	**2.4758**	**5.3673**
urban	**-0.3558**	**-0.4963**	**-0.2152**
ln (per cap income)	0.1088	-0.0938	0.3132
unemploy rate	0.0158	-0.0012	0.0323
rGDP growth	0.9684	-0.2803	2.2133
unionization	**0.9085**	**0.4038**	**1.4188**
% veterans	**1.951**	**0.8046**	**3.097**
gay marriage vote	0.0282	-0.0204	0.0762
electoral vote share	**-0.0773**	**-0.1349**	**-0.0173**
population growth	0.6306	-1.1291	2.443
Cook closeness	**0.0511**	**0.032**	**0.0703**
Texas dummy	-0.0627	-0.1538	0.0267
Mass. dummy	-0.1035	-0.3825	0.1767

Parameter inferences are based on 10,000 sampled values with 10,000 sampled values used as burn-in. Lower 95% and Upper 95% represent the lower and upper 95% credible interval bounds, respectively. Entries in bold represent credible intervals that do not contain zero.

References

Aldrich, John H. "Rational Choice and Turnout." *American Journal of Political Science* 37, no. 1 (Feb. 1993): 246–78.

Ashenfelter, Orley, and Stanley Kelley, Jr. "Determinants of Participation in Presidential Elections." *Journal of Law and Economics* 18, no. 3 (Dec. 1975): 695-733.

Barzel, Yoram and Eugene Silberberg. "Is the Act of Voting Rational?" *Public Choice* 16 (1973): 51-58.

Carter, John R. "Early projections and voter turnout in the 1980 presidential election." *Public Choice* 43 (1984): 195-202.

Chib, Siddhartha. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90, no. 432 (Dec. 1995), pp. 1313-1321.

Chib, Siddhartha and Ivan Jeliazkov. "Marginal Likelihood from the Metropolis-Hastings Output." *Journal of the American Statistical Association* 96, no. 453 (Mar. 2001), pp. 270-281.

Cho, Wendy K. Tam and Thomas J. Rudolph. "Emanating Political Participation: Untangling the Spatial Structure Behind Participation." *British Journal of Political Science* 38, no. 2 (2008): pp. 273-289.

Cutts, David and Don. J. Webber. "Voting Patterns, Party Spending and Relative Location in England and Wales." *Regional Studies* 44, no. 6 (July 2010): 735-760.

Darmofal, David. "The political geography of macro-level turnout in American political development." *Political Geography* 25 (2006): 123-150.

Delaney, John T., Marick F. Masters, and Susan Schwochau. "Unionism and Voter Turnout." *Journal of Labor Research* 9, no. 3 (Summer 1988): 221-236.

Downs, Anthony. *An Economic Theory of Democracy*. New York: Harper Collins, 1957.

Elhorst, J. Paul. "Applied Spatial Econometrics: Raising the Bar." *Spatial Economic Analysis* 5, no. 1 (2010): 9-28.

Feddersen, Timothy J. "Rational Choice Theory and the Paradox of Not Voting." *Journal of Economic Perspectives* 18, no. 1 (2004): 99-112.

Filer, John E. and Lawrence W. Kenney. "Voter turnout and the benefits of voting." *Public Choice* 35 (1980): 575-585.

Gelfand, A.E., S.E. Hills, A. Racine-Poon and A.F.M. Smith. "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling." *Journal of the American Statistical Association* 85 (1990): 972-985.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. "Social Pressure and Voter Turnout: Evidence from a Largescale Field Experiment." *American Political Science Review* 102, no. 1 (February 2008): 33-48.

Geys, Benny. "Explaining voter turnout: A review of aggregate-level research." *Electoral Studies* 25, no. 4 (2006): 637-663.

Hackey, Robert B. "Competing explanations of voter turnout among American blacks." *Social Science Quarterly* 73, no. 1 (March 1992): 71-89.

Key, V.O. Jr. *Southern Politics in State and Nation*. Knoxville, TN: The University of Tennessee Press, 1984.

Kim, Jeongdai, Euel Elliott, and Ding-Ming Wang. "A spatial analysis of county-level outcomes in US Presidential elections, 1988-2000." *Electoral Studies* 22, no. 4 (Dec. 2003): 741-761.

Knack, Steve. "Does rain help the Republicans? Theory and evidence on turnout and the vote." *Public Choice* 79 (1994): 187-209.

Koop, Gary M. Bayesian Econometrics. Chichester, West Sussex, England: Wiley, 2003.

Lacombe, Donald J, "An Introduction to Bayesian Inference in Spatial Econometrics" (July 24, 2008). Available at SSRN: http://ssrn.com/abstract=1244261

Lacombe, Donald J. and Timothy M. Shaughnessy. "Accounting for Spatial Error Correlation in the 2004 Presidential Popular Vote." *Public Finance Review* 35, no. 4 (July 2007): 480-499.

LeSage, James P and R. Kelley Pace. *Introduction to Spatial Econometrics*. Boca Raton: CRC Press, 2009.

Matsusaka, John G. and Filip Palda. "Voter turnout: How much can we explain?" *Public Choice* 98 (1999): 431-446.

Matsusaka, John G. "Explaining voter turnout patterns: An information theory." *Public Choice* 84 (1995): 91-117.

-----. "Election closeness and voter turnout: evidence from California ballot propositions." *Public Choice* 76 (1993): 313-334.

McDonald, Michael P. "The Turnout Rate among Eligible Voters in the States, 1980-2000." *State Politics and Policy Quarterly* 2, no. 2 (Summer 2002): 199-212.

Merrifield, John. "The institutional and political factors that influence voter turnout." *Public Choice* 77 (1993): 657-667.

Nagel, Jack H. and John E. McNulty. "Partisan effects of voter turnout in senatorial and gubernatorial elections." *American Political Science Review* 90, no. 4 (Dec. 1996): 780-793.

Nickerson, David W. "Is Voting Contagious? Evidence from Two Field Experiments." *American Political Science Review* 102, no. 1 (Feb. 2008): 49-57.

O'Loughlin, John, Colin Flint, and Luc Anselin. "The Geography of the Nazi Vote: Context, Confession, and Class in the Reichstag Election of 1930." *Annals of the Association of American Geographers* 84, no. 3 (Sep., 1994): pp.351-380.

Oberholzer-Gee, Felix, and Joel Waldfogel. "Electoral acceleration: the effect of minority population on minority voter turnout." NBER working paper 8252 (April 2001).

Pattie, Charles and Ron Johnston. "'People Who Talk Together Vote Together': An Exploration of Contextual Effects in Great Britain." *Annals of the Association of American Geographers* 90, No. 1 (Mar., 2000): pp.41-66.

Riker, William H. and Peter C. Ordeshook. "A Theory of the Calculus of Voting." *American Political Science Review* 62, no. 1 (March 1968): 25-42.

Robinson, W. S. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15, no. 3 (June 1950): pp. 351-357.

Shachar, Ron and Barry Nalebuff. "Follow the Leader: Theory and Evidence on Political Participation." *American Economic Review* 89, no. 3 (June 1999): 525-547.

Stigler, George J. "Determinants of Participation in Presidential Elections: Comment." *Journal of Law and Economics* 18, no. 3 (Dec. 1975): 743-744.

Timpone, Richard J. "Structure, behavior, and voter turnout in the United States." *American Political Science Review* 92, no. 1 (March 1998): 145-158.

Tollison, R.D. and T. D. Willett. "Some Simple Economics of Voting and not Voting." *Public Choice* 16, no. 1 (1973): 59-71.

Wittman, Donald. "Determinants of Participation in Presidential Elections: Comment." *Journal of Law and Economics* 18, no. 3 (Dec. 1975): 735-741.

Zullo, Roland. "Labor Council Outreach and Union Member Voter Turnout: A Microanalysis from the 2000 Election." *Industrial Relations* 43, no. 2 (April 2004): 324-338.