What Exactly is Going on in the WiscAds Data?

The Evolution of Campaign Ad Strategy Over the 2008 Campaign*

Marc Ratkovic

March 30, 2010

Abstract

Over 1.33 million campaign commercials designed to influence a race for the US House, Senate, or Governorship were run in 2008. This data, gathered from America's top 102 media markets by the Wisconsin Advertising Project, provides a means to illuminate the evolution of ad tone, issue density, and cost over time. The basic problem is one of functional form and variable selection. To account for this uncertainty, I use a series of regression trees to characterize this dataset, selecting locally homogenous subsets of the data, and fitting a linear time trend to each. I find, across models, that behavior on Fox differed in terms of both expenditure and tone, but not issue density. As well, candidates went negative on taxes, but positive on social welfare policy. Evidence also suggested that including a website in an ad was a good predictor of issue density.

*Preliminary draft prepared for the American Politics Workshop, UW-Madison, April 5, 2010. With thanks to Jim Jaczwiecki and Wei-Yin Loh for technical help. Thanks to David Weimer and Amber Wichowsky for useful comments, and to Sarah Niebler, Jake Neiheisel, and Matt Holleque for help with issues regarding the WiscAds Data.

1 Introduction

Do campaign messages converge or diverge? A Downsian model (1957) would suggest convergence, in order to capture the middle, perhaps mitigated by party loyalists with polarized demands (Hirschman 1970). Strategic differentiation, issue ownership, and the the introduction of new issue dimensions in a political discourse can shift the outcome and very landscape of competition (Riker 1996, Baumgartner and Jones 1994, Petrocik 1996, e.g.). In actuality, the tone and issue density of campaign ads are segmented and highly differentiated, aimed at particular audiences, and vary systematically by tone, cost, and issue type.

An estimated 530 million dollars were spent in the 2008 campaigns for House, Senate, and Governor in America's top 102 media markets. Enmeshed in this decision to spend is a series of other strategic decisions, regarding choices of issue, theme, and tone, as well as placement, cost, and time. An ongoing dataset gathered by the Wisconsin Advertising Project[1] has provided a fantastic way to analyze the contours of these decisions. From November 2007-November 2008, every campaign commercial in the top 102 media markets aired on a major affiliate was coded among a variety of dimensions, for above 1.3 million commercials total. The coding methods, summarized below, were crafted to ensure strong internal validity in these necessarily subjective matters.

This data was gathered with the intent to add to the "campaign effects" literature. The major question is whether advertisements, presumably negative, primarily inform, persuade, or mobilize voters (Goldstein & Ridout 2004). There are two key concerns: exposure and persuasion. The key problem is that the subject exposed to the ad may not recall either observing the ad, or may be a poor judge of the ad's effect on their decision-making. In order to have an effect, a campaign ad must be received by the voter through watching television, accepted by the voter so that it is remembered, and later recalled when the decision to vote and for which candidate is made. Each link in this chain has been explored

[1]The data was obtained from a project of the University of Wisconsin Advertising Project includes media tracking data from TNSMI/Campaign Media Analysis Group in Washington, D.C. The University of Wisconsin Advertising Project was sponsored by a grant from The Pew Charitable Trusts. The opinions expressed in this article are those of the author(s) and do not necessarily reflect the views of the University of Wisconsin Advertising Project or The Pew Charitable Trusts.

by many authors. We know voters have rather poor recall about specific ads or positions they perceive (Niemi, et al. 1980; Ansolabehere, et al. 1999), but this only implies that "campaign effects" should not be measured through respondent recalled exposure to the medium. Instead, message reception may be the better measure of exposure (Zaller 1992). When tied directly to turnout, some experimental evidence suggests that negative advertising may reduce voter turnout by as much as 5% (Ansolabehere 1994; Ansolabehere and Iyengar 2005).

Previous research using this data (Coleman and Manna 2000, e.g.) has worked to link the ads to surveys, voter turnout, and campaign outcomes. Due the nature and size of the data, these works have been able to provide only basic summaries of WiscAds data, and most work is done with outcomes aggregated to the race- or media market-level. In short, the works look for subsets of the data that the researcher finds interesting, and proceed to analyze it in pieces. While useful for addressing a given particular concern, this begs the meta-question of identifying the best ways to subset and aggregate the data.

Here, I consider the whole of the dataset, searching for commonalities of tone, issue, and cost. Rather than analyze aggregates of the data, I take each ad as an observation, and examine how they vary in content, cost, and tone over time. Specifically, I search for systematic predictors that can help answer the following two questions:

1. How does ad tone (positive, contrast, negative) vary over time?

2. How does issue density (number of issue areas mentioned) vary over time?

3. When are more costly ads run?

The substantive goal is to provide a description of ad tone and density through the course of the campaign. Rather than look at the impact of ads, I am concerned with the nature of the ads that candidates and other groups air over time. Specifically, I search for subsets of the data that are particularly positive or negative in tone, mention fewer or more issue areas, and are relatively more or less expensive, each over time.

Three basic sets of results emerge. Across subsets, the ads grew more negative over time. Ads on Fox tended to be more negative, while, on the other major networks, social welfare policy was used to promote candidates. Both on and off Fox, mentioning taxes was used to attack the opposition. While not a "Fox Effect," this does imply a selection bias in the placement of ads. On the whole, two different sets of messages appear in campaign advertising. On Fox, taxes in particular, and economic policies in general, were used as attack ads. On the other affiliates, taxes were brought up to promote the favored candidate or attack the opponent. Social welfare policy served as a discriminating variable on non-Fox affiliates, while it was not selected on Fox itself.

Second, issue density varied primarily with whether the ad was encouraging the viewer to vote for or against someone. Negative ads that mentioned taxes mentioned, generally, only that issue. Promotional ads contained more issue areas, and ads that mentioned healthcare and a website contained more issue areas over time. Affiliate effects are less prominent in issue density.

Third, attack ads purchased by campaigns are relatively more expensive than other ads. On Fox, Republicans purchased relatively more expensive ads, while Democrats only did so in races with high commercial advertising expenditures.

2 Method, Data, and Models

The primary statistical problem in analyzing such a complex dataset is determining functional form. The dataset provides a massive amount of data, and a series of plausible covariates, but no real indicator of the functional form linking the two. Even more troublingly, with 1.3 million observations, most any hypothesis will be significant at standard 5% levels, and the data can support a massive number of two-, three-, and higher order interactions. I presuppose that effects are neither linear nor continuous, and instead search the dataset for homogenous pockets that are maximally different from adjacent areas in the parameter space.

The method used here, regression trees, are a standard and conceptually simple datamining method. The method searches through independent variables for ideal splits, maximizing within-partition homogeneity and cross-partition heterogeneity. Regression trees often provide superior predictions to linear models (Loh 2008a, 2008b), while offering an ease of interpretation over standard regression models. The primary shortcoming of trees is their lack of a firm inferential foundation.

The first regression tree algorithms, AID and CHAID, date to the 1960's. These methods, Automatic Interaction Detection, and Chi-squared Automatic Interaction Detection, were designed to help uncover interactive effects in the context of linear models. The modern development of these methods began with the development of CART (Classification And Regression Trees; Breiman, et al. 1984). The CART algorithm is available in R in package RPART (Recursive Partitioning). C4.5, Bayesian Additive Regression Trees, Random Forests and other ensemble methods, and the algorithms of Wei-Yin Loh (GUIDE, LOTUS, CRUISE) all provide extensions and alternatives to this basic idea (Loh 2008a provides an overview and relevant references).

Appendix B illustrates a simple tree. Reading a regression or classification tree is straightforward. If a given condition is met, follow the branch to the left; else, follow the right branch of the tree. Fitted values are simply the sample mean of the terminal nodes. The top node is 1. Each successive node to the left is numbered twice its parent node, while the node to the right is twice plus one.

Three primary concerns arise when considering tree-based methods: pruning, variable selection, and handling categorical variables. Pruning, or when to stop "growing" the tree is generally governed by a cross-validation criterion. The data is fit to 9/10 of the data, used to predict the remaining 1/10, and this is done in turn, at each possible tree size. The tree size that minimizes predictive error selected is that which minimizes cross-validated risk, plus one or one-half of a standard error. This serves as the primary tuning parameter.

Variable selection in CART is done by considering every possible split in each predictor, and splitting at that point. Loh (2008a) summarizes his own work on how this leads to biased selection. Consider two predictors, one that takes on two unique values and one that takes on eighty. The second predictor has more potential splits, and will likely be favored by an exhaustive search. Loh's algorithm, GUIDE (Loh 2008b), provides unbiased selection by selecting the variable first, and splitting second. Variable selection is done through a series of cross-tabulations, with chi-squared values converted to a p-value and then compared on this scale.[2]

A second advantage of GUIDE occurs with categorical variables. In CART, categorical variables are given weights through a "backfitting" algorithm. This involves fixing all categories but one, estimating that one, and cycling through all categories. In practice, this takes quite a bit of time, as variables with ten categories can overwhelm the algorithm. GUIDE generates a linear discriminant to produce category weights at each potential split, greatly reducing the computational load.

Regression trees are exploratory and predictive, rather than inferential, models. In complex datasets, they can illuminate structures that would be lost to linear models. They are not inferential, though, in that no estimates of uncertainty in the tree structure itself are provided. Analyzing predictive models can help uncover the nature of social processes (Friedman 1966, e.g.), though the nonlinearities generate the possibility of models that are "close" in prediction error being dramatically different in structure. Each split in the tree is conditional on all previous splits, so a change in an early split on a bootstrapped or validation sample will affect the splits below it.[3]

For this reason, the analysis presented here is exploratory rather than inferential. Rather than test a given hypothesis, the method searches among all hypotheses for the one with the most evidence in its favor, and iterates the process in each subsequent partition of the data. The emphasis is on discovery, rather than inference (see Kritzer (1996) and Grimmer and King (2009) for similar approaches in political science).

This method has appeared little in political science, with two notable questions. Martin, Quinn, Ruger, and Kim (2004) demonstrate that CART trees are better at predicting

[2] In practice, many of these p-values are often indistinguishable from zero. Instead, the algorithm uses the Wilson-Hilferty approximation to convert a χ^2_v variable to a χ^2_1, and predictors are compared on this scale. Regressions are handled by considering cross-tabs on the signs of the residuals.

[3] GUIDE considers all interactions of variables in the selection step, in order to help ameliorate this tendency.

Supreme Court votes than panels of experts, and Imai and Strauss (2009) use it to formulate optimal get-out-the-vote strategies. These applications are particularly well-suited to trees, as they are both fundamentally prediction problems. Rather than use trees for prediction, I use them here to uncover local patterns in a massive dataset with many covariates.

2.1 Description of Data

The data used here consist of 1,337,320 observed campaign commercials aired in the top 102 media markets on every major affiliate, for 24 total between November 2008 and November 2009. Each specific airing of an ad counts as an observation. The ads in the sample include every ad designed to influence a Senatorial, gubernatorial, or Congressional. This includes ads from the parties, the candidates, and interest groups. Video clips of the commercials were gathered by the Campaign Media Analysis Group (CMAG) in Washington, DC. They included the time and date of the ad; the show, station, and affiliate; and the media market in which the ad is aired. Remaining variables were coded by the Wisconsin Advertising Project. These variables assess the theme, content, and tone of the ad. Internal validity was a key concern. The variables were coded by teams of undergraduates independently, with supervisors mediating when the independent reviewers did not agree. A complete description of the data collection method can be found in Franz, et al. (2006).

A key concern in this analysis is the ignored simultaneity biases. Candidates that spend more on ads are more likely to lose, though the causal explanation runs in the opposite direction. Similarly, negative ads by a challenger may prompt negative or positive ads by an incumbent. A full game theoretic model, and suitable instruments, would be necessary to disentangle these different effects. Yet, neither instruments not the empirical implications of a game theoretic model are robust to model misspecification. Though correlation is not causation, the two are highly correlated,[4] and non-obvious correlations in this data set shed light on broad trends in campaign strategy.

[4] I *wish* this were my line. I think it's Michael Berube's, but I could not track it down.

2.2 Description of Independent Variables

The variables come in four separate groups. *Campaign-level variables* describe the particular campaign being addressed by the ad. *Placement variables* deal with the placement of the ad, and the use of the candidate within the ad. *Content and Tone* variables deal with the content of the specific ad. *Cost Variables* deal with the cost and media market of each variable. Complete descriptions can be found in the appendix. In the first two models, one dependent and 118 independent models were used. In the final model, explaining relative cost (cost of the ad over mean cost of ads for that race), I dropped estimated cost of that ad from the analysis.

The ad tone was relatively balanced, with 39.8% positive, 20.9% contrasting, and 39% negative. Issue density was calculated as the number of policy areas mentioned, with a possible range of zero to six (Economic Policy, Social Policy, Law Policy, Social Welfare Policy, Defense Policy, and Other). Values were observed on zero to five. Each policy area was subsequently broken into specific policy issues, as described in the appendix. Ads mentioning taxes and debt resulted in these variables coded as one.[5] Since some non-zero number of economic policies were mentioned, the economic policy area variable was coded one. Issue density was calculated as the sum of policy areas mentioned, and this was well spread (0=6.8%, 1=36.5%, 2=38.4%, 3=15.2%, 4=2.9%, 5=0.1%). The major networks accounted for 93.6% of observed commercials (ABC=26.7%, CBS=26.5%, NBC=28.5%, FOX=11.9%), with the remainder spread across smaller affiliates (notably UPN=2.6% and WB=1.8%). Other affiliates contained less than one percent of the total ads. ABC, CBS, and NBC agreed closely on the distribution of ad tone. Plus or minus 1.5%, each had about 41% positive, 21% contrasting, and 38% negative. Fox had 35.5% positive, 21.4% contrasting, and 43.0% negative.

Senate races comprised 43.8% of the ads, with House races at 41.9%, and governor races at 14.2%. Of the 940,230 ads that mentioned an economic policy, 519,493 (55%) mentioned

[5] Personal correspondence with Sarah Niebler revealed concerns over inter-coder reliability for particular policy issues. More details will be added on this.

taxes. The modal ad tone for commercials that mentioned one or two issues was negative (41.7%, 43.9%), while the modal category was positive for zero, three, four, or five issues (0=56.8%, 3=48.8%, 4=56.2%, 5=99.9%). In terms of action types mentioned in the ad, 83.0% called for no action, 3.2% called to re-elect or support someone, 0.5% called to vote against someone, and 7.0% called to write, call, or tell someone something.

2.3 Description of Models

Three separate models are presented, as summarized below. Unless otherwise specified, all other options for each model were set to the default option in GUIDE.

- Model 1: Regression Tree for Ad Tone

  Dependent Variable: Tone of Ad; 1 = Positive, 2 = Contrast, 3 = Negative

  Node Variable: Negative days until the election[6]

  Maximal Number of Levels in Tree: 4

- Model 2: Regression Tree for Issue Density

  Dependent Variable: Number of Policy Areas Mentioned, 0-6 (Economic, Social, Law and Order, Social Welfare, Foreign/Defense, and Other)

  Node Variable: Negative days until the election

  Maximal Number of Levels in Tree: 4

- Model 3: Regression Tree for Relative Cost

  Dependent Variable: Cost of ad, divided by mean cost of ads for all expenditures in that race

  Node Variable: Negative days until the election

  Maximal Number of Levels in Tree: 4

As discussed above, the goal of the analysis is exploration rather than prediction. Pruning based off of cross-validation (within-sample predictive error) produced more than forty terminal nodes, with a regression line at each node. To produce interpretable results, the trees were kept to a maximum of four levels.

[6]The variable in the dataset is days until the election. I convert this to a negative number, so it is increasing in time. The value t=0 corresponds with election day, t= -7 corresponds with one week before the election, and so on.

3 Results

The results from Model 1 can be found in Appendix C. Across subsets, the ads grow more negative over time. The first split separates Fox (mean=2.11) from the other affiliates (mean=1.96). More negative ads were placed on Fox than the other affiliates, and this effect is larger than any other singular covariate in this dataset. On Fox, if taxes were mentioned (node 5), the ad was particularly negative (2.29). If taxes were not mentioned, the ad was on average a contrast ad (nodes 8 and 9), with those mentioning a non-tax economic policy growing negative more quickly over time (node 9).

On the remaining affiliates, negative ads appeared in two subsets. Node 12 contains ads on ABC or NBC that called on the viewer to reject or defeat someone, or to send someone a message. Ads on CBS that mentioned taxes but did not mention the favored candidate until the end of the ad (31) were almost uniformly negative (node 31), while those that mentioned the favored candidate during the course of the ad were quite positive (node 30). Ads on ABC and NBC that called for either no action or a supporting action were, on average, contrast ads (node 28 and 29). Ads on CBS that did not mention taxes were also relatively positive (nodes 28 and 29), with ads that mention social policy more positive.

The results from Model 2 are in Appendix D. Issue density follows a different pattern than ad tone. The first split is on whether the ad calls on the viewer to vote against, defeat, or reject someone (to the left) or some other action (to the right). Among ads that call for voting against a candidate, ads that mention social welfare policy also mention some other issue about three fourths of the time (node 5). Ads that mention neither social welfare policy nor taxes (node 8) mention 1.4 issues, with this increasing in time. If taxes, but not social welfare policy, is mentioned (node 9), about 1.4 ads are mentioned, but the time trend is negligible. Mentioning taxes, but not social welfare policy (node 9) is stable over time.

For ads that are positive or call for no action, the next split is on health care. Those ads that do not mention health care, but mention education (node 13), about 2.4 issues are discussed, with a sizable time trend. Ads that mentioned neither health care nor education (nodes 24 and 25) mentioned about 1.5 issues. Ads that mentioned healthcare also mentioned about 1.4 additional issue areas (nodes 28, 29, 15). Interestingly, whether a website was mentioned was selected. Of ads on CBS and NBC that mentioned health care, mentioning a website resulted in a higher issue density as the campaign neared (node 29, the least squares coefficient of 5.8).

Appendix E contains the results for Model 3. In terms of relative cost of each ad, the Fox vs. ABC/NBC/CBS split is made again. Ads on Fox are less expensive (to the right of node 1), with Republicans and Libertarians spending relatively more on Fox (node 6), at a rate increasing in time. Among Democrats spending on Fox (nodes 14 and 15), those in races with high aggregate expenditures (node 15), Democratic expenditures are similar to Republicans. In races with lower aggregate expenditure (node 14), Democrats place relatively cheaper ads. The variable for aggregate expenditure, ESTSPENDsum, is selected instead of the variable for mean expenditure by race, ESTCOSTMEAN2. This suggests that Democrats are purchasing more expensive ads on Fox when the aggregate expenditure on ads in the race is high.

Ads on CBS, NBC, and ABC were relatively more expensive than those on Fox. Those that urged an action to vote against someone or send someone a message (node 4) were the most costly, relative to all ads in the race. The estimated spending on these ads, though, decreased over time. For those ads that encouraged a vote for someone, or no action, those that mentioned campaign finance and Iraq were relatively costly (nodes 11 and 21), while those that did not mention either issue were of average cost (node 20). On the major networks, attack ads were the most expensive, followed by those that focused on campaign finance and Iraq.

4 Conclusion

Generalizing from a particular election year is difficult. In many ways, 2008 was sui generis, with young voters activated in record numbers, a charismatic Democratic presidential candidate, and a vastly unpopular President with a respectable collection of bungled wars and mismanaged hurricane relief efforts to his name. The 2006 Democratic tide was continuing to roll in, with Obamas coat-tails carrying Democrats into both houses of Congress.

Yet, few of the variables selected in the models were particular to 2008. Health care and Iraq, perhaps, but the remaining selected variables will not lose their salience in future years or races. In the abstract, a reasonable picture appears. Ads on Fox that mention taxes will most likely continue to be, on average, negative. As a haven for conservatives, the salient issues on the station (taxes and economic policy) mirror the concerns of the viewership. Ads that mention taxes elsewhere will be positive if the favored candidate is mentioned, and negative otherwise. Candidates promoted themselves on social welfare policies, but attack on taxes. They spent more on the major networks, in particular on attack ads. Expenditures on Fox varied by party, with Democrats spending more in competitive races. Across models, a persistent theme was the different behavior on Fox, by either tone, issue, or party. More generally, as viewers self-select into affiliates based on political predilection, campaigns should continue to differentiate, playing to the concerns of each affiliate.

Ads that call on the viewer to vote against, defeat, or reject someone mentioned fewer issues than those that call for no or a supportive action–and campaigns purchased more expensive ads of this type. Opponents were attacked on a single issue (taxes, social welfare policy), but promoted on several. Interestingly, the web was selected as a better predictor than many other plausible variables (incumbency, race type, and so on), and this subset had a high issue density. In ads that call for no or supportive action of a candidate, a website mention will most likely continue signify an issue dense ad. Using the web as a means of information dissemination signified high issue density in 2008.

Splits in ad tone and issue density are motivated by a particular issue, taxes, and a particular issue area, social welfare policy (which subsumes health care and education). Once the decisions to air on a particular affiliate and call for a type of action have been made, social welfare policy and taxes provide the best means to cleave campaign ads in terms of issue density and tone. Mentioning campaign finance or Iraq provided a good predictor of relative expenditure, with more money for each ad being spent on either issue on the major networks.

A Description of Variables, by Category

- Dependent Variable[7]

    ISSUEDENS: Number of issue areas mentioned.*

    ADTONE: Tone of ad.

    RELCOST2: Spending on ad over the mean spending for the race.*

- Node Variable

    DaysTo: Negative days to the general election.

- Split Variables

    Campaign and Candidate Variables

    RACETYPE: House, Senate, or Governor.

    INCUMBENT: Favored candidate an incumbent.*

    ESTSPENDsum: Sum spent in that race.*

    ESTCOSTMEAN2: Mean ad cost in race.*

    TOTADSRACE: Total ads in the race.*

    PARTY: Party of candidate.

    MINWEEKS: Latest week in which that candidate placed an ad.*

    Placement Variables

    ESTSPENDING: Estimated spending on the ad.

    WeeksTo: Weeks to the election.

    MonthsTo: Month to the election.

    SPONSOR: Sponsor of spot.

    DAYPARTNUM: Part of day.

    AFFILIATE2: Affiliate.

    LENGTH: Length of ad.

    PFBYN: Does the ad say who paid for it?

    ACNYN:Does the ad direct the viewer to take any action (as opposed to merely providing information)?

    ACNTYP: What is the action?

    MAGWRD: Does the ad mention any of the following specific words or phrases: vote for, election, support, cast your ballot, [Smith] for Congress, vote again, defeat, or reject?

    PHONE: Does the ad provide a phone number?

    MAILAD:Does the ad provide a mailing address?

    WEBSITE: Does the ad provide a website address?

[7] Variables not marked with an asterisk, and their descriptions, from the unpublished Wisconsin Advertising Project Codebook 2008. The data with an asterisk were calculated by the author.

APPROVE: Where does the candidate's oral approval of the spot appear?

APVID: Does the candidate physically appear on screen and speak to the audience during authorization?

FCMNTN: Is the favored candidate mentioned in the ad?

FCAPER:Does the favor candidate appear on screen narrating his or her ad?

OPMNTN: Is the favor candidate's opponent mentioned in the ad?

PRTYMN:Does the ad mention the party label of the favored candidate or the opponent?

HUMOR: Is the ad funny or intended to be humorous?

SUPSRC: Does the ad cite supporting sources to bolster various claims?

OPAD: Is an opponent's ad mentioned or shown on screen?

ENEGMENT: Does the ad mention negative or dirty campaigning by opponents?

PERPLY: In your judgment, is the primary focus of the ad on personal characteristics of either candidate or policy matters?

LANG: What is the primary language of the ad?

FLAG: Does an American Flag appear in the ad?

PRSMENT: Is George W. Bush mentioned or pictured in the ad?

BCLINTMT: Is Bill Clinton mentioned or pictured in the ad?

CONGMT: Is (the Democratic) Congress mentioned or pictured in the ad?

DEMNOMMT: Is the Democratic nominee mentioned or pictured in the ad?

REPNOMMT: Is the Republican nominee mentioned or pictured in the ad?

ENDRSE: Does the ad include any endorsements?

NEWPAP: Endorsement by Newspaper

POLICE: Endorsement by Law Enforcement

POLS: Endorsement by Politician

UNIONEN: Endorsement by Labor Union

INTGRPS: Endorsement by Interest Group

TEACHERS: Endorsement by Teacher Group

CELEB: Endorsement by Celebrity

ORDCIT: Endorsement by Ordinary Citizen

OTHEREND: Endorsement by

OCLEB: Is a political figure or celebrity featured in the ad in a way that associates them with the opponent?

Is911: Specifically mentions September 11th

TERROR: Specifically mentions Terror

IRAQ: Specifically mentions Iraq

SURGE: Specifically mentions Surge

GOD: Specifically mentions God

HOPE: Specifically mentions Hope

CHANGE: Specifically mentions Change

EXPERIENCE: Specifically mentions Experience

*ECPOLICY***: Mentions any of the Economic Policies Below[8]

TAXES: Taxes

DEBT: Debt

GOVSPEND: Government Spending

RECESS: Recession/Stimulus

MINWAGE: Minimum Wage

FARM: Farming (Friend of)

BUSINESS: Business (Friend of)

UNION: Union (Friend of)

JOBS: Employment/Jobs

POVERTY: Poverty

TRADE: Trade/Globalization

HOUSING: Housing/Sub-prime Mortgages

*SOCPOLICY***: Mentions any of the Social Policies Below

ABORTION: Abortion

GAYS: Homosexuality/Gay & Lesbian Rights

MORAL: Moral/Family/Religious Values

TOBACCO: Tobacco

AFFIRACT: Affirmative Action

GAMBLING: Gambling

EUTH: Assisted Suicide/Euthanasia

GUNS:Gun Control

PRIVACY: Civil Liberties/Privacy

RACE: Race Relations/Civil Rights

*LAWPOLICY***: Mentions any of the Law and Order Policies Below

CRIME: Crime

DRUGS: Narcotics/Illegal Drugs

CAPPUNISH: Capital Punishment

SPCOURT: Supreme Court/Judiciary

*SWPOLICY***: Mentions any of the Social Welfare Policies Below

EDUCTION: Education/Schools

[8]Variables marked with "**" are policy areas. Variables below each comprise their specific issues.

LOTTERY: Lottery for Education

CHILDCAR: Child Care

HCARE: Health Care (not prescription drugs)

PRESDRUG: Prescription Drugs

MEDICARE: Medicare

SOCSEC: Social Security

WELFARE: Welfare

WMHEALTH: Women's Health

*DEFPOLICY***: Mentions any of the Foreign or Defense Policies Below

DEFNIRQ: Defense Military (not Iraq)

FORNIRQ: Foreign Policy (not Iraq)

VETS: Veterans

FORAID: Foreign Aid

NUKES: Nuclear Proliferation

CHINA: China

MIDEAST: Middle East

IRAN: Iran

AFGHAN:Afghanistan

*OTHPOLICY:*** Mentions any of these Other Issues Below

GLOBWARM: Global Warming

ENERGY: Energy Policy

OTHENV: Other Environmental

CAMFIN: Campaign Finance Reform

GOVETHIC: Government Ethics/Scandal

COFRAUD: Corporate Fraud

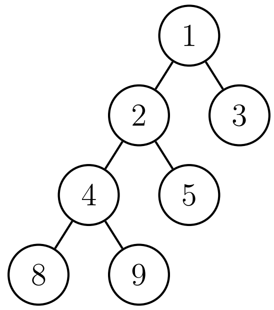
TERMLIM: Term limits

PLEDGE: Pledge of Allegiance (restrictions on use of)

KATRINA: Hurricane Katrina

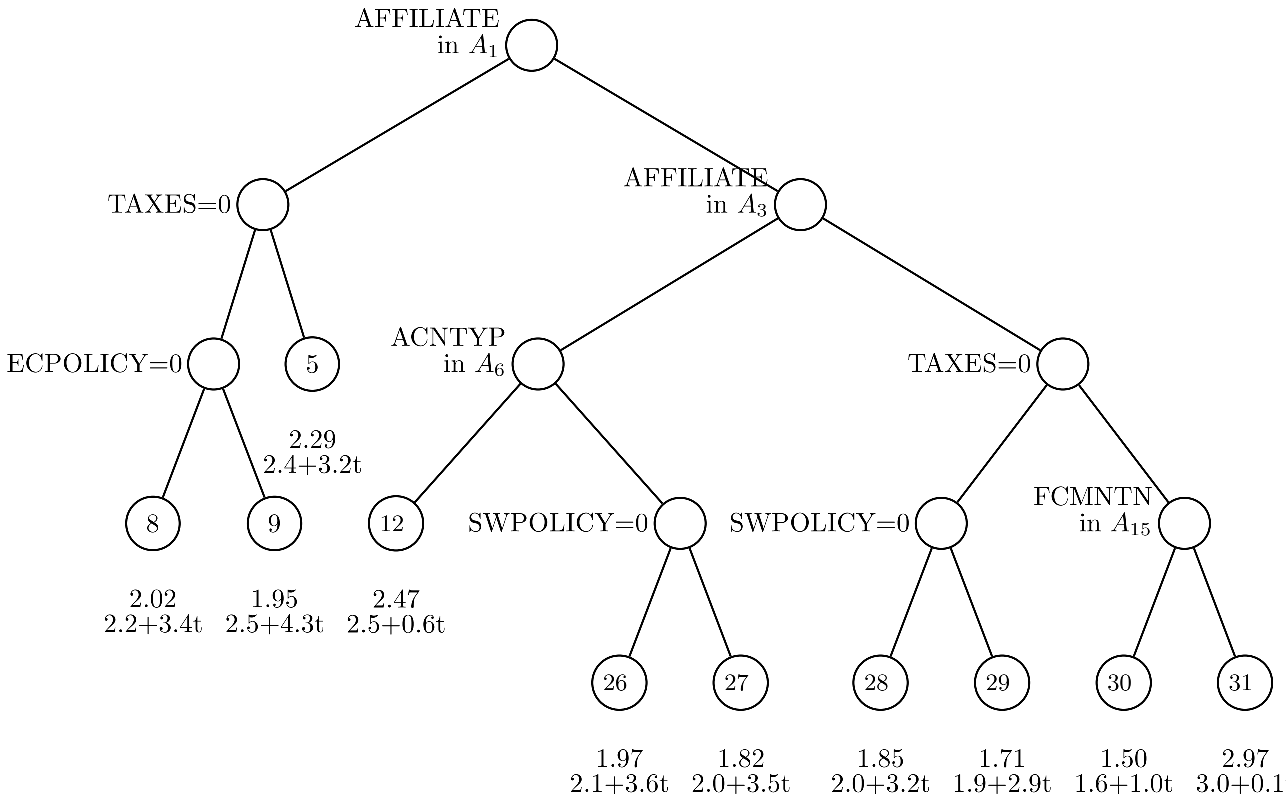
LOCALISS: Local Issues

B An Example of Regression Tree Node Numbering



An example of the numbering of regression tree nodes.

C A GUIDE Regression Tree for Campaign Ad Tone



GUIDE piecewise multiple linear least-squares regression tree model. Terminal nodes are numbered for identification. At each intermediate node, a case goes to the left child node if and only if the condition is satisfied. Beneath each leaf node are the sample mean of ADTONE and the coefficients from the least squares fit in that node. **Key**

Dependent Variable:

ADTONE = Ad Tone; 1 = Promote, 2 = Contrast, 3 = Attack

Node Variable:

t: time, negative thousandths of days until the general election. As t increases, the ad is closer to the election.

Split Variables:

ACNTYPE = Action Type

SWPOLICY, *ECPOLICY* = Social Welfare or Economic Policy mentioned

Split Sets:

A1: Go left if Affiliate is FOX, CW, INC, IND, LIFE, MYTV, ND, UPN, WB, WHAM, WV.

A1: Go right if Affiliate is ABC, CBS, NBC, FOX/ABC, SYN, TEL, TLF, UNI, UPN/WB, WTVT

A3: Go left if Affiliate is ABC or NBC.

A3: Go right if Affiliate is CBS, FOX/ABC, SYN, TEL, TLF, UNI, UPN/WB, WTVT

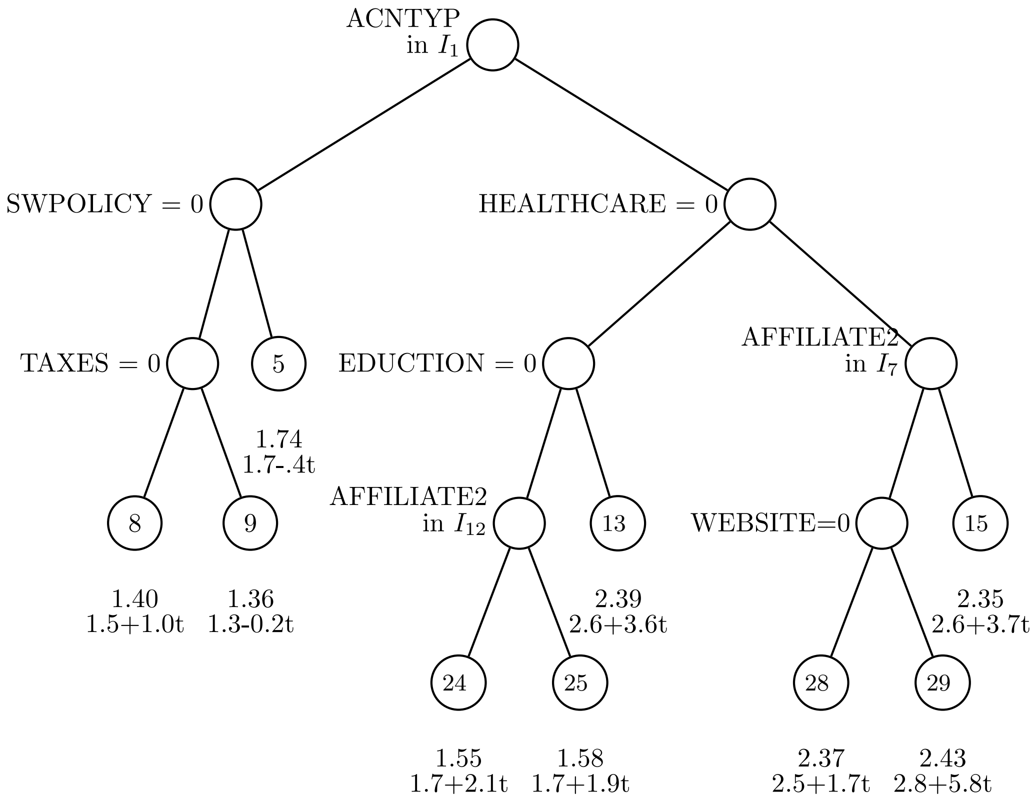
A6: Go left if Action Type is to vote against, defeat, or reject someone; to send a message to someone; or other.

A6: Go right if Action Type is no action; to vote, elect, re-elect, or support someone; or to urge action on a particular matter.

A15: Go left if the favored candidate is mentioned by name, pictured, or mentioned and pictured

A15: Go right if the favored candidate is mentioned only in approving the ad, or DK/NA.

D A GUIDE Regression Tree for Campaign Issue Density



GUIDE piecewise multiple linear least-squares regression tree model. Terminal nodes are numbered for identification. At each intermediate node, a case goes to the left child node if and only if the condition is satisfied. Beneath each leaf node are the sample mean of ISSUEDENS and the coefficients from the least squares fit in that node.

Dependent Variable:

ISSDENS = Issue Density; Number of issues mentioned, 0-6 possible. Min=0, max=5.

Node Variable:

t: time, negative thousandths of days until the general election. As t increases, the ad is closer to the election.

Split Variables:

ACNTYPE = Action Type

SWPOLICY, *ECPOLICY*, *EDUCTION* = Social Welfare, Economic, or Education Policy mentioned

WEBSITE = A website address is provided

Split Sets:

I1: Go left if Action Type is vote against, defeat, or reject someone

I1: Go right if Action type is NA; vote for, elect, or re-elect someone; urges action, send a message, or write, call, or tell someone something

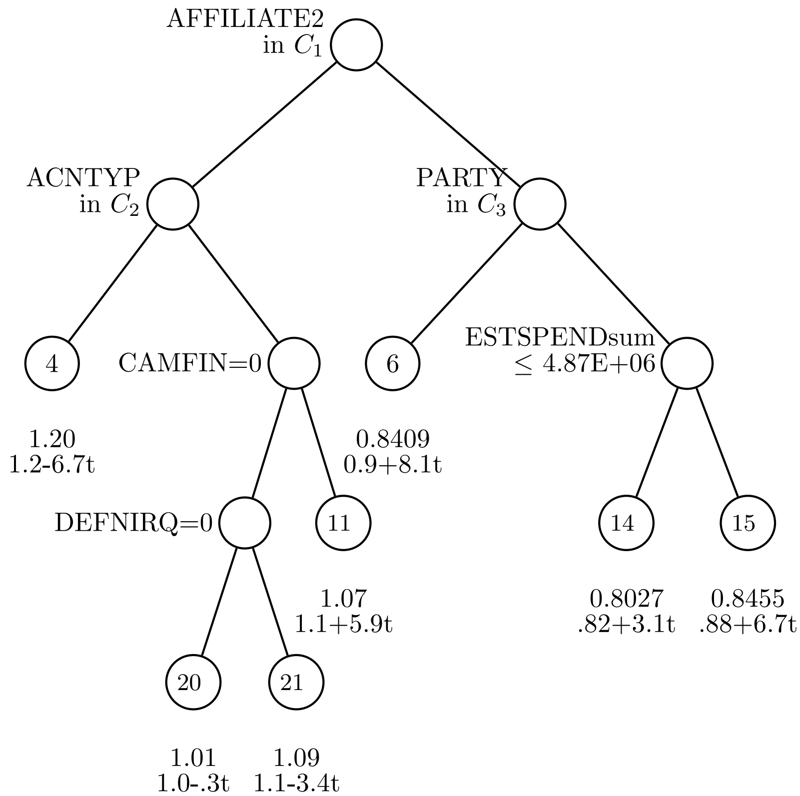
I7: Go left is Affiliate is CBS, NBC, CW, FOX/ABC, SYN, TLF, UPN, WB/UPN, WHAM, WV

I8: Go right if Affiliate is ABC, FOX, INC, IND, CBX, LIFE, ND, TEL, UNI, UPN/WB, WB, WTVT

I12: Go left if Affiliate is NBC, IND, MYTV, SYN, TEL, UNI, UPN, UPN/WB, WTVT

I12: Go right if Affiliate is ABC, CBS, FOX, CW, INC, LIFE, ND, TLF, WB, WB/UPN, WHAM

E A GUIDE Regression Tree for Ad Relative Cost



GUIDE piecewise multiple linear least-squares regression tree model. At each intermediate node, a case goes to the left child node if and only if the condition is satisfied. Beneath each leaf node are the sample mean of RELCOST2 and the name of the regressor.

Dependent Variable:

RELCOST2 = Relative cost of ad; Cost of ad divided by mean cost of ads for that race. Values above one are above the average cost of all candidates' ads in that race; values below one are of below average cost.

Node Variable:

t: time, negative ten thousandths of days until the general election. As t increases, the ad is closer to the election.

Split Variables:

ACNTYPE = Action Type

CAMNFIN, *DEFNIRQ* = Campaign finance, Iraq mentioned

PARTY = Political Party

ESTSPENDsum = Estimated spending by all candidates in that race

Split Sets:

C1: Go left if Affiliate is ABC, CBS, NBC, FOX/ABC, INC, LIFE, TEL, UNI, WB, WHAM, WV

C1: Go right if Affiliate is FOX, MYTV, ND, CW, IND, SYN, TLF, UPN, UPN/WB, WV

C2: Go left if Action Type is to vote against someone, defeat someone, tell someone something, or send a message to someone.

C2: Go right if Action Type is no action, to urge support of someone or to vote for someone, to urge action on a particular matter, join an organization, contribute money, or other

C3: Go left if Party is Republican, Libertarian, or other

C3: Go right if Party is Democrat, Independent, Green, Libertarian, or missing

References

Ansolabehere, Stephen D., Shanto Iyengar & Adam Simon. 1999. "Replicating Experiments Using Aggregate and Survey Data: The Case of Negative Advertising and Turnout." *The American Political Science Review* 93(4):901–909.

Ansolabehere, Stephen & Shanto Iyengar. 1997. *Going Negative*. Free Press.

Ansolabehere, Stephen, Shanto Iyengar, Adam Simon & Nicholas Valentino. 1994. "Does Attack Advertising Demobilize the Electorate?" *The American Political Science Review* 88(4):829–838.

Baumgartner, Frank & Bryan Jones. 1993. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Breiman, L., J. H. Friedman R. A. Olshen & C. J. Stone. 1984. *Classification and Regression Trees*. Belmont: Wadsworth.

Coleman, John J. & Paul F. Manna. 2000. "Congressional Campaign Spending and the Quality of Democracy." *Journal of Politics* 62:757–789.

Downs, Anthony. 1957. *An economic theory of democracy*. HarperCollins, New York.

Goldstein, Kenneth & Travis Ridout. 2004. "Measuring the effects of televised political advertising in the United States." *Annual Review of Political Science* 7.

Grimmer, Justin & Gary King. 2009. "Quantitative Discovery from Qualitative Information: A General-Purpose Document Clustering Methodology.".
*http://gking.harvard.edu/files/discov.pdf

Hirschman, Albert O. 1970. *Exit, voice, and loyalty : responses to decline in firms, organizations, and states*. Cambridge, Mass. : Harvard University Press.

Imai, Kosuke & Aaron Strauss. 2009. ""Planning the Optimal Get-out-the-vote Campaign Using Randomized Field Experiments".".
*http://imai.princeton.edu/research/choice.html

Kritzer, Herbert M. 1996. “The Data Puzzle: The Nature of Interpretation in Quantitative Research.” *American Journal of Political Science* 40(1):1–32.

Loh, W.-Y. 2008*a*. *Classification and Regression Tree Methods*. Chichester, UK: Wiley pp. 315–323.

Loh, W.-Y. 2008*b*. “Regression by parts: Fitting visually interpretable models with GUIDE.” pp. 447–469.

Martin, Andrew D., Kevin M. Quinn Theodore W. Ruger & Pauline T. Kim. 2004. “Competing Approaches to Predicting Supreme Court Decision Making.” *Perspectives on Politics* 2(4):761–767.

Niemi, Richard G., Richard S. Katz & David Newman. 1980. “Reconstructing Past Partisanship: The Failure of the Party Identification Recall Questions.” *American Journal of Political Science* 24(4):633–651.

Petrocik, John R. 1996. “Issue Ownership in Presidential Elections, with a 1980 Case Study.” *American Journal of Political Science* 40(3):825–850.

Riker, William. 1986. *The Art of Political Manipulation*. New Haven: Yale University Press.