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

A disaggregate approach to economic models of voting in U.S. presidential elections: forecasts of the 2008 election

Stephen Haynes
Department of Economics, University of Oregon

Joe Stone
Department of Economics, University of Oregon

Abstract

By examining disaggregate state-level data, we address two weaknesses of prior estimates of economic voting models in U.S. Presidential elections. First, our disaggregate approach substantially improves statistical power, thus reducing the danger of "over- fitting." Second, our analysis demonstrates systematic differences in voting behavior across states, which have been ignored: voters in higher-income states respond significantly to inflation, changes in the Dow-Jones stock market average, the number of terms the incumbent party has held office, and measures of national security concerns, yet voters in lower-income states respond significantly only to economic growth. Our forecasts for the 2008 U.S. Presidential election predict a statistical dead-heat overall, but a systematic preference for Senator John McCain in lower-income states and for Senator Barack Obama in higher-income states.

The authors acknowledge helpful comments for Bob O'Brien, and excellent research assistance from Kiwako Sakamoto.
Citation: Haynes, Stephen and Joe Stone, (2008) "A disaggregate approach to economic models of voting in U.S. presidential elections: forecasts of the 2008 election." *Economics Bulletin*, Vol. 4, No. 28 pp. 1-11
Submitted: October 6, 2008. **Accepted:** October 19, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume4/EB-08D70032A.pdf

1. Introduction

Well before the 1982 William Clinton campaign phrase "it's the economy, stupid," Kramer 1971, Stigler 1973, and Fair 1978 proposed that voting in Presidential elections is largely determined by economic factors. These models emphasize economic growth, price stability, and the role of parties, and despite very limited degrees of freedom, have significant predictive power for the popular vote for the President. For some elections, however, the predictions of these models go awry, including two recent elections.

In the election of 1992, the models falsely predicted a landslide victory for the incumbent, President George H.W. Bush. Instead, he lost in a close election to Governor William Clinton. In response to this errant forecast, subsequent studies (e.g., Gleisner 1992, Haynes and Stone 1994, and Fair 1996) introduced additional determinants, e.g., the number of consecutive terms the incumbent party held the Presidency, the rate of change in the Dow-Jones stock market average, and changes in the proportion of the population in the military (a proxy for national security concerns). These determinants improved estimates and forecasts of voting models, yet each newly proposed variable raised the danger of “overfitting" given the small number of elections.

The 2004 election again appeared to pose a puzzle. As in 1992, the models (e.g., Fair 2004) predicted a landslide victory in the popular vote for the incumbent President, ironically President George W. Bush, the son of the former President George H.W. Bush, yet the pre-election polls were close and the incumbent President Bush won by a small margin. One obvious omitted factor was the ongoing conflict in Iraq. To address this omission, in Haynes and Stone 2004 we introduced two factors, working in opposition, to account more fully for the potential role of armed conflicts and national security: the first, a “rally round the flag” proxy, which would increase support for an incumbent President, and the second, a proxy for the economic cost of national defense, which can draw support away from an incumbent. We showed that this second factor outweighed the first one in the 2004 election, reducing President Bush's predicted vote share and thereby narrowing the divergence between the model's prediction and both the pre-election polls and the final vote.

A major limitation with all previous tests of models of Presidential elections is the reliance on aggregate voting data, with very few Presidential elections (most estimates are based upon only 20 to 25 observations). Nevertheless, researchers have attempted to address perceived model limitations by introducing additional determinants of voting, e.g., for the 1992 and 2004 elections, which further increases the danger of "overfitting." In this note, we reexamine traditional economic voting models of U.S. Presidential elections by exploring disaggregate state-level data for the U.S. from 1916 through 2008. Our results reaffirm the general findings in previous aggregate estimates, but also reveal novel monotonic patterns in the disaggregate estimates, including that voters in high income states respond to inflation but voters in low income states respond to real growth. We also show that these income-contingent voting patterns have dramatic implications for forecasts of the upcoming 2008 Presidential election between Senator Barack Obama and Senator John McCain.

2. Model

The point of departure for our disaggregate analysis is the general frameworks of Fair (1978, 1988) and Gleisner 1992, where voter utility (U) is determined by economic performance (E) and non-economic factors (N).

$$U = U[E, N] \tag{1}$$

The voter then chooses between the Democratic candidate D or Republican candidate R based on expected outcomes of E and N for each party. Thus, the probability that a Democrat is chosen over a Republican depends on the difference between the corresponding expected utilities for the two parties (for a general derivation, see Judge et. al. 1985, p. 769):

$$VOTE = \text{prob}\,[U(R) < U(D)] \tag{2}$$

We interpret VOTE as the Democratic share of the two-party vote at the aggregate level. VOTE is a continuous variable, but is bounded between zero and one. However, the log-odds transformation of VOTE is unbounded, permitting estimation using ordinary least squares (OLS):

$$V = \log\,[VOTE/(1-VOTE)] \tag{3}$$

In our state-level panel sample, both observed and predicted values of VOTE in the linear specification are concentrated around 0.5, with none outside the bound, supporting OLS estimation of the linear model. For simplicity of interpretation, we therefore present results in linear rather than log-odds form; estimates for the two specifications are in fact very similar.

3. Estimation Equations and Data

Our base equation follows Haynes and Stone 2004, modified for estimation with disaggregate state-level data. The model combines the primary components of Fair (1978, 1996, 2002), as extended by Gleisner 1992 to include a stock market variable and Haynes and Stone 2004 to add variables on the number of consecutive terms the incumbent party has been in power and on national security. Eq. (4) is the resulting estimation equation, where expected signs are listed above the regressors.

$$\begin{array}{c} \overset{+}{} \quad \overset{-}{} \quad \overset{+}{} \quad \overset{+}{} \quad \overset{-}{} \\ VOTE = f\,[PARTY,\ DURATION*P,\ DOWJONES*P,\ GROWTH*P,\ INFLATION*P, \\ \overset{+}{} \quad \overset{-}{} \\ ARMY*P,\ ARMYSPEND*P] \end{array} \tag{4}$$

where VOTE is the Democratic share of the two-party Presidential vote;
PARTY (P) is 1 if the incumbent is a Democrat, and -1 if a Republican (all regressors are interacted with P to permit symmetric treatment of the two parties);
DURATION is the number of consecutive terms the incumbent party has been in power;

DOWJONES is the annual rate of change in the Dow-Jones stock market index, January to October of the election year;
GROWTH is the annual growth rate of real per capita GNP (GDP) in 2nd and 3rd quarters of the election year;
INFLATION is the absolute value of the annualized inflation rate (GNP/GDP deflator) in the two-year period prior to the election;
ARMY is the annualized percentage change of the proportion of the population in the armed forces in the two-year period prior to the election; and
ARMYSPEND is the annualized percentage change in the proportion of government spending devoted to national security in the two-year period prior to the election.

Our sample begins with the 1916 U.S. Presidential election, consistent with Fair 1996. Data on the regressors, which are aggregate U.S. series, are from Haynes and Stone 2004, updated as detailed in the Data Appendix. Aggregate data on the dependent variable VOTE are also from Haynes and Stone 2004, and disaggregate state-level data on VOTE are described in the Data Appendix.

4. Estimates

The dependent variable VOTE is the Democratic share of the two-party Presidential vote. Each regressor is interacted with Party (P) so that interpretation of the impact of a change in a regressor on the dependent variable is symmetric whether the incumbent party is Democratic or Republican. The first column in Table 1, using aggregate data, presents OLS estimates of eq. (4), with t-statistics based on heteroskedasticity-corrected (White) standard errors in parentheses. The remaining columns in Table 1, using state-level panel data, present GLS estimates of eq. (4) based on cross-section weights, include state fixed effects, and report t-statistics based on White cross-section standard errors in parentheses.[1] Reestimation with state-level panel data using OLS and/or excluding state fixed effects (omitted for brevity) yields qualitatively similar estimates, and computing t-statistics without the White cross-section correction yields higher, although biased, t-statistics (see Moulton 1990).

The first column in Table 1, entitled Aggregate, displays significant coefficients with the correct sign on all regressors, and closely replicates the aggregate estimates in Haynes and Stone 2004, but is based on only 23 observations.[2] The second column, entitled Panel, repeats the estimates for the same specification as in column one, but is based on the state-level panel data with 1126 observations. The Panel estimates in column two are generally very similar to those in column one, with the exception of the loss of significance for the

[1] GLS cross-section weights are used because of the large differences in the sizes of the 50 U.S. states; fixed-effects permits intercept differences across the states; and White cross-section standard errors are necessary to prevent downward bias in the standard errors resulting from the use of identical regressors across states (see Moulton 1990 for a theoretical discussion of this problem).

[2] The two estimates are not exact because the samples differ: the estimate in Haynes and Stone 2004 is from 1908 through 2000, while the column one estimate is from 1916 through 2004.

coefficient on Army. The close similarity of the estimates in these two columns suggests that the findings and conclusions in the many aggregate voting estimates presented in the literature the past fifteen years are reliable despite the fact they are based on only 20 or so observations of data.

The final five columns in Table 1, labeled Y1 through Y5, summarize estimates of eq. (4) based on quartile subsets of the state-level panel data after ordering the states by their per capita personal income at the mid-sample year of 1960. Thus, the Y1 column is eq. (4) estimated for the ten lowest income states, Y2 for the next ten higher income states, up to Y5 for the highest ten income states.[3] Comparison of the coefficients on each regressor across income quartiles shows a consistently monotonic income-contingent voting pattern. The coefficient on duration smoothly declines from insignificant to significantly negative as income increases (i.e., shifting from the estimate in column Y1 to that in column Y5), indicating that only voters in the higher-income states are more adversely influenced by the number of terms the incumbent party has been in power. The coefficient on DowJones smoothly increases from insignificant to significantly positive as income increases, and the coefficient on Inflation smoothly declines from insignificant to significantly negative as income increases, patterns suggesting that higher-income citizens, unlike lower-income citizens, vote based on economic outcomes that impact the value of their asset holdings. Alternatively, the coefficient on Growth declines from significantly positive to insignificant as income increases, indicating that lower-income citizens, unlike higher-income citizens, vote based on economic outcomes that impact their employment status. Finally, both military variables become significant as income increases (positive for Army, and negative for ArmySpend), supporting the importance of both dimensions of national defense determinants only for higher-income voters. These novel monotonic patterns in the regression coefficients across income quartiles, with all coefficients correctly signed when significant, remain robust to alternative estimation methods.

5. Forecasts of the 2008 Election

Out-of-sample forecasts for the upcoming 2008 election are reported in the next-to-last row in Table 1, where the most updated values for the regressors are used in the forecast (see the Data Appendix for these values). The Aggregate and Panel election forecasts of the overall Democratic vote share are extremely close to 50.0 percent, a predicted dead-heat. Given the similarity in the regression coefficients in the aggregate and the state-level panel equations, the similarity in these forecasts is not surprising.

[3] Y1 states are Alabama, Arkansas, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, South Dakota, Tennessee, and West Virginia; Y2 states are Idaho, Louisiana, Maine, North Dakota, Nebraska, New Mexico, Oklahoma, Utah, Virginia, and Vermont; Y3 states are Arizona, Florida, Hawaii, Iowa, Kansas, Minnesota, Missouri, Montana, New Hampshire, and Texas; Y4 states are Colorado, Indiana, Maryland, Ohio, Oregon, Pennsylvania, Rhode Island, Washington, Wisconsin, and Wyoming; and Y5 states are Alaska, California, Connecticut, Delaware, Illinois, Massachusetts, Michigan, New Jersey, Nevada, and New York.

The pattern in the election forecasts across income quartiles differs consistently from the bottom to the top quartiles. The Democratic vote shares (in percent), predicted for the five income quartiles beginning with Y1, are 44.60, 44.66, 50.40, 51.09, and 53.35, hence increase monotonically as income increases.[4] The bottom row summarizes the fraction of the total popular vote attributed to each income group based on voting data from the 2004 Presidential election. The highest fraction of votes comes from the highest income quartile, Y5, which is not surprising since two states in Y5 are the high population states of California and New York.

In sum, the out-of-sample forecasts predict a statistical dead-heat in the U.S. popular vote for the 2008 Presidential election, but a systematic monotonic preference for Senator John McCain in lower-income states and for Senator Barack Obama in higher-income states.[5]

6. Concluding Comments

This note examines the reliability of aggregate estimates of U.S. voting in Presidential elections, typically based on only 20 or so observations, by re-estimation with state-level panel data, and reports forecasts of the 2008 U.S. Presidential election with these data. We present three conclusions. First, aggregate estimates using data from 1916 to 2004, with only 23 observations, are very similar to state-level panel estimates for the same years based on 1126 observations, suggesting surprising reliability of the findings in the aggregate Presidential voting literature despite limited degrees of freedom.

Second, after partitioning the U.S. states into quartiles based on income level, we find a novel and monotonically consistent income-contingent pattern across the coefficients on each regressor. For the higher income states, increases in the number of consecutive terms a party has been in office reduces their probability of reelection. Also, for higher-income states, increases in the change in the Dow-Jones stock market index improves the incumbent's probability of reelection, and increases in inflation reduces this probability. Conversely, only for lower-income states, increases in real growth improves the incumbent's probability of reelection. Finally, national defense factors, both the positive dimension ("rally-round-the-flag" motive) and the negative dimension (the costs of sustaining a war), are significant for higher-income states but not lower-income states.

---

[4] The 2008 values of the regressors used in the forecasts were taken from data through September 17, 2008. If these regressor values were computed two weeks later, deeper into the U.S. financial crisis, the forecast percents would have been higher than 51.09 for Y4 and 53.35 for Y5 because of the dramatic drop over those two weeks in the Dow-Jones stock market index.

[5] Every election includes idiosyncratic determinants unique to that election. For the 2008 election, these would in part include the race and gender of the candidates, Black turnout, new voter registration, and the financial crisis of fall 2008 (beyond its direct impact on the economic regressors in the model). Of course, no model, no matter how elaborate, can fully capture in advance all factors that influence voters the day of the election.

These income-contingent patterns in the regression coefficients represent an important yet neglected dimension of voting behavior for the U.S. President. The pattern of significance of the inflation and the Dow-Jones variables only for higher-income states, where presumably wealth preservation is a stronger motive than employment stability, in combination with significance of real growth only for lower-income states, where employment stability is likely a stronger motive than wealth preservation, is consistent with evidence in Joyadev (2006, p.71), who in a different context concludes that "the poor are more likely than the rich to prefer that unemployment be controlled rather than inflation (they are less relatively inflation averse)".

Our third general finding concerns forecasts for the 2008 U.S. Presidential election. Both aggregate and state-level panel data predict a statistical dead-heat. However, disaggregation shows that forecasts systematically differ by state income level. The lowest-income quartile of states clearly prefers Senator John McCain by a significant margin (55.4% to 44.6%); yet the middle-income quartile of states is almost evenly split between the two candidates; and finally the highest-income quartile of states prefers Senator Barack Obama by a substantial and significant margin (53.35% to 46.65%). This evidence suggests that predictions of 2008 voting in key "swing" states should include income-contingent differences in state-level voting preferences, especially in the context of the fall 2008 U.S. financial crisis.

In sum, this research presents a potentially rich avenue to understand voting behavior in U.S. Presidential elections which has heretofore been ignored, and many questions remain. On the empirical side, our findings should be explored using other groupings of states and samples, other estimation methods such as interaction of the regressors with state income, and inclusion of control factors such as education level and race to understand the role of income in our evidence. On the theoretical side, the dramatic and monotonic income-contingent patterns in the voting coefficients should be investigated and related to other determinants of voting in U.S. Presidential elections.

References

Fair, R.C. (1978) "The Effect of Economic Events on Votes for President" *Review of Economics and Statistics* **60**, 159-173.

_____ (1988) "The Effect of Economic Events on Votes for President: 1984 Update" *Political Behavior* **10**, 168-179.

_____ (1996) "Econometrics and Presidential Elections" *Journal of Economic Perspectives* **10**, 89-102.

_____ (2002) "The Effect of Economic Events on Votes for the President: 2000 Update" *unpublished* http://fairmodel.econ.yale.edu/RAYFAIR/FDF/2002DHTM.HTM

_____ (2004) "Presidential Vote Equation--July 31, 2004" *unpublished* http://fairmodel.econ.yale.edu/vote2004/vot0704.htm

Gleisner, R.F. "Economic Determinants of Presidential Elections: The Fair Model" *Political Behavior* **14**, 383-394.

Haynes, S.E. and J.A. Stone (1994) "Why Did Economic Models Falsely Predict a Bush Landslide in 1992?" *Contemporary Economic Policy* **12**, 123-130.

_____ (2004) "'Guns and Butter' in U.S. Presidential Elections," *Economics Bulletin* **1:5**, 1-8.

Jayadev, A. (2006) "Differing Preferences Between Anti-Inflation and Anti-Unemployment Policy Among the Rich and Poor," *Economics Letters* **91:1,** 67-71.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.C. Lee (1985) *The Theory and Practice of Econometrics*, John Wiley and Sons: New York.

Kramer, G.H. (1971) "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964," *American Political Science Review* **65,** 131-43.

Moulton, B.R. (1990) "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *The Review of Economics and Statistics* **72,** 334-338.

Stigler, G.J. (1979) "General Economic Conditions and National Elections," *American Economic Review* **63,** 160-67.

Bureau of Economic Analysis, *Current-Dollar and Real Gross Domestic Product.* http://www.bea.doc.gov/bea/dn/home/gdp.htm

Bureau of Economic Analysis, *Regional Economic Accounts: State Annual Personal Income.* http://www.bea.gov/regional/spi/cfm?statable=SA05

Dow Jones and Co, *DowJones Indexes*. http://www.djindexes.com/jsp/index.jsp.

U.S. Census Bureau, Statistical Abstract of the United States, Section 9: Federal Government Finances and Employment. http://www.census.gov/prod/www/statistical-abstract-02.html

U.S. Census Bureau, Statistical Abstract of the United States, Section 10: National Defense and Veterans Affairs. http://www.census.gov/prod/www/statistical-abstract-02.html

U.S. National Archives and Records Administration. http://www.archives.gov/federal-register/electoral-college/popular_vote.html

U.S. Presidential Elections. http://www.multied.com/elections/states/list.html

Table 1
Democratic Share of US Presidential Two-Party Vote -- Eq. (4) -- 1916-2004

Variable	Aggregate	Panel	Y 1	Y 2	Y 3	Y 4	Y 5
Intercept	48.76**	48.81**	51.60**	43.99**	48.19**	48.48**	50.08**
	(46.63)	(49.92)	(28.45)	(39.36)	(44.90)	(48.96)	(49.11)
Party (P)	7.34**	7.24**	-2.10	3.50	6.69**	10.67**	13.37**
(+)	(3.39)	(3.05)	(-0.59)	(1.32)	(2.63)	(4.00)	(4.95)
Duration*P	-1.66*	-1.71*	0.11	-1.39*	-1.74*	-2.24**	-2.46**
(-)	(-1.84)	(-2.03)	(0.11)	(-1.78)	(-1.85)	(-2.48)	(-2.30)
DowJones*P/100	11.53*	8.72*	-8.19	3.00	6.98	16.34**	15.95**
(+)	(2.44)	(1.84)	(-0.75)	(0.53)	(1.22)	(3.32)	(2.52)
Growth*P	0.50**	0.73**	2.00**	1.40**	0.91**	0.18	-0.18
(+)	(2.83)	(4.66)	(7.69)	(9.11)	(5.45)	(0.95)	(-0.81)
Inflation*P	-0.73*	-0.66*	0.81	-0.09	-0.53*	-1.17**	-1.69**
(-)	(-2.46)	(-2.17)	(1.29)	(-0.24)	(-1.60)	(-3.38)	(-4.34)
Army*P/100	3.88*	0.14	-7.39	-4.28	-1.05	3.32*	9.25**
(+)	(2.26)	(0.09)	(-1.91)	(-2.08)	(-0.53)	(1.85)	(4.04)
ArmySpend*P/100	-8.10**	-6.94*	8.15	-2.31	-6.73*	-11.05**	-17.30**
(-)	(-2.78)	(-2.25)	(1.96)	(-0.73)	(-2.05)	(-3.19)	(-4.58)
Adjusted R-squared	0.75	0.55	0.54	0.61	0.58	0.64	0.65
SE of Regression	3.54	9.59	11.81	9.46	8.49	6.18	6.34
Number Observ.	23	1126	228	230	219	230	219
Pred. Vote 2008	49.98	50.14	44.60	44.66	50.40	51.09	53.35
(Conf. Interval)	(±3.95)	(±1.58)	(±4.47)	(±3.87)	(±3.84)	(±2.50)	(±3.04)
Y Group Weight			0.147	0.090	0.220	0.219	0.323

**Significant at one percent level, and *significant at five percent level, one-tail tests. Dependent variable is the Democratic share of the two-party vote. Aggregate estimate is based on total US Democratic share; Panel estimate is based on state-by-state shares; and Y1 through Y5 panel estimates are state quartile groupings ordered by per capita real GDP, where Y1 is the lowest ten income states, Y2 is the next higher ten income states, etc. See text for explanation of the regressors; expected signs are summarized below these variables. The aggregate equation is estimated with ordinary least squares, and robust (White) t-statistics are in parentheses. All panel equations are estimated with generalized least squares based on cross-section weights, include cross-section fixed effects, and robust (White cross section) t-statistics are in parentheses.

DATA APPENDIX FOR TABLE 1

YEAR	VOTE	PARTY	DURA-TION	DOW JONES	GROWTH	INFLA-TION	ARMY	ARMY-SPEND
1916	51.682	1	1	12	6.38	7.73	2.33	4.04
1920	36.119	1	2	-23.5	-6.14	8.01	-107.6	11.24
1924	41.756	-1	1	6.0	-2.16	0.62	-3.38	-23.05
1928	41.240	-1	2	31.3	-0.63	0.81	-0.48	10.15
1932	59.140	-1	3	-25	-13.98	10.01	-2.97	-37.56
1936	62.458	1	1	24.9	13.41	1.36	7.6	28.86
1940	54.999	1	2	-12.9	6.97	0.53	16.79	8.33
1944	53.774	1	3	9.0	6.88	1.98	53.1	17.16
1948	52.37	1	4	6.3	3.77	10.39	-38.82	-86.56
1952	44.595	1	5	-1.8	-0.34	2.66	43.89	71.59
1956	42.240	-1	1	2.4	-0.69	3.59	-9.93	-14.34
1960	50.090	-1	2	-13.9	-1.92	2.16	-4.1	-8.44
1964	61.344	1	1	15.8	2.38	1.73	-3.68	-5.88
1968	49.596	1	2	10.0	4.0	3.94	0.06	6.28
1972	38.21	-1	1	5.4	5.05	5.17	-11.91	-19.71
1976	51.05	-1	2	3.0	0.78	7.64	-2.56	-20.15
1980	44.697	1	1	12.4	-5.69	8.99	-1.37	-0.44
1984	40.83	-1	1	-6.9	2.69	3.68	-0.22	7.38
1988	46.07	-1	2	12.6	2.43	3.3	-1.58	-1.09
1992	53.455	-1	3	-0.9	1.34	3.15	-7.33	-10.11
1996	54.736	1	1	24.54	3.08	1.95	-5.62	-12.67
2000	50.265	1	2	-5.02	2.95	1.8	-2	1.83
2004	48.586	-1	1	-8.01	3.49	2.50	-0.51	14.91
2008		-1	2	-30.7*	2.10*	3.70*	-0.87*	0.41*

*Preliminary estimate, with computation date of September 17, 2008.

Notes: Above are aggregate data, and through 2000 are from Haynes and Stone (2004, p. 11), where the incumbent share of the two-party vote is changed to VOTE, the Democratic share of the two-party vote. For 2004 and 2008 data, DOWJONES are from Dow Jones and Co.; GROWTH and INFLATION from the Bureau of Economic Analysis: Current-Dollar and Real Gross Domestic Product; ARMY from U.S. Census, Statistical Abstract of the United States, Section 10, National Defense and Veterans Affairs; and ARMYSPEND from U.S. Census Bureau, Statistical Abstract of the United States, Section 9: Federal Government Finances and Employment. Disaggregate state-level data on VOTE, not presented in this appendix, are computed from popular vote data in U.S. Presidential Elections and U.S. National Archives and Records Administration. Per capita personal income by state for mid-sample year 1960 are from the Bureau of Economic Analysis: Regional Economic Accounts.