THE HONORABLE JOHN E. BRIDGES

SUPERIOR COURT OF THE STATE OF WASHINGTON
FOR CHELAN COUNTY

Timothy Borders et al.,

Petitioners,

v.

King County et al.,

Respondents,

and

Washington State Democratic Central Committee,

Intervenor-Respondent.

NO. 05-2-00027-3

DECLARATION OF DAVID J. BURMAN REGARDING WSDCC EXPERT ADOLPH

DAVID J. BURMAN declares under penalty of perjury as follows:

1. I am one of the attorneys representing Intervenor-Respondent Washington State Democratic Central Committee ("WSDCC") in this litigation. I am competent to

Perkins Coie LLP
1201 Third Avenue, Suite 4800
Seattle, Washington 98101-3099
Phone: (206) 359-8000
Fax: (206) 359-9000

testify and make this declaration based on my personal knowledge and the files and records in this matter.

2. Pursuant to the Court's ruling of May 2, 2005, WSDCC hereby submits the report of WSDCC's expert Christopher Adolph, Ph.D.

3. Due to the late disclosure of various supplemental reports by Petitioners' experts and supporting and related materials, and because the Court may reject certain categories of allegedly invalid ballots or challenges to individual allegedly invalid ballots, Respondent WSDCC reserves the right for its experts to apply the basic approaches set forth in their reports to the evidence submitted at trial.

SIGNED at Seattle, Washington, this 12th day of May, 2005

s/ David J. Burman
DAVID J. BURMAN

Perkins Coie LLP
1201 Third Avenue, Suite 4800
Seattle, Washington 98101-3099
Phone: (206) 359-8000
Fax: (206) 359-9000

Report on the 2004 Washington Gubernatorial Election

Christopher Adolph

May 12, 2005

1 Qualifications

I am an assistant professor of political science at the University of Washington, Seattle. I am also a core faculty member of the Center for Statistics and Social Sciences, and an adjunct assistant professor of statistics. I have been hired by legal counsel as a private individual for my relevant expertise, and do not represent the university.

I have an extensive background assisting consultants and expert witnesses in conducting statistical analyses of voting behavior relevant to numerous legal proceedings on legislative redistricting and Voting Rights Act challenges. I have published several scholarly articles on ecological inference, the key statistical issue in this lawsuit. I also conduct research, teach courses, and advise graduate students in the broader field of quantitative political methodology.

I received my Bachelor of Arts degree from Rice University, and my Master of Arts and Doctor of Philosophy degrees, both in political science, from Harvard University. My curriculum vitae is provided as an appendix to this report.

2 Introduction

I have been asked by legal counsel to evaluate the claims in the expert reports of Jonathan Katz and Anthony Gill, and to assess the ability of statistical techniques to uncover the vote choices of invalid voters in the 2004 Washington gubernatorial election.

The 2004 Washington gubernatorial election was extraordinarily close, with Democratic candidate Christine Gregoire narrowly defeating Republican Dino Rossi by 129 votes in the final manual recount, out of 2,810,058 votes cast (0.005 percent).[1] The petitioners argue that the outcome of the election would have been different if 1,183 ballots which they allege to be invalid were not counted. Petitioners claim the invalid ballots consist of those cast by disenfranchised felons, the deceased, people who voted in multiple states or precincts, non-citizens, and people casting invalid provisional ballots. It is important to remember that invalid ballots comprise a very small fraction of the total votes cast (0.042 percent, if we consider the petitioners' 1,183

[1]Figures in this paragraph rely on data from the Washington Secretary of States office, and do not include votes for write-in candidates, under- and over-votes, or spoiled ballots.

invalid ballots). In their reports, both Gill and Katz claim that it is possible to say how this small minority of voters cast their secret ballots, that a substantial majority of them cast ballots for Gregoire, and that the election outcome would have been different had these ballots not been counted.

I have reviewed the reports and data used by Gill and Katz, as well as new data provided by the respondents. Based on my own analyses of these data, I have three main conclusions:

- Neither Katz nor Gill employ currently accepted methods for solving the problem at hand, known in social science as the *ecological inference* problem.
- Application of currently accepted methods of ecological inference would inform us that the behavior of invalid voters in this election cannot be inferred from the aggregate data Katz and Gill use, or indeed, from any data currently available.
- Even if one accepts the (flawed and non-standard) methods used by Gill and Katz, their own methods find the election outcome would not change, once problems with their data are corrected.

The issues in this case turn on subtle but important—and in many cases, very long recognized—social science and statistical concepts. To help the court wade through these issues, I will proceed through my review of Gill and Katz, and my re-analysis of the problem, in three stages. In each of these three stages, I illustrate a different fatal problem in these experts' work. Each of these problems completely undermines the results offered by Katz and Gill. Rather than stop after the first issue, in each section, I ask "If the previous issues were somehow corrected or ignored, would Gill and Katz's findings still stand?" In each case, I find it would not.

The three independent problems in Gill and Katz reports are:

1. Use of a non-random, non-representative, and incomplete sample of invalid votes, which is useless for answering questions about the net effect of all invalid votes on the state-wide election outcome.
2. Use of a flawed and error-prone method of ecological inference. Modern, accepted methods would provide strong warnings to the researcher that ecological inference in this case is impossible.
3. Combination of flawed methods and unrepresentative data. Applying Katz and Gill's methods to more complete data results in no change in the election outcome from excluding invalid votes.

As I show below, correcting *any* of these flaws would show that invalid ballots cannot be found to have affected the election outcome with any certainty.

3 Serious flaws in the petitioners' data undermine their claims

The methods adopted by Katz and Gill will almost surely fail to produce accurate estimates unless applied to a full census, random sample, or representative sample of invalid ballots. For example, if there were a systematic bias in the collection of their data, such that invalid ballots were more likely to be included if they were in certain counties or precincts, then we would say that their sample is not representative, nor randomly drawn, nor a complete census. It would be systematically biased, and that bias would carry over to their results. In particular, Gill's and Katz's estimates of the likely impact of invalid votes, and Katz's confidence intervals, require an unbiased sample to produce meaningful results. (Of course, an unbiased sample does not guarantee meaningful results by itself, as we shall see below when we consider the virtues of Gill and Katz's methodology.)

3.1 Petitioners' data are a non-random sample

Are the petitioners' data a census, a random sample, or a representative sample of invalid ballots? Neither Gill nor Katz claim they are. This omission is unusual and conspicuous. The first thing another scholar would ask about these reports is "Where did the data come from? Are the data a valid sample?" Unless the petitioners can make and support the claim that these data are a representative sample of invalid ballots across the whole state, it will be impossible to make even minimally valid scientific claims regarding the likely effect of these alleged invalid ballots on the election outcome. In particular, if petitioners over-sampled precincts that voted overall for Gregoire, then their method will tend to produce biased results suggesting, perhaps incorrectly, that Gregoire benefited from the inclusion of invalid votes.

To see if the petitioners' data appeared to be a representative sample, I first divided Washington voting precincts into four categories: those voting heavily for Gregoire (i.e., where Gregoire's margin over Rossi was at least 23 percentage points), those moderately favoring Gregoire (i.e., where Gregoire's held a margin of less than 23 percentage points), those moderately favoring Rossi (i.e., where Rossi held a margin less than 18 percentage points), and those heavily favoring Rossi (i.e., where Rossi's margin was greater than 18 points). Each of these categories roughly corresponds to a quartile of Washington voters.

Table 1 shows that the large majority of petitioners' invalid ballots were found in pro-Gregoire precincts, and especially in heavily Gregoire precincts. By itself, this is not *prima facie* evidence of cherry picking, since we do not know the true distribution of invalid votes. However, because the petitioners have an interest in finding invalid ballots precisely where they did find them, and because their experts resolutely refuse to vouch for the data, we should proceed cautiously.

The easiest way to check whether petitioners have systematically over-sampled invalid ballots from Gregoire precincts is to look for missing invalids in Rossi precincts. Legal counsel have provided me with a list of 743 felons, omitted from the petitioners' list, who appear to have cast ballots in the 2004 Washington gubernatorial election.[2] Most of these felons cast ballots in

[2]The total number of invalid ballots, and the standards for what constitutes an invalid ballot, appear

Precincts ranked by margin	% of ballots cast	Invalid ballots found by petitioners	Invalid ballots overlooked by petitioners
Gregoire by 23% or more	24.8 %	541	2
Gregoire by less than 23%	27.8	308	7
Rossi by less than 18%	22.7	251	234
Rossi by 18% or more	24.6	83	498

Table 1: Petitioners' invalid vote sample appears skewed towards precincts voting heavily for Gregoire. Petitioners found far more invalid ballots per ballot cast in precincts that voted heavily for Gregoire. Such a distribution is *a priori* possible. However, respondents have found large numbers of invalid ballots in precincts which Rossi won, strongly suggesting that the petitioners' data are neither a random sample, nor a representative sample, nor a census of invalid ballots. (The column for respondent data is missing two Benton County ballots for which I lack precinct voting data).

precincts leaning towards Rossi.

Adding these new data helps balance the petitioners' skewed dataset. For what it is worth, the combined data appear roughly uniformly distributed across the precincts. But we still cannot be confident that the data gathered collectively by both sides is either a census or representative of the whole state. As a result, we should view with skepticism any claims that the election might have turned out differently, had invalids been excluded, based on these data.

3.2 Petitioners' methods fail when applied to a non-random sample

Gill and Katz's methods assume that invalid votes were cast in the same proportion as other votes in a given precinct. Thus if we had a complete statewide census of invalid votes, but excluded from the analysis invalid votes cast in precincts won by Dino Rossi, the Katz-Gill method would conclude that Gregoire benefited from the inclusion of invalid votes. On the other hand, if precincts Christine Gregoire won were excluded, it would appear that Rossi benefited from the invalid votes, as Gill illustrates when he excludes King County from his analysis. Only a complete statewide census, or a random or representative sample, has any hope of offering an unbiased estimate of the effect of invalid votes.

To the extent we are unsure that the petitioner's data is truly complete or representative—or, to put it another way, to the extent the petitioners have cherry-picked invalid votes in Gregoire country—we should strongly discount the conclusions of the Gill and Katz analyses. Only if the Rossi campaign put as much effort into searching for invalid votes in precincts they won (where, by their own logic, finding such votes would hurt them) as in precincts they lost, can we begin to trust their experts' reports.

to be evolving as the case progresses. All the methods in this report can be re-run with no essential changes on revised data as they become available.

		Voting Decision			
		Gregoire	Rossi	Other	
Status of	Invalid	?	?	?	1,926
Individual Ballots	Valid	?	?	?	2,808,132
		1,373,361	1,373,232	63,465	2,810,058

Table 2: The Ecological Inference Problem applied to invalid voters. The goal of ecological inference is to infer from the marginal entries (each of which is the sum of the corresponding row or column) to the cell entries. It is in general a very difficult problem. To succeed, we will need either tight deterministic bounds, or strong assumptions. (Voting data above are from the Washington Secretary of States office, and do not include votes for write-in candidates, under- and over-votes, or spoiled ballots. These ballots could easily be lumped in with the Other category, however. Invalid ballots are the sum of those claimed by petitioners and respondents.)

The possibility that some invalid ballots are missing in a non-random way directly undermines a key thesis of Gill's report. Gill claims that there are "a sufficient number of uncontested invalid ballots in King County alone to believe the 2004 Washington State gubernatorial election may have resulted in a victory for Mr. Dino Rossi had those invalid ballots not been cast ..." (Gill, Supplemental Report, page 2). Because Gill's own method is to deduct a proportion of invalid ballots from the candidates' totals based on their precinct-wide support, the net change in the statewide margin of victory will depend, again according to Gill's method, on the number of invalid ballots cast in each and every precinct in the state. Because Gill provides no evidence or even any assurance that his data are complete, we cannot discount the possibility that further invalid ballots may be discovered outside of King County, and in particular in precincts won by Dino Rossi. Under Gill's own method, such discoveries would mitigate or overwhelm the effects of invalid ballots cast in King County. For Gill to suggest that conclusions on statewide outcomes may be drawn solely on the basis of King County is intuitively and statistically preposterous. Furthermore, this is not merely a hypothetical point: the data used by Gill appear to have systematically missed invalid ballots cast in precincts won by Rossi, and inclusion of these ballots would change the results of his analysis, as we shall see below.

So far we have seen that the petitioner's data are suspect. In the next section, we will see that their methods are equally dubious.

4 Petitioner's experts commit the ecological fallacy

The statistical problem in this case can be illustrated with a simple table. Table 2 shows what is known and unknown. From the final recount, we know the total number of ballots cast for Gregoire, for Rossi, or for other candidates. Setting aside our doubts about the completeness of the data, we "know" the number of ballots which were invalid or valid. These numbers form the

row and column sums, or marginals, of a 2×3 table, the interior cells of which are unknown. The *ecological inference problem* concerns making inferences from the marginal totals to the contents of the cells. It has long been recognized as a very difficult problem.

Unfortunately, it is the problem we face. The question is simply which candidate got the most valid votes, or in terms of the diagram, which box in the lower row of "?"'s is largest. Therefore, the scientific problem before us is intrinsically the ecological inference problem, and any method failing to solve the ecological inference problem will give a dubious and scientifically unacceptable estimates of the election outcome.

Rather than face the ecological inference problem head on, Gill and Katz assume it away. At the core of the reports of Gill and Katz is a single, critical assumption: that the vote choices of any sub-population within a voting precinct can be assumed to be the same as the average vote choice in the precinct as a whole. This assumption is strong, implausible, and unwarranted. Moreover, it flies in the face of decades of warnings from social scientists and statisticians to avoid committing the "ecological fallacy".

In this section, I first explain the ecological fallacy and efforts to overcome it. I then apply modern methods to the problem of inferring the voting behavior of Washington's invalid voters, and show that the aggregate data used by Katz and Gill cannot reveal individual voters' choices. Finally, I address the impact of the ecological inference problem on the specific methods used by Gill and Katz.

4.1 Examples of the ecological fallacy

Since the publication of a landmark paper by William S. Robinson in 1950, social scientists have recognized the hazards in making inferences about a particular individual's behavior using only information about the average behavior of groups (Robinson, 1950).[3] Indeed, it is only in desperation at a lack of individual data that one would attempt ecological inference, and it is intuitively obvious that such inferences will very often be badly wrong. This danger is known as the *ecological fallacy*, and it is at the heart of this case.

An example will help illustrate the fallacy. Suppose you were asked to estimate the batting average last season of a particular Seattle Mariner, Ichiro Suzuki, but you know only that the team as a whole had a batting average of 0.270 in 2004. If we were to leap from the fact that 0.270 was exactly the American League average to the conclusion that all Mariners had mediocre seasons at the plate, we would make a badly mistaken inference. In fact, Ichiro Suzuki batted 0.372, the highest average of anyone in the American League.[4]

The study of voting is rife with examples of the ecological fallacy. We need look no further

[3] Robinson's paper is one of the most cited in all social science, and the second most cited in the history of the *American Sociological Review* (Jacobs, 2005). Robinson noted that at the aggregate level, the number of immigrants and the number of literate residents in a state appear to be correlated. This might lead a researcher to mistakenly infer that immigrants are more likely to literate than those born in the US, when at the time the reverse was known to be true from surveys of individuals.

[4] Data taken from `www.espn.com`.

than the 2004 presidential election. After the election outcome was announced, some commentators in the mass media noted that the so-called "red" states—which voted, on average, for President George W. Bush—tend to have lower average incomes than the "blue" states—which voted, on average, for challenger John Kerry. These commentators were puzzled by this aggregate behavior, because it seemed to contradict the well-known tendency of individual high-income voters to be more likely to vote for Republican candidates. Some wondered whether the aggregate pattern implied that the connection between income and partisanship had broken down, or even reversed itself.

A group of scholars, led by statistician/political scientist Andrew Gelman, point out that this "inference" is simply the ecological fallacy, still rearing its head 55 years after its discovery (Gelman et al., 2005). Gelman et al first note that the aggregate Republican vote share appears strongly inversely related to average state income; see, for example, the cases of Mississippi, Ohio, and Connecticut in Figure 1A. On the basis of this aggregate relationship, careless observers made inferences about the behavior of individual voters. However, by analyzing individual level data from exit-polls, Gelman et al found that, in each state, higher income voters were more likely to vote for Bush, compared to lower income voters, as the upward sloping lines in Figure 1B show. But because the *average* voter in Mississippi was more likely to vote for Bush than the *average* voter in Connecticut, the aggregate data masked this clear and strong individual-level relationship (Figure 1C).

The lesson is clear: if we eschew or lack individual level data, and attempt to make inferences from aggregates, we risk being badly wrong; even completely backwards in our conclusions. To be sure, the recognition of the ecological fallacy in 1950 did not completely stop unwitting social scientists, journalists, courts, or ordinary folks from making unwarranted inferences about individual behavior from aggregate data, and it would not be difficult to find news articles, legal opinions, or even articles published in prestigious journals which commit the ecological fallacy. Still, it is a fallacy, and when recognized in scholarship, tends to be fatal to the publication prospects of a scientific study.

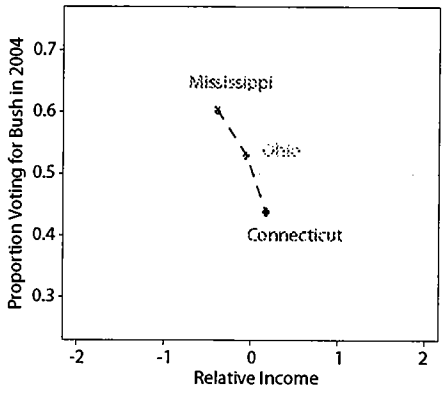
4.2 Early methods of ecological inference

After Robinson identified the ecological fallacy, the race was on to find methods of ecological inference that would avoid or overcome it. Until the late 1990s, there were two competing methods, neither particularly satisfactory. The first, known as the method of bounds (Duncan and Davis, 1953), noted that ecological data sometimes place logical limits on the range of individual data. To understand the method of bounds, it helps to have an example.

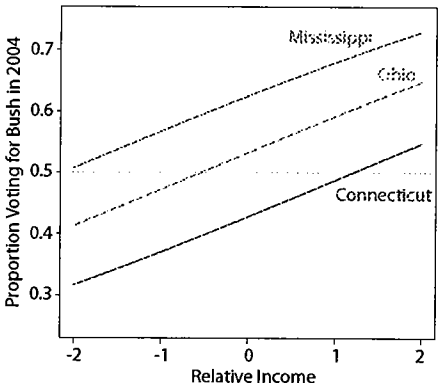
Suppose we were studying racial patterns of voting in a electoral district with a voting age population consisting of 40,000 black adults and 60,000 non-black adults. Suppose further than 80,000 people turned out to vote. Can we say what proportion of black and non-black adults voted?

Not with any precision. But we can state a logical range bounding the possible answers to the question. We start with the following accounting identity, which is mathematically true no

A. State-level averages of income and vote choice, and apparent inverse relationship between income and votes for Republican candidate Bush



B. Individual-level data on income and vote choice, and the actual *positive* relationship between income and votes for Republican candidate Bush



C. The ecological fallacy revealed: Aggregate data are consistent with, but give no obvious indication of, the true individual level data & relationship

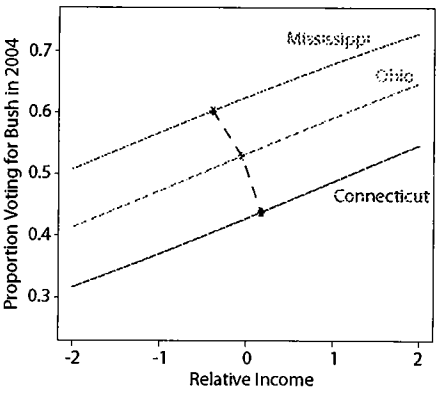


Figure 1: An illustration of the ecological fallacy from the 2004 presidential election. The essence of the ecological fallacy is that aggregate data often are a very misleading guide to underlying relationships. Source: Hierarchical logistic regression from Gelman et al. (2005).

matter what for each precinct:

$$\begin{aligned}\text{Overall turnout rate} \;=\; &\text{Black turnout rate} \times \text{Black fraction of VAP} \\ &+ \text{Non-black turnout rate} \times \text{Non-black fraction of VAP}\end{aligned}$$

or in symbols

$$T_i = \beta_i^B X_i + \beta_i^N (1 - X_i) \tag{1}$$

We can use this equation to figure out the black turnout rate if we are first given the white turnout rate. In other words, we can use this equation to find the minimum and maximum mathematically allowed black turnout rate, simply by plugging in the maximum and minimum possible white turnout rate. Hence, if all 60,000 non-black adults had voted, it follows logically that only 20,000 blacks voted. On the other hand, it is logically possible that all 40,000 blacks voted, in which case exactly 40,000 non-blacks must have turned out.

To summarize, the logically possible range of black turnout rates lies between 50% and 100%, while the logically possible non-black turnout rates lies between $66\frac{2}{3}\%$ and 100%. Any values outside these bounds are mathematically impossible. Unfortunately, we are left with rather large ranges. The method of bounds by itself does not tell us any more precisely what the turnout actually was; only what it could be. Still, we have gathered some information that helps narrow the ecological inference problem.

A second technique dating from the 1950s is ecological regression, also known as Goodman's regression, after its discoverer (Goodman, 1953). Ecological regression assumes that the black turnout rate in each precinct is exactly identical, and then treats Equation 1 as a regression model. Unlike the method of bounds, ecological regression produces a single answer to ecological inference problems. Unfortunately, this answer is often wrong, and will generally be wrong if the black turnout rate varies across precincts, as it almost certainly does. Indeed, Goodman's regression often produces impossible results: it is not unusual to see ecological regression estimates claiming 130 percent of blacks voted Democratic, or that -30 percent of non-blacks voted Democratic. In general, Goodman estimates fail to respect or take advantage of the logically true information contained in the bounds.

Although at one time, Goodman's regression, or the even cruder proportional analysis used by the petitioner's experts, may have been used to "solve" ecological inference problems, the poor performance of early methods has long been recognized in scholarly work (see, e.g., Achen and Schively 1995), and for the most part scholars long avoided tackling problems of ecological inference because the inherent difficulties seemed insurmountable. Today, these methods are obsolete—to the extent they were ever even accepted—and are not generally accepted solutions to ecological inference problems in scholarly work.

4.3 Recent advances in ecological inference

Scientific progress has been made in recent years on the ecological inference problem. In particular, the new method of King (1997) set off an explosion of work on ecological inference, after decades of neglect. The central insight in this new work is that ecological inference is possible by combining the method of bounds and new statistical techniques (where the latter can be thought of as more sophisticated versions of ecological regression). These new methods work surprisingly well in solving some ecological inference problems in voting, but only when the problem can be broken down into precincts, some of which contain high concentrations of each "type" of voter. From these concentrations, the analyst can calculate tight deterministic bounds on the quantities of interest, and then conduct a statistical analysis within the limits given by these bounds. In this way, every modern method of ecological inference builds on the information in the bounds—and every method of ecological inference fails to produce any useful information when these bounds are universally uninformative.

One of the biggest contribution of King's approach to ecological inference is the recognition that in some cases ecological inference is practicable, while in other cases it clearly is not. King offers useful tests for whether ecological inference will yield meaningful inferences. These tests do not guarantee success, of course, but when they are negative, failure is almost certain.[5]

The strengths of King's approach, and the importance of exploiting the information in the bounds, has been recognized in many ways. On one hand, many studies using King's methods have been published in top journals (Burden and Kimball, 1998; Gay, 2001; Gimpel and Schuknecht, 2002; Liu, 2001; Lublin and Voss, 2002; Voss and Miller, 2001). On the other hand, King's method serves as the foundation for many new methods of ecological inference (see, for example, Wakefield, 2004; Rosen et al., 2001, and numerous contributions in King et al., 2004). Tellingly, all of these new papers and methods incorporate the deterministic bounds in the manner prescribed by King, and in this critical sense are all variations on his basic model. Finally, King's method has been used extensively in court cases where racial voting patterns were an issue. A federal judge held that King's method "is the best method currently available to measure racial bloc voting" (*Mallory, et al, v. State of Ohio, et al.*, USDC Southern District Court of Ohio, case no. C-2-95-381), a view upheld on appeal (*Mallory, et al, v. State of Ohio, et al.*, United States Court of Appeals, Sixth Circuit, No. 97-4425, 4/13/99). Subsequently, King's method has been used in numerous court cases, and has been accepted everywhere it has been

[5]Other scholars concur in this assessment. University of Wisconsin political scientist Charles Franklin told the New York Times

> It's far better to know that you don't know as much as you think you do. ... When the new method fails, it tends to produce evidence that shows you you're not doing a good job ... The opposite happens with the older method: it tends to tell you too small a margin of error when it's doing badly.

See Karen Freeman, "Statistician Builds What May Be a Better Data Mousetrap," *New York Times*, July 8, 1997, C8. On the importance of having tight bounds in ecological inference, see also Voss (2004) and Adolph and King (2003).

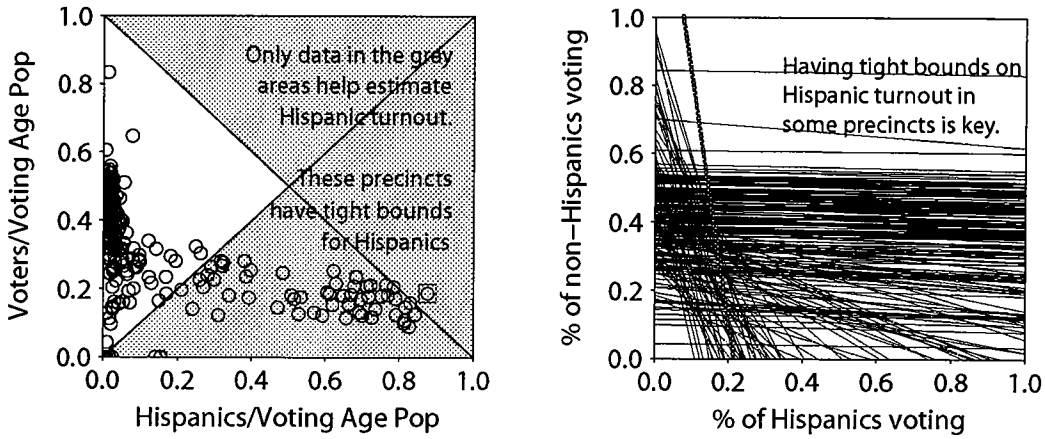


Figure 2: An example of a promising case for ecological inference: The 1990 Pennsylvania State Senate (2nd district) election. The key to successful ecological inference is having some informative precinct-level deterministic bounds on the quantity of interest. Here, the quantity of interest was the rate of turnout among Hispanics. The presence of some precincts with high concentrations of voting age Hispanics allows inferences to be made with modern techniques like that of King. An example of such a precinct is boxed in blue in the left plot. The precinct has bounds shown as a blue line in the right plot. Source: King (1997).

introduced.

Katz himself has used King's method as an expert witness in legislative redistricting lawsuits—for example, in Texas, *Perry v. Del Rio*, 66 S.W.3d. 239 (Tex. 2001). It is puzzling that he does not employ, or even cite, those methods in his report. Indeed, both examples of ecological inference Katz cites (Burden and Kimball, 1998; Gay, 2001) use King's method, and it is difficult to imagine these studies would have been published—much less published in the top journal in political science—had they used earlier techniques, such as Goodman's regression, or dubious methods, like "proportional analysis" or Katz's "principle of insufficient reason". Instead, ecological inference is now possible, and on the research agenda, because King's method provides either good estimates, or clear warnings that good estimates cannot be obtained from a particular dataset.

Although the details of King's method are quite technical, the general idea and crucial pre-estimation diagnostics can be described in plain English. Figure 4.3 illustrates an example of successful ecological inference using modern methods. The data come from the 1990 Pennsylvania 2nd Senate district general election, and the quantity we wish to infer is the proportion of voting age Hispanics turning out to vote. The two plots shown in Figure 2 are diagnostics introduced by King (1997) to aid researchers in deciding whether ecological inference is feasible for a given dataset.

The plot at left, known as a scattercross plot, graphs the turnout rate for all voters against the proportion of Hispanics in each precinct. The large X divides the plot into four triangles.

Precincts in the top, bottom, and right triangles all have informative bounds on the Hispanic turnout rate. Precincts in the left triangle have so few Hispanics that any turnout from 0 to 100% is possible. As a rule, ecological inference will be fruitless unless some of the precincts have informative bounds. In this case, such precincts exist; one example is boxed in blue.

The plot at right, known as a tomography plot, graphs a line for each precinct. This line contains every possible combination of Hispanic and non-Hispanic turnout rates allowed by the bounds (i.e., every combination of β's allowed by Equation 1). The bounds and these lines are deterministic in the sense that they rely on no statistical assumptions and are not the least bit uncertain (provided the underlying data are correct). It is logically required that the true turnout rate lie on this line. Steep lines pin down the Hispanic turnout rate in a precinct very precisely. The blue line, which corresponds to the boxed precinct in the left panel, is a precinct where there are enough Hispanics to determine the turnout rate from the bounds alone.

If we find some precincts with informative bounds, we can proceed to the next step, which is applying a statistical model to the distribution of tomography lines; this statistical model would produce estimates of turnout by ethnicity made robust by the information in the bounds. As a rough intuitive guide, note that the estimates of Hispanic and non-Hispanic turnout will tend to cluster near the intersection of the precinct-level tomography lines, though of course the estimates for each precinct will be restricted to lie on the portion of the precinct's tomography line nearest this intersection.

That is what an analyst would do in a good case, and the remarkable thing is we can identify probable good cases from known information (the bounds). On the other hand, if there were no informative bounds—i.e., no precincts outside the left triangle, or with steep tomography lines—then we would stop. In such cases, any effort at ecological inference, under any accepted method, is so dependent on unverifiable assumptions as to be practically useless.

4.4 The deterministic bounds in this case are very wide

Now, let's turn from the Pennsylvania example, where inference is possible, to the 2004 Washington election. We will use plots like those in Figure 4.3 as diagnostics to see whether ecological inference—of any sort—is possible in this case. The data I will consider are the combined invalid votes found by petitioners and respondents; however, the conclusions in this section do not materially change when these two datasets are considered separately.[6]

Recall that without some precincts with substantial invalid voters, there will be no precincts with narrow or informative bounds on invalid vote choice, and hence no possibility of successful inference of invalid vote choice from aggregate data. The scattercross plot for the invalid vote

[6]In his second supplemental report, Katz introduces a "calculated" residual vote total to account for write-ins, under- and over-votes, and spoiled ballots. I have attempted to reconstruct this quantity from data provided by the petitioners, but lack an exact description of the calculation used by the petitioners. Thus, the results given in this section and the next may be slightly different from those used by petitioners. However, I strongly doubt the calculation of residual votes will substantively affect any conclusions, as the findings here are unambiguous.

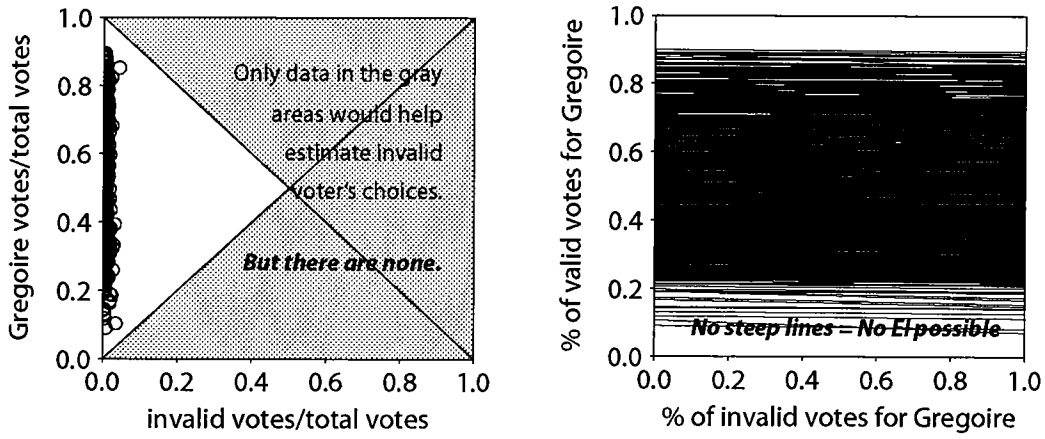


Figure 3: Standard diagnostics show the invalid ballots data are a hopeless case for any form of ecological inference. A key advance in recent work on the ecological inference problem is the ability to diagnose cases where inference is likely to fail. This is one of those cases.

data reveals that prospects for any kind of ecological inference are grim (Figure 3, left panel). It shows that none of the precincts will have informative bounds, a message that is hammered home by the tomography plot, which has no steep lines—meaning any proportions are possible in every any precinct (Figure 3, right panel).

Finally, in Table 3, I list the logically true bounds on the fraction of invalid ballots cast for Gregoire for each of the 1344 precincts which the petitioners or respondents identified as having at least one invalid ballot. *In each and every one of these 1344 precincts, the bounds on the vote choice of invalid voters are very wide. The bounds show it is possible that every invalid vote in every precinct was cast for Rossi. Alternatively, every vote may have been cast for Gregoire—or any combination in between, without contradicting any of the data.* Because there are too many precincts to list in this report, I list only the first and last five of these precincts (ordered by county and precinct number), as well as the two precincts most favorable for the petitioners: the precinct with the largest number of invalid ballots (Pierce County precinct 432, with a mere 8 invalid ballots out of 901 cast), and the precinct with the largest proportion of its ballots considered to be invalid (King County precinct 1595, with 4%). Even in these cases, the fraction of ballots that are invalid is far too small for us to infer anything about the behavior of the invalid voters from these aggregate data.

I have studied racial voting patterns in hundreds of elections using ecological inference techniques, and the statistical problem here is the same, as Katz notes on page 6 of his first report. A crucial step in every analysis I conduct is determining whether the bounds contain sufficient information for inference to proceed. Sometimes, I find that the bounds are simply uninformative, usually because no precincts contain a significant concentration of the group of interest. In such cases, I do not offer estimates, because the only results I could provide would be based on

County	Precinct	Number of invalid ballots, according to petitioners	Total votes cast, from manual recount	% of invalid votes cast for Gregoire Lower Bound	Upper Bound
Adams	212	1	128	0 %	100 %
Benton	1090	1	65	0	100
Benton	1400	2	77	0	100
Benton	1600	1	316	0	100
Benton	1645	1	151	0	100
⋮	⋮	⋮	⋮	⋮	⋮
King	1595	6	147	0	100
⋮	⋮	⋮	⋮	⋮	⋮
Pierce	432	8	901	0	100
⋮	⋮	⋮	⋮	⋮	⋮
Yakima	4602	2	287	0	100
Yakima	4606	1	505	0	100
Yakima	5003	1	720	0	100
Yakima	5016	2	276	0	100
Yakima	5020	1	311	0	100
Statewide bounds on the fraction of invalid ballots cast for Gregoire				**0 %**	**100 %**

Table 3: Wide deterministic bounds on the percent of invalid votes cast for Gregoire make this a hopeless case for any form of ecological inference. The above results are based on the combined data of petitioners and respondents. To save space, all but 12 of the 1344 precincts containing invalid votes (according to the petitioners) have been omitted from the table; the full table would run to 32 pages. However, I calculated the bounds for each of these 1344 precincts, and in every case, they range from 0 percent to 100 percent.

pure assumption, rather than the data.[7]

In my experience, I have seldom, if ever, seen a less promising case for ecological inference than the invalid voters of the 2004 Washington gubernatorial election. The precinct-level bounds contain no information whatsoever. If one were to fit a model to these data—whether that model were Goodman's regression, or "proportional analysis", or King's ecological inference method,

[7]To satisfy the reader's curiosity, I note that when I did press on, naively applying King's method to these unsuitable data, the method found that on average across the state, invalid voters voted no more Democratic than valid voters. In particular, it found that 49.9 percent of invalid voters voted Democratic, compared to 47.8 percent of valid voters, too small a margin to affect the election outcome. The standard errors for these quantities are less than 1 percent in each case. I emphasize, however, that these results are highly sensitive to model assumptions that cannot be verified, given the lack of informative bounds. The right thing to do in this case is to decline to attempt ecological inference.

or a variant of the same—the results one would find would highly sensitive to the assumptions made, and there would be no way to check if those assumptions had been validated by the data.

Rather than make untenable, untestable assumptions, I find that the best statement of the uncertainty in this case is given by the deterministic bounds. The only confidence interval around invalid voters' behavior that I am comfortable with states that it is possible that none of them voted for Gregoire, or that all of them voted for Gregoire, or any combination in between.

4.5 Gill ignores the ecological inference problem

To see how Anthony Gill's report addresses the problem of ecological inference, we must take a step back. As Gill lays out the problem, there are several things we would like to know to establish that the invalid votes changed the outcome of the election:

- First, we need to know how many net votes Gregoire can afford to lose before her margin is exhausted. This amount, obviously, is 129 votes.
- Second, we might like to know how many invalid votes in each precinct would be enough to overwhelm this margin, *given perfect knowledge of the rate at which invalid votes in each precinct were cast for Gregoire, for Rossi, and for other candidates.* Given such knowledge, the number of invalid votes needed to change the election (Gill's "tipping point") is also very easy to calculate.
- Third, and crucially, we need to know for whom the invalid votes were cast. While items one and two are extremely easy to calculate using basic arithmetic, this final piece of information is extraordinarily hard to estimate. Without this knowledge, items one and two will not avail us in determining the correct election outcome.

Gill devotes the bulk of his report to item two, an easy question. Gill thus focuses attention on a point not really in dispute. On the other hand, he devotes little space to the key issue of the case, which is estimating how invalid votes were actually cast. Gill simply assumes that these votes were cast in the same proportion as the other votes in the corresponding precinct. (Implicitly, Gill assumes that felons and law-abiding citizens in a single precinct are more alike with respect to voting than, say, law-abiding citizens in two different precincts.)

As we have seen, this assumption is a clear example of the ecological fallacy. Yet Gill makes no mention of the ecological inference problem, nor does he acknowledge the long history of scholarship on this difficult problem. Contrary to standard practice, he does not even offer an estimate of the uncertainty of his estimates on the vote choices of invalid voters. On face, Gill seems to suggest that the allocation of invalid votes to the Rossi or Gregoire column is a simple problem that can be solved with little difficulty or doubt.

Instead, there should be great doubt as to whether estimates from the "proportionality method" have any validity or use whatsoever in this case. Using his "proportionality method", not only would Gill conclude that Ichiro Suzuki had a mediocre year in 2004, he would also conclude, erroneously, that nationally, high-income voters vote more Democratic than low-income

voters. But let's give Gill's method a third chance, and ask whether it can correctly identify *known* characteristics of individual invalid voters from the aggregate characteristics of their precincts.

Suppose the statistical problem in this case was to infer, from aggregate data, the sex of the felons claimed by the petitioners to have voted illegally. Let us apply Gill's proportionality method to this problem, where we can verify its performance. We simply assume that the fraction of voting felons who are male matches the fraction of all voters in the precinct who are male. This is exactly how Gill models voting; we have merely applied the technique to gender, in a problem that is the same from a statistical point of view. Applying the proportionality method at the precinct level and summing across the state leads the analyst to conclude that felon voters must have been 46.9 percent male. Unfortunately, this inference is badly wrong. Based on individual level data, we know that the felons who voted were actually 75.3 percent male. The naive assumption of "proportionality" between aggregate and individual level data has led us far astray of the truth.

A basic test of whether a statistical method is valid is whether it produces accurate results under controlled conditions, where the true answer is known and can be compared to the answer given by the method. Judged on this venerable standard, "proportionality analysis" fails, and should be discarded. If the ecological inference problem were so easy to solve, social scientists and statisticians would not have spent half a century wrestling with it.

4.6 Katz fails to solve the ecological inference problem

Katz acknowledges that the problem at issue is one of ecological inference. Yet Katz uses the same method as Gill to allocate invalid votes—he assumes they were cast in the same proportions as the valid votes in the same precinct. Thus, like Gill, Katz's method strikes out swinging in the three examples discussed above.

Unlike Gill, Katz acknowledges the need to give measures of uncertainty for statistical estimates. I applied Katz's method of calculating confidence intervals for proportionality method estimates to the problem of estimating felon gender. Like Gill, Katz would predict that 46.9 percent of illegal felon voters are male. He would also offer a 95 percent confidence interval around this estimate ranging from 43.8 to 50.1 percent. Needless to say, the true proportion of males among the felon voters, 75.3 percent, lies far outside the 95 percent confidence interval.

Why has Katz's 95 percent confidence interval made us so over-confident in a flawed answer? The answer is simple: uncertainty about the *assumptions* of Katz's model is not included in Katz's confidence interval. Katz's confidence interval is valid only if his assumption that invalid voters are the same as valid voters in the same precinct is true. If this assumption is false, Katz's confidence intervals are incorrect, and could lead to dangerous overconfidence in his estimates.

Because the key uncertainty in this case is how invalid voters voted, it is disturbing that Katz and Gill can offer no evidence in favor of their proportionality assumption, nor any estimate of how certain this assumption is to be true. The uncertainty Katz does capture is trivial by comparison, and results from the fundamental variability of binomial (or multinomial) variates.

I discuss two further objections to Katz's methods in Section 6. Based on the examples and evidence above, however, it should be clear that his is a wholly inadequate approach to the problem.

4.7 This is a hopeless case

Based on modern, accepted statistical techniques, all we can be sure of is that we have no idea how the invalid voters cast their votes. There is no direct evidence on their behavior, and no existing method of ecological inference can extract information on these voters from the voting returns. Based on my experience in ecological inference, and the current standards in the field, I consider efforts to estimate from aggregate data the vote choices of invalid voters in this election to be a hopeless, quixotic endeavor.

5 Petitioner's own methods imply no change in election outcome

I do not find the data or methods presented by the petitioners' experts to meet the currently accepted standards in social science. Had this been an article under review for publication, as a referee, I would stop here and recommend rejection of the research. However, the court may be interested in whether the petitioner's results would hold up if we accept their methods, but apply them to the best currently available data. In this section, I do just that, with the caveat that I consider the methods used herein to be completely inadequate to the task.

I begin by replicating the results from Katz's first supplemental report. That is, I use Katz's own data and methods, and confirm that they lead to the results he presents. I was able to replicate Katz's results exactly using the same Normal approximation to the Binomial. I also reached nearly the same answers using a slightly more accurate simulation technique (the details of which can be found in the Appendix to this report). I present the slightly more accurate simulation results in Table 4, row 1. These show that under Katz's data and assumptions, the outcome of the election would almost certainly change.

Next, I correct two problems. The first is minor, and addressed by Katz in his second supplemental report. Prior to that report, Katz assumed that invalid voters had a choice between only two candidates, Gregoire and Rossi. In other words, while only 95.8 percent of all Washington voters cast ballots for Gregoire or Rossi—the remaining 4.2 percent chose Ruth Bennett, a write-in, or no one—Katz, in his first two reports, assumes that 100 percent of invalid voters choose Gregoire or Rossi. Under this assumption, the "proportionality method" slightly exaggerates the effect of invalid ballots on the election outcome.[8]

[8] In note 9 on page 7 of his initial report, Katz explains his choice to ignore third party votes:

> Through out the analysis I will be using two party share. This is essentially assuming none of the invalid ballots were cast for Ruth Bennett, the Libertarian candidate who garnered only 2.25% of the statewide vote. It is standard in most voting analysis to ignore minor candidates who win less than 5% of the vote. This assumption is not material to my

Correcting Katz's mistake is easy enough: rather than supposing each voter faced a binary choice, and hence assuming the total votes by candidate in a precinct are Binomially distributed, we recognize that each voter faces three choices (Gregoire, Rossi, and Other), and assume the total votes in each precinct are Multinomially distributed.[9] As with the Binomial, standard simulation techniques allow us to calculate confidence intervals around the estimates from this model (again, see the Appendix for details).

It turns out that the more appropriate multinomial distribution does not make too much difference in the results. However, this better assumption does shrink the Rossi margin slightly, as the second line in Table 4 shows. Indeed, this is the only criticism in this report which does not reverse the findings of the petitioners' experts, and it is the only criticism here addressed by Katz's supplemental reports.

A far bigger problem is the biased sample to which Katz applies his method. Applying almost any statistical method to a biased sample will yield biased results. In this case, the potential bias is enormous. Suppose that *instead* of the petitioner's seemingly cherry-picked data, we used only the data collected by the respondents. Applying Katz's method to these data yields the conclusion that Gregoire's margin would actually increase if the invalid votes were excluded. Moreover, using only the new data, the probability that the election outcome would change is close to zero.

Of course, using only the respondent's data is no more valid than using only the petitioner's data. Neither dataset is a census, random sample, or representative sample. The petitioners' data were collected primarily from precincts which voted overall for Gregoire; the respondents' efforts to supplement these data were collected primarily from precincts voting for Rossi (if only because those are the precincts the petitioners missed). Using either sample alone gives diametrically opposed results under the petitioners' experts' own methods. Neither set of results

analysis or findings.

Of course, what is "standard" in analysis of a typical election may not be sufficiently rigorous for a very close election. And in an election decided by 0.005 percent, which petitioners hope to overturn on the basis of a similarly tiny fraction of the ballots cast, 0.042 percent, we will need to be as careful and precise as possible. It is hard to justify treating 2.25 percent of the ballots as a negligible error in this context.

[9]The Multinomial distribution is a straightforward generalization of the Binomial distribution, appropriate when more than two different outcomes are possible from a given trial. For further details, see any basic text on probability; e.g., DeGroot (1989). Lumping together Bennett with write-ins, under- and over-votes, and spoiled ballots does not affect our conclusions, since in the end we want to isolate just Gregoire and Rossi's invalid votes. A more difficult issue is determining how many residual votes (write-ins, under- and over-votes, and spoiled ballots) were cast in each precinct. At this time, these data do not appear to be available, and Katz in his second supplemental report offers a "calculation" of them. Working from Katz's data, I have attempted to reproduce this calculation, but because he does not spell out his method, I am not certain I have done so in the same manner, nor am I yet convinced that his calculation is the right one. Naturally, any results in this section could be recalculated using Katz's method, once it becomes known, or any other method, and given the likely small number of residual votes, I would be surprised if any substantive conclusions were altered.

	Estimated Gregoire margin of victory	95% confidence interval		Probability excluding invalid votes would change election outcome
		lower	upper	
Petitioner's data, Katz 1st method	−133.3	−195.6	−71.0	0.99
Corrected for multinomial choice	−115.9	−177.0	−55.0	0.99
Respondent's data only	325.4	272.0	378.0	0.00
Combined data	66.5	−13.0	145.0	0.05

Table 4: Using the best available data, the Gill-Katz method finds no change in election outcome. The first line of this table shows the results from Katz second report, which I replicated exactly from Katz's data. The second line shows that the strength of these results weakens a bit when the correct statistical formulas are used. The third line shows that if one uses only the respondents' data, the effects found by Katz completely reverse. The fourth line shows that Katz's result largely disappears when more comprehensive data on invalid ballots are considered. The difference between Gregoire and Rossi's totals when the correct data are used is no longer significant by any commonly used statistical standard.

is statistically meaningful, or consistent with the generally accepted standards for quantitative work in the social sciences.

If we insist on using the (flawed) method of "proportional analysis", the most sensible thing would be to combine all the known invalid ballots into a single combined dataset. Unfortunately, we cannot be confident the combined data are representative or complete, but they are clearly less biased than either sides data taken separately. The fourth line of Table 4 shows that when the combined data are used, the proportionality method of Gill and Katz predicts that Gregoire would still win the election (by about 67 votes) if all invalid ballots were excluded. There is some chance Gregoire would still lose (the 95 percent confidence interval overlaps zero), but it appears that this chance is only about 5 percent.

Finally, recall the figure from Katz's initial report. This figure, reproduced in the left panel of my Figure 4, purports to find a correlation between the number of felons in a precinct and the overall share of the vote going to Gregoire. This exercise is a textbook example of the ecological fallacy: even if there is a correlation between these two aggregates, it says nothing about whether illegally voting felons actually voted Democratic. It is just like the classic example from Robinson (1950) cited in my note 3 above, and should set off alarm bells for any social scientist.

Still, it would be interesting to update this graphic, to include the invalid voters the petitioners' data leave out. I have done this in the right panel of Figure 4. Using a more complete dataset, we see that felons voted illegally in all kinds of precincts: Democratic and Republican-leaning. Of course, this still tells us nothing about how those felons voted.

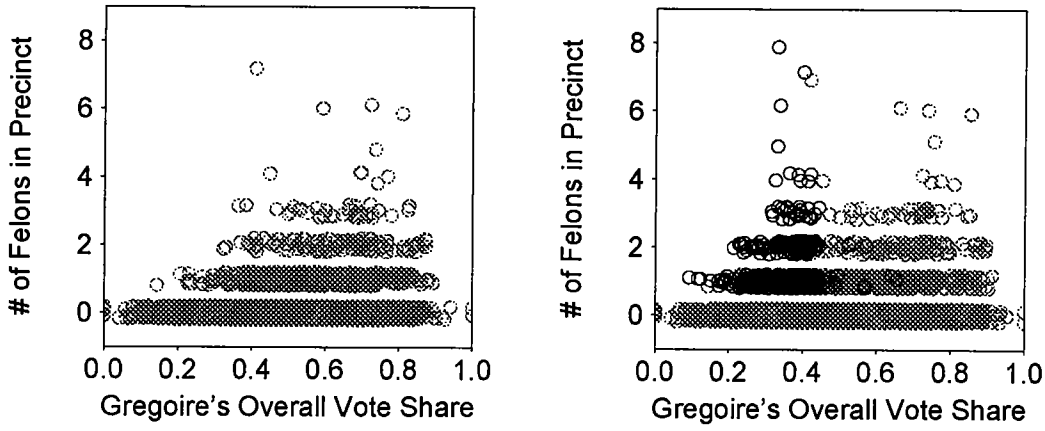


Figure 4: More complete data show no correlation between felons and Democratic vote share. The plot at left replicates Figure 1 in Katz's initial report, using only Katz's data, as updated in his supplements. The plot at right adds (in blue) the additional invalid votes discovered by respondents. The apparently correlation noticed by Katz turns out to result from biased data collection. Either way, making inferences from such plots is a textbook ecological fallacy, and should be avoided.

6 Other issues raised by petitioners' experts

The foregoing sections stand on their own to demonstrate the flaws in Gill and Katz's data, methods, and conclusions. This section turns to two side issues in the Gill and Katz reports.

6.1 The "principle of insufficient reason"

On page 7 of his initial report, Katz makes the following curious statement:

> Since we can not distinguish an invalid voter from valid voter from a valid geographic unit, then a randomly selected voter of either type *needs* to cast his or her vote for the Democratic candidate with an identical probability. Formally in statistics this is known as the principle of insufficient reason: when there is no way to distinguish two events or types, they need to have the same probability. [emphasis added]

From the tone and context, one would be forgiven for assuming that Katz means for the word "needs" to imply logical necessity: that in some mathematically required sense, the probability that a valid voter will choose a Democratic candidate is the same as the probability an invalid voter will vote Democratic. However, this is not logically required. Indeed, one could replace the word "needs" with "could", or "might", or even "might not" and obtain a perfectly valid statement. The only person who "needs" these probabilities to be the same is Katz, for he needs this assumption to draw the conclusion that Rossi would win the election had certain votes not been counted.

Katz tries to justify this assumption with an appeal to an old statistical argument known as the "principle of insufficient reason". His argument is brief—the quote above is all he has to say on the matter—and rather vague, and could be interpreted in several ways, none satisfactory. I will assume that Katz is treating the vote choice of the invalid voters as the random event to be modeled. Another way to read his argument is that we are somehow drawing invalid and valid votes from the universe of votes, and those draws are the random event. I consider the latter interpretation inconsistent with the facts of the case and the manner in which Katz's data were collected, and do not discuss it here. In fact, ambiguities like this one have historically made the principle of insufficient reason controversial, and it is seldom appealed to in applied social science research.

It appears, moreover, that Katz's application of the principle is incorrect. The principle of insufficient reason claims that when one does not have any information about the probability of the outcome of an experiment, one might as well suppose that all possible outcomes are equally likely. It is a recommendation for how to deal with situations of *complete* ignorance about probabilities of outcomes. It does not apply to situations where we have any information whatsoever about the probabilities of the events.

To see this, consider the following example. Suppose you are in an casino, and are challenged to a game by a stranger. The stranger proposes to roll a single, six sided die that he has brought with him. If he gets a 6, you will pay him \$5. If he gets rolls a 5 or less, he will pay you \$5 dollars. Should you take the bet?

To figure this out, you need to discover, or assume, the probability that the die will come up 6 when thrown. A balanced die will come up 6 once in six throws, on average. But the stranger may be a cheater, who has weighted his die to come up 6 more often. As he is a complete stranger, and he won't let you touch his die, you are remain ignorant of the probabilities of each outcome, and decide to apply the "principle of insufficient reason". That is, you *assume*

$$\text{P(Roll 1)} = \text{P(Roll 2)} = \text{P(roll 3)} = \ldots = \text{P(roll 6)} = \frac{1}{6},$$

where P(X) reads "the probability that event X will occur". Under this assumption, the bet sounds like a good deal.

Now, suppose as you are opening you mouth to take this stranger's bet, a man passes by and says, "Find another sucker?" This information changes everything. You now suspect that the die is an unfair one, and that P(roll 6) $> \frac{1}{6}$—though you still don't know the exact probability of a 6. You no longer can rely on the principle of insufficient reason to make your decision. In fact, it is irrelevant, because you are no longer completely ignorant of the probabilities of the outcomes of the experiment, and the principle of insufficient reason only applies to situations of complete ignorance.

If we applied the principle of insufficient reason to the probability that invalids voted for each candidate, we would assume

$$\text{P(Gregoire vote)} = \text{P(Rossi vote)} = \text{P(Bennett vote)} = \frac{1}{3}$$

Yet Katz does not make this assumption. Instead, he notes several pieces of information:

1. Far fewer people voted for Bennett than Rossi or Gregoire, so he assumes invalid votes for Bennett are negligible.

2. At the precinct level, there is an accounting identity which must hold.

Neither piece of information helps us pin down the probabilities with any precision, but it is no longer the case that we are completely ignorant of the probabilities. Thus Katz does not use the principle of insufficient reason (which would lead to the 1/3 solution). Instead, he *assumes* that any subset of the population behaved in the same manner as the average person in the precinct. This is not a result of the principle of insufficient reason, but instead is its own assumption. Rather than the principle of insufficient reason, I would call it the "principle of homogeneity".

Homogeneity is *not* derived from the principle of insufficient reason. It is instead made purely for convenience. Katz has two unknown quantities, and no satisfactory way to estimate them. It would be convenient for Katz to assume they exactly equal two other known probabilities. But there is no evidence supporting this assumption, no statistical principle he can point to in support of this assumption, and indeed, a long literature warning against just this sort of assumption as leading directly to the ecological fallacy.

6.2 Do we have outside evidence on felon voting behavior?

Both Gill and Katz cite a paper, Uggen and Manza (2002), which they claim gives evidence that felons tend to vote Democratic. It is important to note that neither Gill nor Katz incorporate Uggen and Manza's work into their own analyses. Instead, both Gill and Katz assume that felons are exactly as likely to vote Democratic as other voters in their precinct. It is also important to note that Uggen and Manza do not provide any evidence on the voting behavior of actual felons, nor do they solve the ecological inference problem which bedevils Gill and Katz. Simply citing Uggen and Manza does nothing to bolster Gill and Katz analysis.

Uggen and Manza *assume* that felons with given demographic characteristics *would* vote Democratic at the same rate as non-felons with those demographic characteristics. Then, based on the fact that a disproportionate fraction of felons *nationwide* are black, and the fact that black non-felons tend to vote Democratic, they hypothesize that if felons were allowed to vote, they would tend to vote Democratic.

Note the words "assume", "would", and "nationwide" in the above summary. Uggen and Manza have not set out to demonstrate how felons vote, but instead want to speculate on the electoral consequences of a particular hypothesis about how they might vote. They do not study

the voting behavior of actual disenfranchised felons, precisely because felons are disenfranchised. Moreover, their data are drawn from, and their conclusions are aimed at, average nationwide behavior in historical elections, not future elections in particular states.

To put it all together, Uggen and Manza conducted a study of the voting behavior of non-felons across America in federal elections prior to 2002. Gill and Katz suggest this tells us something about the behavior of felons in Washington in the 2004 gubernatorial election.

I have already noted the danger in assuming felons are identical, even on average, to the rest of their precincts. Assuming that felons are identical to non-felons across the nation is just as dangerous. The potential ecological fallacies are multiplied by the many assumptions needed to make this comparison work:

1. Neither Gill nor Katz make any effort to control for the potentially different demographics of Washington felons. The key variable driving Uggen and Manza's analysis (indeed, the only variable which is statistically significant in all their logistic regressions) is voter race. If Washington state felons—more specifically, Washington state felons who voted illegally—include fewer blacks than the national population of felons, then Uggen and Manza's estimates would need to be adjusted for this fact. From Uggen and Manza's Appendix Tables A and B, we see that in 2000, 11.6 percent of Washington ex-felons were black, compared to 33.3 percent nationwide. This is a huge difference, and would likely have a large impact on any predictions of party preference under the Uggen and Manza model.

2. Even when matched on race and any other available demographic variables, felons and non-felons still differ in at least one crucial respect: felony status. We simply have no evidence on whether, say, white female felons and white female non-felons vote the same. To assume so is to commit the ecological fallacy once again.

3. Moreover, Uggen and Manza did not study state elections, and caution that

   > the biggest drawback of the NES series [used to estimate party preferences] is that while it asks respondents how they voted in presidential and congressional elections, there are too few respondents ($N < 2500$) to permit meaningful state-level analyses. (Uggen and Manza, 2002, 784)

   In other words, we have no evidence that the opinions of Uggen and Manza's non-Washington non-felons on the national Democratic party are the same as the opinions of Washington felons on the Washington State Democratic party. Clearly, there are many assumptions made in such a comparison, and no evidence that these assumptions are correct.

4. Indeed, there is evidence that Washington voters, at least in the 2004 gubernatorial election, bucked the national trends. Uggen and Manza rely on the fact that non-white voters ordinarily prefer Democratic candidates. Yet exit polls in this specific election showed that

male non-white voters strongly preferred the Republican Dino Rossi over Gregoire, by a 54 percent to 42 percent margin.[10] Each election is different, and making inferences from the average election to a specific election is an ecological fallacy.

5. Candidate characteristics are often a powerful factor explaining vote choice. There is a reasonable argument that felons would be systematically opposed to Gregoire, who is a former Attorney General. To dismiss this possibility, we would need evidence on the voting preferences of felons in elections involving a former Attorney General.

6. Another potentially relevant characteristic is candidate gender. A reasonable hypothesis, supported by evidence from other states (Sanbonmatsu, 2002), is that male voters tend to prefer male candidates, while female voters tend to prefer female candidates. Thus, the application of Uggen and Manza's conclusions to the Washington gubernatorial race may be doubly difficult: even if we decided felons prefer the average Democratic candidate, we cannot be sure they prefer female Democratic candidates, like Christine Gregoire.

7. Even if we had a study of the political preferences of all felons in Washington, we would still face an ecological inference problem in distinguishing the preferences of felons who tried to vote from those who chose not to vote. The invalid voters in the 2004 Washington gubernatorial are unique in many ways, and only individual level data on these particular voters can tell us how they voted. Otherwise, we are still guessing how voters behaved based on unverifiable and potentially faulty comparisons to other people.

8. Finally, it is worth noting that Uggen and Manza's research is based on all felons, not just disenfranchised ex-felons. The authors suggest that their results might change if the question were ex-felon behavior (Uggen and Manza, 2002, 790).

There are many pitfalls in applying Uggen and Manza's research on a nationwide sample of non-felons to a particular group of Washington felons. This is why we actually hold elections, rather than rely on opinion polls to decide who will be elected. To extrapolate from Uggen and Manza's study to the outcome of the 2004 Washington gubernatorial election would be to decide an election using opinion polls about different elections conducted in other states, years before Dino Rossi and Christine Gregoire became candidates in Washington. It is difficult to see the use of such an exercise, or quantify the uncertainty of any conclusions we would draw from it.

Yet both Gill and Katz claim that the existence of Uggen and Manza's paper somehow makes their own work more "conservative". It is not clear what this would mean. If Gill and Katz believe Uggen and Manza's approach is better than their own, they should have based their analysis on Uggen and Manza. Instead, they chose to assume not that felons vote like demographically similar people across the nation, but instead, that they vote exactly like

[10]Data from an exit poll of 2,152 Washington voters, available from CNN at `www.cnn.com/ELECTION/2004/pages/results/states/WA/G/00/epolls.0.html`.

everyone else in the neighborhood.[11] The only estimates Gill and Katz provide rely on the latter assumption. Based on Gill and Katz's reports, we simply don't know how to turn Uggen and Manza's paper into a statement of who won the Washington gubernatorial election. I do not think we should try.

7 Conclusions

A central concern of statistics is the quantification of uncertainty. Statisticians are not in the business of making wild guesses. Instead, statisticians and political methodologists present estimates of the uncertainty of their findings. Accordingly, in the case of the vote choice of invalid voters in the 2004 Washington gubernatorial election, I cannot offer an estimate of how those invalid voters voted, precisely because currently accepted methods show that any such estimate would be so uncertain as to be useless.

The only confidence intervals that do not rest on strong, dubious, or unwarranted assumptions are the deterministic bounds I provide in Section 4.4. These confidence intervals tell us what is intuitively obvious to most observers: we do not know how invalid voters voted in the 2004 gubernatorial election, we cannot conclude with any confidence that their exclusion would change the outcome of the election, and we will not be able to resolve this question using data and statistical methods currently available.

In contrast, the petitioners' experts would have us believe that we can estimate the behavior of invalid voters, and do so precisely enough to make confident statements about how their exclusion would change the outcome of the election. Yet the data they rely on is unrepresentative and their methods are inadequate. At the heart of the Gill and Katz reports is the assumption that invalid voters are identical to other voters in the same precinct. Neither expert offers evidence for this assumption, and it is this assumption alone that drives their results. Accordingly, using the petitioners data, it would be a leap of faith to assert the election outcome would change with a 95 percent probability, or in fact with any probability at all. Moreover, simply using more complete data, along with Gill and Katz's flawed assumptions, would lead to the conclusion that excluding invalid votes would not change the election outcome at all.

[11] Uggen and Manza (2002, 783) clearly do not support the assumptions of Gill and Katz. Instead, they emphatically reject the homogeneity assumption behind "proportional analysis":

> If we could assume unit homogeneity—that felons would have voted in the same numbers and with the same preferences as non-felons—we could simply count the disenfranchised felons and apply national turnout and party-preferences averages. But because felons differ from non-felons in ways that are likely to affect political behavior, this sort of blanket assumption is likely untenable.

Appendix: Details of Multinomial simulation

In this appendix, I explain the technical details of the calculating the quantities in Table 4.[12]

First, define the number of invalid voters in precinct i as I_i, define the fraction of all voters choosing Gregoire in precinct i as V_i^G, define the fraction of all voters choosing Rossi in precinct i as V_i^R, and define the fraction of all voters making another choice in precinct i as $1 - V_i^G - V_i^R$. Now we can calculate any desired quantity from the Multinomial distribution using the following steps:

Step 1: Following Katz's model, for each precinct i, draw the vote choices of invalid voters from a multinomial distribution with I_i trials, P(Gregoire vote) $= V_i^G$, P(Rossi vote) $= V_i^R$, and P(Other vote) $= 1 - V_i^G - V_i^R$.

Step 2: Sum over all precincts in the state to obtain a set of three simulated invalid vote totals, one for Gregoire, one for Rossi, and one for Other. Call these totals $\tilde{S}^G$, $\tilde{S}^R$, and $\tilde{S}^O$, respectively.

Step 3: Calculate the difference between the simulated Rossi total and the simulated Gregoire total plus 129 (the margin in the manual recount), $\tilde{S}^R - \tilde{S}^G + 129$. This is the simulated Gregoire margin of victory if invalid votes are excluded, $\widetilde{MV}^G$.

Step 4: Repeat Steps 1 to 3 many (e.g., 10,000) times, each time recording $\widetilde{MV}^G$, to obtain a list of 10,000 simulated margins of victory.

Step 5: Calculate summaries from the list of $\widetilde{MV}^G$'s. For example, the average $\widetilde{MV}^G$ is the expected Gregoire margin under Katz's model. The 95 percent confidence interval can be found by taking the 250th largest and 9750th largest $\widetilde{MV}^G$'s. And the probability of a change in the outcome of the election from excluding invalid voters is simply the fraction of $\widetilde{MV}^G$'s which are less than 0.

This simulation routine produces the same quantities as Katz's own methods, without resorting to Normal approximations of the Binomial or Multinomial distributions.

[12]This algorithm works whether one is simulating Binomial or Multinomial data, as the Binomial distribution is a special case of the Multinomial.

References

Achen, Christopher H. and W. Phillips Shively. 1995. *Cross-level Inference.* Chicago: University of Chicago Press.

Adolph, Christopher and Gary King. 2003. "Analyzing second stage ecological regressions." *Political Analysis* 11(65–76).

Burden, Barry and David Kimball. 1998. "A new approach to the study of ticket splitting." *American Political Science Review* 92(3):533–544.

DeGroot, Morris. 1989. *Probability and Statistics.* 2 ed. Reading, MA: Addison Wesley.

Duncan, Otis Dudley and Beverly Davis. 1953. ""An alternative to ecological correlation." *American Sociological Review* 18:663–666.

Gay, Claudine. 2001. "The effect of black Congressional representation on political participation." *American Political Science Review* 95(3):589–602.

Gelman, Andrew, Boris Shor, Joseph Bafumi and David Park. 2005. "Rich state, poor state, red state, blue state: whos voting for whom in Presidential elections?" Presented at the Annual Meetings of the Midwest Political Science Association, Chicgo, IL, April 8, 2005.

Gimpel, James G. and Jason E. Schuknecht. 2002. "Political and demographic foundations for sectionalism in state politics: The Connecticut case." *American Politics Research* 30:193–213.

Goodman, Leo. 1953. "Ecological regressions and the behavior of individuals." *American Sociological Review* pp. 663–664.

Jacobs, Jerry A. 2005. "ASR's Greatest Hits." *American Sociological Review* 70(1):1–3.

King, Gary. 1997. *A Solution to the Ecological Inference Problem.* Princeton: Princeton University Press.

King, Gary, Ori Rosen and Martin A. Tanner, eds. 2004. *Ecological Inference: New Methodological Strategies.* Cambridge: Cambridge University Press.

Liu, Baodong. 2001. "The positive effect of black density on white crossover voting: Reconsidering social interaction theory." *Social Science Quarterly* 82:600–613.

Lublin, David and D. Stephen Voss. 2002. "Context and Francophone support for sovereignty: An ecological analysis." *Canadian Journal of Political Science* 35:75–101.

Robinson, William S. 1950. "Ecological correlations and the behavior of individuals." *American Sociological Review* 15:351–357.

Rosen, Ori, Wenxin Jiang, Gary King and Martin A. Tanner. 2001. "Bayesian and Frequentist inference for ecological inference: The $R \times C$ case." *Statistica Neerlandica* 55(2):134–156.

Sanbonmatsu, Kira. 2002. "Gender stereotypes and vote choice." *American Journal of Political Science* 46(1):20–34.

Uggen, Christopher and Jeff Manza. 2002. "Democratic contraction? Political consequences of felon disenfranchisement in the United States." *American Sociological Review* 67:777–803.

Voss, D. Stephen. 2004. Using ecological inference for contextual research. In *Ecological Inference: New Methodological Strategies*, ed. Gary King, Ori Rosen and Martin A. Tanner. Cambridge: Cambridge University Press pp. 69–96.

Voss, D. Stephen and Penny Miller. 2001. "Following a false trail: The hunt for white backlash in Kentucky's 1996 desegregation vote." *State Politics and Policy Quarterly* 1:63–82.

Wakefield, Jonathan C. 2004. "Ecological inference for 2×2 tables (with discussion)." *Journal of the Royal Statistical Society, Series A* 167:385–445.

Curriculum Vitae

CHRISTOPHER ADOLPH

Center for Statistics and the Social Sciences
C-14-E Padelford Hall / Box 354320
University of Washington
Seattle, WA 98195-4320

Education

HARVARD UNIVERSITY
Ph.D., Political Science, March 2005
A.M., Political Science, June 2000

Dissertation: 'The Dilemma of Discretion: Career Ambitions and the Politics of Central Banking.' Advisors: Torben Iversen (chair), James Alt, and Michael Hiscox.

RICE UNIVERSITY
B.A., *summa cum laude*, Phi Beta Kappa, Political Science, Economics, and History, May 1998

Employment

UNIVERSITY OF WASHINGTON September 2004–
Assistant Professor of Political Science
Adjunct Assistant Professor of Statistics
Core Member of the Center for Statistics and the Social Sciences

Fellowships

Center for Basic Research in the Social Sciences Dissertation Completion Fellowship, 2002-2003
Harvard Graduate Society Merit Fellowship (declined)
Doctoral Fellow of the Multidisciplinary Program in Inequality and Social Policy;
Recipient of National Science Foundation Training Grant, 2001-2002
Harvard Summer Language Fellowship, 1999
National Science Foundation Graduate Fellowship, 1998-2001
Jacob Javits Fellowship (declined)
Faculty-Undergraduate Research Grant, Rice University, 1998

Publications

'Visual Interpretation and Presentation of Monte Carlo Results.' *The Political Methodologist*, 2003, 11.2: 31–35.

'A Consensus on Second Stage Analyses of Ecological Inference Models.' (Christopher Adolph and Gary King, with Michael Herron and Kenneth Shotts). *Political Analysis*, 2003, 11: 86-94.

'Analyzing Second Stage Ecological Regressions.' (Christopher Adolph and Gary King). *Political Analysis*, 2003, 11: 65-76.

'Party Schools: How the Republicans Skew State Aid to Education.' *The American Prospect.* December 18, 2000. 16–17.

Working Papers

'What Time Series Cross-Section Estimator Should I Use *Now*? Guidance from Monte Carlo Experiments.' (joint work with Daniel Butler and Sven Wilson)

'Three simple tests of career influences on monetary policy'

'Ambition is universal: Career effects on inflation in developing countries'

'Succession in the Temple: Central Banker Careers and the Politics of Appointment'

'The Uses of Autonomy: Central Bankers' Careers, Institutional Context, and Economic Performance.'

'Paper Autonomy, Private Ambition: Theory and Evidence Linking Central Bankers' Careers and Economic Performance.'

'Parties, Unions, and Central Banks: An interactive model of unemployment in OECD countries.'

'Playing Favorites: How Parties Distribute School Finance by Income and Race.'

'Homeowner Association Foreclosures and Property Values in Harris County, 1985-2001'

Teaching

University of Washington: I have taught graduate level quantitative methods courses listed in four departments (Political Science, Statistics, Sociology, and the Center for Statistics and the Social Sciences), covering maximum likelihood estimation, categorical data analysis, models of counts, contingency tables, and simulation. I have also offered an undergraduate seminar on comparative political economy. Extensive teaching materials are available at my web site.

Harvard University: Teaching fellow for courses in advanced quantitative political methodology (Gary King) and the visual display of quantitative information (Alyssa Goodman). Principal advisor to four thesis writers, including a Hoopes Prize winner and three recipients of highest departmental honors (probability of random occurrence $\approx .002$). Teaching assistant for Empirical Implications of Theoretical Models workshop, 20002, Harvard University (assisting Robert J. Franzese).

Presentations

'What Time Series Cross-Section Estimator Should I Use *Now*? Guidance from Monte Carlo Experiments.' (Joint with Daniel Butler and Sven Wilson), Paper presented at the Annual Meetings of the Midwest Political Science Association, Chicago, April 10, 2005.

'Succession in the Temple: Central Banker Careers and the Politics of Appointment' Paper presented at the Annual Meetings of the Midwest Political Science Association, Chicago, April 10, 2005; at the Center for Statistics and the Social Sciences Seminar, March 9, 2005; poster presented at Summer Meetings of the Society for Political Methodology, Stanford University, July, 2004.

'The Dilemma of Discretion: Career Ambitions, Policy Preferences, and the Politics of Money' Poster presented at Multidisciplinary Program on Inequality and Social Policy Summer Institute, Cambridge, MA, June 19, 2004.

'The Uses of Autonomy: Central Bankers' Careers, Institutional Context, and Economic Performance.' Presented at the Annual Meeting of the Midwest Political Science Association, Chicago IL, April 16, 2004.

'Paper Autonomy, Private Ambition: Central Bankers' Careers and the Economy.' Presented at Yale University (November 2003), Rice University (November 2003), Penn State University (December 2003), the University of Washington (December 2003), New York University (February 2004), the Annual Meeting of the American Political Science Association, Philadelphia, PA, August 2003, and at the Comparative Political Economy Workshop, Cornell University, October 25-26, 2002.

'Central Bankers' Careers and Inflation: Results from 14 Countries.' Presented at the Annual Meeting of the American Political Science Association, Boston MA, August 2002.

'Parties, Unions, and Central Banks: An interactive model of unemployment in OECD countries.' Presented at 'The Comparative Political Economy of Developed and Less Developed Countries,' Yale University, May 4-5, 2001.

'Economic and Racial Bias in Parties' Allocation of State Educational Aid.' Presented at the Inequality and Social Policy Workshop, September 25, 2000.

'Parties, Institutions, and Unemployment.' Presented at the Rational Choice/Political Economy Lunch, November 12, 1999.

'Determining Legislative Outcomes: The Effects of Committee Specialization on Legislative Committee Performance.' With Keith Hamm, Ronald Hedlund, and Nancy Martorano. Presented at the Annual Meeting of the American Political Science Association, Boston, MA, September, 1998.

Professional Activities

Member of American Political Science Association, Midwest Political Science Association, American Statistical Association, and American Economic Association.

Referee for *Party Politics* and *Political Analysis.*

Discussant for Comparative Political Economy Poster Session, Annual Meetings of the Midwest Political Science Association, Chicago, April 10, 2005; Chaired 'New Approaches to the Welfare State' Panel at the European Political Economy Workshop, Center for European Studies, Harvard University, November 5-6, 1999.

Foreign Languages: Spanish (advanced reading), French (intermediate reading)
Computer Languages: Gauss, R/Splus, Perl