E C O N O M I C S   B U L L E T I N

Electoral behavior of US counties: a panel data approach

Stanislav Anatolyev
New Economic School

Abstract

This note proposes an econometric framework for studying electoral returns using aggregate voting and socioeconomic panel data. Along with usual covariates, the model includes electoral unit effects, electoral subunit effects and time effects, and features nested groupings and heteroskedasticity. We apply the framework to model the electoral behavior of US counties in congressional elections.

My thanks go to Mahmoud El−Gamal, Grigory Kosenok, former peer students at the University of Wisconsin−Madison and participants of the Panel Data Methodological Seminar in Moscow in Fall 2001 organized by the Economics Education and Research Consortium. The usual disclaimer applies.
Citation: Anatolyev, Stanislav, (2002) "Electoral behavior of US counties: a panel data approach." *Economics Bulletin,* Vol. 3, No. 9 pp. 1−10
Submitted: April 3, 2002. **Accepted:** June 28, 2002.
URL: http://www.economicsbulletin.com/2002/volume3/EB−02C50004A.pdf

1. Introduction

The US electoral data provide rich opportunities for studying the individual as well as aggregate behavior. The data usually span quite long periods of time; legislative branch elections occur bi-yearly and presidential elections – every four years. This means that the time dimension of electoral panels may be quite large. At the same time the cross-sectional dimension may be huge if one wishes to consider the behavior of very small electoral units like wards, townships and precincts. The negative fact is that for such small units there exists no statistics on socioeconomic indicators that are likely to appear as regressors in a reasonable econometric model. More precisely, "many of the sources of data for subcounty units are little known outside their local areas... The costs of collecting data, as a consequence, rise dramatically when research is focused upon such subcounty units and when reliance must be placed upon scattered and idiosyncratic data sources" (Austin, Clubb and Traugott 1981). But studying the behavior of larger electoral units, counties say, relaxes these problems since there are many sources of socioeconomic data for them. These data are though often incomplete and contaminated.

Another dimension of analysis is possible due to the fact that elections into different offices usually occur simultaneously. This means that returns of presidential, federal and state legislative elections may be studied together. Also, usually there are more that two parties standing for election, which permits broader analysis than that of returns for one of the parties. Finally, one may take into account a voter turnout (which seems to be the primary concern in political studies) having in mind the selection problem – difference in willingness to vote among, say, democrat and republican sympathizers. It is worthwhile to note that one may take quite a different approach to the issue of electoral behavior – analyzing individual behavior rather than that of communities of voters. Difficulties arise due to the balloting system where one cannot trace characteristics of the person responsible for a particular vote, so one has to heavily rely on the assumption of random selection. There exist such data collected by the National Election Study (NES), but nothing can guarantee random sampling as well as the absence of misreporting and non-responses. Besides, these data are pseudo-panels rather than true panels, since, as a rule, different people are inquired in different years.

In spite of availability of data and tools, the quantitative electoral literature mostly attempts to draw conclusions from the observed partial correlations. One of the most serious treatments is Clubb, Flanigan and Zingale (1981), which though does not go into the quantitative analysis beyond studying connections between the individual behavior and group data (Iversen 1981). The present note tries to go farther in the quantitative analysis of electoral panels. It follows the aggregate approach, and takes into account not only time and cross-sectional dimensions of the panels, but also the nesting effects caused by division of states into counties. The unit of analysis is a county, but counties are grouped in a natural way into larger divisions – states, and the behavior of two intra-state counties have more in common than that of two interstate ones. Although this does cause certain technical complications, the idea seems reasonable.

2. The model

2.1. The dependent variable and covariates

As mentioned in the introduction, the units of analysis are counties, which are grouped into states. As the dependent variable I take the percentage of votes of participated eligible electorate for the Democratic Party representatives in congressional elections (held once every two years), which is conventional in the literature. I account for division of the rest of votes among other parties assuming that, in aggregate, independence of irrelevant alternatives holds. Towards this end, I assume that only Democrats and Republicans are the "relevant" alternatives, and so change the original return to the proportion of democratic votes in the "relevant" ones:

$$\text{Return (\%)} = \frac{\text{\% for D}}{\text{\% for D + \% for R}}.$$

For simplicity, I ignore attrition, which, even if exists, must be negligible. The two parties do not distinguish themselves to a such extent that people behave strategically. This would have to be taken into account in analyzing electoral behavior in such country as Russia where the presence of the attrition bias is evident.

Not all county covariates matter for determination of election returns. While the race composition or average educational level of a county do matter, some socioeconomic indicators matter only on the state level. For example, one may argue that the distribution of income across counties is quite even for a given state, while the economic performance of the whole state affects the electoral preferences, the per capita income of states indeed exhibiting strong variation. Thus I divide the regressors into two parts: the first, county-level covariates, and the second, state-level covariates. The composition of both parts will be discussed in section 4.

2.2. State, county and time effects

The panel of data includes election return y_{ct} for county *c* in year *t*, vector of county-specific covariates x_{ct} for county *c* in year *t*, and vector of state-specific covariates z_{st} for state *s* in year *t*. The model, which is assumed to be linear, is written as

$$y_{ct} = a + x'_{ct}b + z'_{st}g + w_{sct},$$

where w_{sct} is unobservable disturbance term, which is comprised of several components:

$$w_{sct} = m_s + l_t + u_{sct},$$

where m_s account for fixed state effects, l_t denote fixed time effect, and u_{sct} are stochastic disturbances composed of the two random components: $u_{sct} = u_{st} + e_{ct}$, where u_{st} are state-level stochastic disturbances, and e_{ct} are remainder county-level stochastic disturbances. Both u_{st} and e_{ct} are assumed to be independent of each other and of vectors of covariates x_{ct} and z_{st} for all *t*, *s* and *c*. In the panel-data language, the error has four components and nested groupings. Two of the unobservable effects are fixed and two others are random. The two fixed are state effects m_s and time effects l_t.

A state effect may be thought of as overall, net of accounting for *z*'s, conservatism or progressism of the state where the county belongs, which obviously affects election results. The rough analogy would be dividing the country into two large groups – North and South, which is often made in the growth literature. I allow subtler division of states permitting each state to have its own "index" of political preference, and expect the m_s to exhibit clustering. Note that because the intercept term is already included, a constraint like $\sum_{s=1}^{S} m_s = 0$ or $m_S = 0$ must be imposed. The time effects are quite comprehensible phenomena in electoral time-series data. They account for nationwide shifts in political preferences due to global events like outstanding performance of a president from a certain party or the end of the Cold War. Again, a constraint like $\sum_{t=1}^{T} l_t = 0$ or $l_S = 0$ should be imposed.

The two random effects are state-time and county-time disturbances. The former represents all deviations of a state from nation-average behavior caused by, say, the behavior of state authorities. The latter is a remainder disturbance for the smallest unit of analysis. In an alternative specification the state-time effects u_{st} might be assumed fixed as well, but this would bring a loss of many degrees of freedom without clear advantage.

It is also worth noting that the above specification is not the most general as far as effects are concerned. Pure county effects free of time index are not present in the model. That is, I assume that a county is just a subdivision of a larger unit – state, and does not have political preference bias not connected with the county's covariates. All deviations are modeled by e_{ct}, and are assumed to be uncorrelated across time and counties.

2.3. Nested effects and heteroskedasticity

The division of the units of analysis into larger groups in panel data has not been extensively discussed in the literature, which usually assumes equal "cell" numbers (Ghosh 1976, Baltagi 1993). But in the problem under consideration this is certainly not the case: the number of counties within a state varies from one state to another. Although estimation methods become more complicated from technical perspective, they may nevertheless be generalized in a straightforward way.

Nothing yet has been said about the marginal distribution of u_{st}- and e_{ct}-disturbances. Since u_{st} result from actions of state authorities, there is no reason to assume dependence of their marginal distribution on the size of a state. Where size does matter, it is in the e-components that represent deviations from the average behavior of the county's electorate. Therefore it is reasonable to link the variance of e_{ct} to the county population, or, which is more appropriate, the number of participated voters, i.e. eligible electorate multiplied by the turnout rate. Since the dependent variable is measured in percentages of votes, one may appeal to the Binomial distribution. Let for one person the probability of voting for Democrats be *p*, then for *n* voters the variance of the election return is *p(1-p)/n*. This is in no way a precise argument (after all, *p* varies across voters who do not behave independently, etc.) but rather a justification of the following form of heteroskedasticity: $Var(e_{ct}) = s_e^2 / n_{ct}$.

2.4. The model in full

To recapitulate, the model is

$$y_{ct} = a + x'_{ct} b + z'_{st} g + m_s + l_t + u_{st} + e_{ct}, \; t = 1,\ldots,T, \; s = 1,\ldots,S, \; c = 1,\ldots,C,$$

where $u_{st} \sim iid\left(0, s_u^2\right)$, $e_{ct} \sim inid\left(0, s_e^2/n_{ct}\right)$, and u_{st} and e_{ct} are independent of each other and of x_{ct} and z_{st}; m_s are *S* fixed state effects, $m_S = 0$; l_t are T fixed time effects, $l_S = 0$; x_{ct} is a $k_1\times1$-vector of county-specific characteristics, z_{st} is a $k_2\times1$-vector of state-specific characteristics, b and g are $k_1\times1$- and $k_2\times1$-vectors of parameters, a is a common intercept, n_{ct} is a participation level in congressional elections in county *c* in year *t*, and y_{ct} is the election return. In total there are $k_1 + k_2 + S + T - 1$ regression parameters and 2 variance parameters. In the matrix form the model looks as follows:

$$\mathbf{Y} = \mathbf{R}d + \mathbf{U} \text{ with } \mathrm{E}\,\mathbf{U} = \mathbf{0},\ \mathrm{E}\,\mathbf{UU}' = \Omega,\ \mathrm{E}\,\mathbf{UR} = \mathbf{0},$$

where $\mathbf{R} = \left[i_{CT} \quad \mathbf{X} \quad \mathbf{Z} \quad \mathbf{E}_m \quad \mathbf{E}_l\right]$, $d = \left[a \ b' \ g' \ m_1 \cdots m_{S-1} \ l_1 \cdots l_{T-1}\right]$, $\mathbf{Y} = \left[\left[y_{ct}\right]_{c=1}^{C}\right]_{t=1}^{T}$,

$\mathbf{X} = \left[\left[\mathbf{x}'_{ct}\right]_{c=1}^{C}\right]_{t=1}^{T}$, $\mathbf{Z} = \left[\left[i_{C_s} \otimes \mathbf{z}'_{st}\right]_{s=1}^{S}\right]_{t=1}^{T}$, $\mathbf{E}_m = i_T \otimes \begin{bmatrix} \mathbf{diag}\left\{i_{C_s}\right\}_{s=1}^{S-1} \\ \mathbf{0}' \end{bmatrix}$, $\mathbf{E}_l = \begin{bmatrix} \mathbf{I}_{T-1} \\ \mathbf{0}' \end{bmatrix} \otimes i_C$, $C = \sum_{s=1}^{S} C_s$,

i_k is $k\times1$ vector of ones, $\mathbf{I}_k$ is $k\times k$ identity matrix, $\mathbf{J}_k$ is $k\times k$ matrix of ones.

3. Estimation technique

The asymptotic approximation is taken as $S \to \infty$, $C \to \infty$ and *T* stays fixed. This allows consistent estimation of the structural parameters b and g, as well as the time effects l_t. The state effects are not estimated consistently, but this inconsistency does not carry over to the structural parameters and time effects (Baltagi 2001). Of course, the number of states is fixed, but it is sizable ($S = 47$), and some of the states have just a few counties (e.g., Delaware). This makes estimates of the individual effects not quite reliable.

Let $\mathbf{N} = \mathbf{diag}\left\{\mathbf{diag}\left\{n_{ct}\right\}_{c=1}^{C}\right\}_{t=1}^{T}$. Then the variance of the error term is

$$\Omega = s_n^2 \mathbf{I}_T \otimes \mathbf{diag}\left\{\mathbf{J}_{C_s}\right\}_{s=1}^{S} + s_e^2 \mathbf{N}^{-1}.$$

The spectral decomposition of Ω (Baltagi 2001) is

$$\Omega = \left[s_n^2 \mathbf{I}_T \otimes \mathbf{diag}\left\{C_s \mathbf{I}_{C_s}\right\}_{s=1}^{S} + s_e^2 \mathbf{N}^{-1}\right]\mathbf{P} + s_e^2 \mathbf{N}^{-1}\mathbf{Q},$$

where $\mathbf{P} = \mathbf{I}_T \otimes \mathbf{diag}\left\{C_s^{-1}\mathbf{J}_{C_s}\right\}_{s=1}^{S}$ and $\mathbf{Q} = \mathbf{I}_{CT} - \mathbf{P}$. The matrices P and Q are symmetric, idempotent and mutually orthogonal. This allows one to easily determine the inverse of Ω and its square root (see Baltagi 2001 for the details of this trick):

$$s_e^2 \Omega^{-1} = \mathbf{diag}\left\{\mathbf{diag}\left\{q_{ct}\right\}_{c=1}^{C}\right\}_{t=1}^{T} \mathbf{P} + \mathbf{NQ},\quad s_e \Omega^{-1/2} = \mathbf{diag}\left\{\mathbf{diag}\left\{\sqrt{q_{ct}}\right\}_{c=1}^{C}\right\}_{t=1}^{T} \mathbf{P} + \mathbf{N}^{1/2}\mathbf{Q},$$

where $q_{ct} = \frac{s_e^2}{s_n^2 C_s + \frac{s_e^2}{n_{ct}}}$ and $\mathbf{N}^{1/2} = \mathbf{diag}\left\{\mathbf{diag}\left\{\sqrt{n_{ct}}\right\}_{c=1}^{C}\right\}_{t=1}^{T}$.

For applying the feasible GLS procedure we have to consistently estimate s_u^2 and s_e^2. This preliminary step does not affect the asymptotic distribution of the estimate of the vector d. We follow the approach of Balestra (1973). Note that the second term in the spectral decomposition of Ω depends only on s_e^2. This means that we can estimate s_e^2 by getting rid of the first term with the aid of transformation $\mathbf{Q}\Omega = s_e^2\mathbf{N}^{-1}\mathbf{Q}$. Since $\mathbf{U} \sim (\mathbf{0}, \Omega)$, $\mathbf{QU} \sim (\mathbf{0}, \mathbf{Q}\Omega\mathbf{Q}') = (\mathbf{0}, s_e^2\mathbf{N}^{-1}\mathbf{Q})$ due to the symmetry of Ω and idempotency of $\mathbf{Q}$. It follows that an estimator of s_e^2 may be formed as $\hat{s}_e^2 = \dfrac{\hat{\mathbf{U}}'\mathbf{N}^{-1}\mathbf{Q}\hat{\mathbf{U}}}{T(C-S)}$, where $\hat{\mathbf{U}}$ is a vector of OLS residuals. Having obtained a consistent estimator for s_e^2, one can proceed to estimation of s_u^2. If one applies **P** to **U**, one gets

$$\mathbf{PU} \sim (\mathbf{0}, \mathbf{P}\Omega\mathbf{P}') = \left(\mathbf{0}, \left[s_u^2\mathbf{I}_T \otimes \mathbf{diag}\ \left\{C_s\mathbf{I}_{C_s}\right\}_{s=1}^{S} + s_e^2\mathbf{N}^{-1}\right]\mathbf{P}\right),$$

which allows one to estimate s_u^2 by $\hat{s}_u^2 = \dfrac{\hat{\mathbf{U}}'\mathbf{P}\hat{\mathbf{U}} - \hat{s}_e^2\ tr\left(\mathbf{N}^{-1}\mathbf{P}\right)}{TC}$. If $\hat{s}_u^2 < 0$ one may set $\hat{s}_u^2 = 0$.

Denoting by $s_e\hat{\Omega}^{-1/2}$ with imputed estimates for s_u^2, s_e^2 and computing estimates of q_{ct} for all *c* and *t*, one runs OLS on the transformed equations

$$s_e\hat{\Omega}^{-1/2}\mathbf{Y} = s_e\hat{\Omega}^{-1/2}\mathbf{R}d + s_e\hat{\Omega}^{-1/2}\mathbf{U}$$

to get a consistent and asymptotically efficient estimate of d and of its variance matrix. One may iterate this procedure using previous step GLS residuals for estimating s_u^2 and s_e^2 until convergence.

4. The data

The greatest difficulty of implementing the model is the need to have electoral, demographic and socioeconomic data. Since these different categories of data have been collected separately by different agencies, the task of combining them proves to be difficult. First, one has hard time finding periods when the congressial elections were held and for which demographic and socioeconomic records are available. Second, usually data sets contain different variables for different periods of time, and even if not, the units of measurement of a variable or its precision may vary over time, not to mention the fact that different studies give different values for the same variable. Also, the data sets are often highly blanked.

In this paper I use four data sources: Clubb, Flanigan and Zingale (1986), Cohen and Gardner (1992), Inter-university Consortium for Political and Social Research (1970) and Condon-Rall (1989). While the first and fourth sources do seem to be reliable, the other two are full of the abovementioned drawbacks. Therefore, I have had to balance between the desire to draw more information from the data sets and resistance to bring the problem to a one of lower dimension (e.g., shrinking it to just a cross-sectional one). Especially this concerns the choice of a set of regressors (see below).

The data I use cover three periods of time: years 1930, 1940 and 1950 (so that T = 3). Not only are the electoral data quite complete for these periods, but also common basic demographic and socioeconomic characteristics are available. The decade gap between the periods serves additionally as a source of larger variation in the covariates. The data set spans C = 2748 counties in S = 47 states (Alaska, District of Columbia, Hawaii and Georgia are excluded because they do not have electoral data for some of the years). The set of county-specific regressors consists of: the gender factor (share of male population), the race factor (share of non-white population), the average age adjusted by the average age of the nation, a square of the previous regressor, and the education factor (share of illiterate population above certain age). The only state-specific variable I use is per capita income in a state relative to that of the nation. Also, I have complete data on total numbers of voters for all counties.

5. The results

For performing calculations GAUSS 3.1.19 for Windows was used. The feasible GLS procedure was iterated 24 times until the convergence up to 5 decimals was reached. The estimates for the structural coefficients and time effects are gathered in Table I. All structural coefficients look natural: the male factor acts strongly in favor of Republicans, the race factor acts in favor of Democrats, although not to such extent as one might imagine. The effect of the education factor is that the illiterate people prefer Democrats, and very strongly. A better relative economic performance of a state makes good to the Democratic Party if one is willing to set the significance level larger than 9%. Strangely enough, the county average age is not statistically significant when the control over the other covariates is taken.

Since the algorithm takes year 1950 as a "base" one, the signs and magnitudes of time effects tell us that 1930 was a year of a relative Democratic glory (just before the Great Depression) while 1940 was one of a relative Democratic fall (at the beginning of World War II). Finally, the state effects present an interesting pattern (see Table II). These pure state effects may be interpreted as intrinsic political preferences of states' population. What a naked eye can catch is that the m_s, taken at face value from the above tables, reveal clustering exhibiting certain neighborhood gravitation. Figure 1 represents the pattern of the division of states by the state effects – indices of political influence of the parties (remember that Georgia fell out of the data set).

A look at the figure shows that although the evidence for neighborhood gravitation is strong, there are some exceptions. For example, Utah does not obey tastes of its neighbors. All exceptions have various (for instance, historical) reasons, which go beyond what the control of covariates. Utah was born as a Mormon religious state, and this is reflected in its intrinsic political preference. In their turn, the clusters of states may have similar reasons to differ from other "political" regions. For example, that the southern states are Democratic may be explained by the fact that this party was the one of the South side during the Civil War era.

References

Austin, E.W., J.M. Clubb, and M.W. Traugott (1981) "Aggregate units of analysis" in *Analyzing electoral history: a guide to the study of American voter behavior* by J.M. Clubb, W.H. Flanigan, and N.H. Zingale, Eds., Beverly Hills: Sage Publications

Balestra, P. (1973) “Best Quadratic Unbiased Estimators of the Variance-Covariance Matrix in Normal Regression” *Journal of Econometrics* **2**, 17–28

Baltagi, B.H. (1993) “Nested Effects” *Econometric Theory* **9**, 687–688

Baltagi, B.H. (2001) *Econometric Analysis of Panel Data*, New York: John Wiley & Sons

Clubb, J.M., W.H. Flanigan, and N.H. Zingale, Eds. (1981) *Analyzing electoral history: a guide to the study of American voter behavior*, Beverly Hills: Sage Publications

Clubb, J.M., W.H. Flanigan, and N.H. Zingale (1986) *Electoral data for counties in the United States: presidential and congressional races, 1840-1972* [Computer file] ICPSR ed. Ann Arbor: Inter-university Consortium for Political and Social Research [producer and distributor]

Cohen, W. and J. Gardner (1992) *Demographic characteristics of the population of the United States, 1930-1950: county-level* [Computer file] ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor]

Condon-Rall, M.E. (1989) *Historical statistics of the United States: colonial times to 1970*, Government Printing Office.

Ghosh, S. K. (1976) “Estimating from a More General Time-Series Cum Cross-Section Data Structure” *American Economist* **20**, 15–21.

Inter-university Consortium for Political and Social Research (1970) *Historical, demographic, economic and social data: the United States, 1790-1970* [Computer file] Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor]

Iversen, G.R. (1981) “Group data and individual behavior” in *Analyzing electoral history: a guide to the study of American voter behavior* by J.M. Clubb, W.H. Flanigan, and N.H. Zingale, Eds., Beverly Hills: Sage Publications

Figure 1. The pattern of clustering by influence of the parties net of covariates

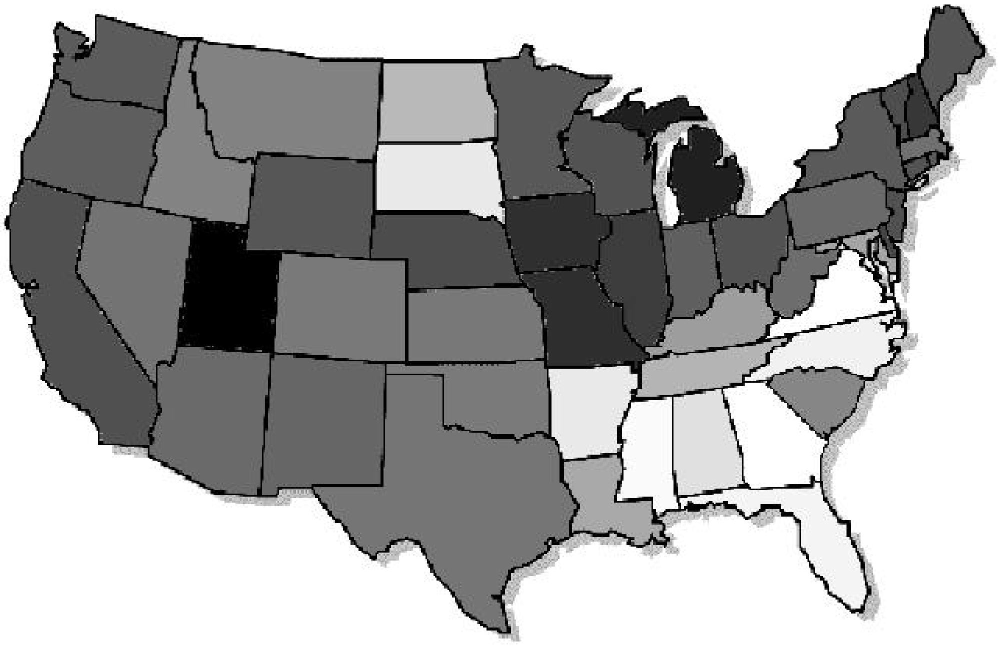

Table I. The structural parameters and time effects

Coefficient	Estimate	StError	t-stat	P-value
β_1 (male)	-.773	.110	-7.01	0%
β_2 (N/W)	.0785	.0138	5.68	0%
β_3 (age)	.0048	.0098	.49	62%
β_4 (age^2)	-.0001	.0002	-.65	52%
β_5 (unedu)	.364	.064	5.68	0%
γ_1 (income)	.0362	.0213	1.70	9%
λ_1 (1930)	.0296	.0157	1.88	6%
λ_2 (1940)	-.0570	.0172	-3.31	0%

Table II. The state effects

State	C_s	Value	State	C_s	Value	State	C_s	Value
			Iowa	99	-.146	Kentucky	38	.217
Connecticut	8	-.026	Kansas	105	.096	Maryland	23	.241
Maine	16	-.025	Minnesota	87	-.010	Oklahoma	77	.102
Massachusetts	14	.067	Missouri	114	-.148	Tennessee	93	.293
New Hampshire	10	-.120	Nebraska	93	-.051	West Virginia	55	.060
Rhode Island	5	-.008	North Dakota	53	.309	Arizona	14	.050
Vermont	14	-.046	South Dakota	67	.467	Colorado	63	.097
Delaware	3	-.073	Virginia	97	.538	Idaho	44	.129
New Jersey	21	-.038	Alabama	65	.435	Montana	56	.126
New York	62	-.004	Arkansas	73	.467	Nevada	17	.077
Pennsylvania	67	.049	Florida	67	.503	New Mexico	31	.034
Illinois	102	-.104	Louisiana	64	.252	Utah	29	-.300
Indiana	92	.013	Mississippi	82	.523	Wyoming	23	-.019
Michigan	83	-.197	North Carolina	100	.487	California	58	-.038
Ohio	88	-.031	South Carolina	44	.140	Oregon	36	.009
Wisconsin	71	-.032	Texas	186	.083	Washington	39	.000