PREDICTIONS OF THE BUSH-CLINTON-PEROT PRESIDENTIAL RACE FROM THE PRESS

David P. Fan, University of Minnesota
Dept. Genetics and Cell Biology, 1445 Gortner Ave., St. Paul, MN 55108

KEY WORDS: Election, mass media, methodology

In recent years, forecasters of American presidential elections have used regression methods to make quantitative estimates of the percentage of the vote for the two major party candidates on election day. The explanatory variables have included economic growth and July presidential approval (Abramowitz, 1988; Campbell, 1993; Fair, 1988; Lewis-Beck and Rice, 1992). The models have had varying degrees of success for the 1992 election (Campbell and Mann, 1992; Greene, 1993; Campbell, 1993). Another forecasting method has been the Iowa Political Stock Market (Forsythe, Nelson, Neumann and Wright, 1992) in which where individuals trade futures contracts representing the presidential candidates. As with the regression models, the goal has been the electoral outcome. In addition, the futures method generates a time trend of likely vote shares on election day throughout the time period before polling day.

Instead of aiming just for the election result, this paper models the entire series of opinion polls preceding the election. The election then provides just another measure of public opinion and constitutes one additional measure of the dependent variable, candidate preference. The basic tenet of this paper is that the mass media form the only information conduit which is rapid and extensive enough to transmit the breaking news which can affect the election, whether that information is about the economy or other topics. Although there are both paid and unpaid forms of the print and electronic media, print news is archived on electronic databases and is hence easy to retrieve and analyze. Therefore, the experimental question was the extent to which the print press could capture the salient messages influencing the 1992 election.

Data and content analysis

The basic method (Fan, 1988; Fan and Tims, 1989) was to score press coverage prior to the election for the numbers of paragraphs favorable and unfavorable to the three main candidates, George Bush, Bill Clinton and Ross Perot. Then the scores were entered into the mathematical model of ideodynamics to compute a media share time trend. This trend was computed assuming that candidate preference was driven entirely by press information. The model was tested by comparison with actual survey points.

Poll data on the election came from the POLL electronic database archived at the Roper Center, University of Connecticut. Since the goal was to find public opinion which corresponded best to electoral results, a time series was constructed from 61 polls asking about the chances of voting for Bush, Clinton or Perot in a three way race. For comparability to the actual vote, responses were restricted to those including people leaning to the candidates. After the vice-presidential running mates were selected, the questions also included their names together with those for the presidential candidates. Questions were excluded if they posed conditions such as Ross Perot being unlikely to win. The undecideds and don't knows constituted 13 percent or less of the total in all but two of the 61 polls and were therefore subtracted before renormalization of the remaining numbers to 100 percent.

Since the first polls comparing all three candidates including leaners were reported toward the end of March, 1992, mass media messages were sought from March 15, 1992 to election day. To focus on information relevant to the election, the Major Paper library of the NEXIS electronic database was searched for stories containing at least two of the last names Bush, Clinton and Perot. At the time of the study, this library archived stories from 16 major newspapers around the United States. A retrieval was made of the texts from 3394 of the 24,213 stories identified by the search from March 15, 1992 to election day, November 3, 1992.

The 13 million characters of retrieved text were scored by computer (Fan, 1988; Fan and Tims, 1989; see the author for access to the software). The computer began by selecting just those paragraphs mentioning at least one of the words Bush, Clinton or Perot. Then the computer scored these paragraphs in two dimensions. In one, a count was made of the numbers of paragraphs favorable and unfavorable to each of the three candidates. In the other, enumerations were made of paragraphs both mentioning Bush, Clinton and/or Perot and discussing individual campaign issues (see Table 1 for listing).

The result of the content analysis was scores for the numbers of paragraphs on each day favoring each of the chosen ideas. These scores were visualized as persuasive force functions (Fan, 1988, Appendix A) in which all paragraphs were given their scored value on the date of the story. On each succeeding day, that score decreased by half corresponding to consistent findings of exponential drops in a story's persuasive ability with a one day half-life (Fan, 1988).

Table 1: Subtopics of campaign issues

Issue	Examples of subtopics
Social Policy	Lifestyle, religion, family, social, values, abortion, crime, guns, flag burning
Education	Education, schools, colleges, universities, students, teachers
Character	Trust, character, scandal, infidelity, flip flop, sleaze, slickness
Foreign	Foreign countries and regions, crisis management, Iran Contra
Military	Military, arms, defense spending, bombers, submarines
Health	Health, medical, AIDS
Environment	Environment, clean air, acid rain, ozone, nuclear power, global warming
Relatives	Relatives, wives, children, etc. by name
Miscellaneous	Immigration, MIAs, reform of any type
Disadvantaged groups	Economic disadvantages of urban, inner city areas, homeless, the poor, welfare, ethnic, racial, gay, lesbian
Other subgroups	Women, middle class, wealthy, farmers, labor, elderly
Abstract ideas	Ideological like conservative or liberal or radical, insider, outsider, change, status quo
Economy	Budget deficit, business, financial, taxation, spending, saving, trade unemployment, regulation

Scores for all paragraphs on good and bad news were added together to give the plots in Fig. 1. The Bush paragraphs, both pro and con, remained low from March 15 through the beginning of the Democratic convention in July. After mid-April, when effective opposition to Clinton had ended, the major spikes in coverage corresponded to news about Clinton during the Democratic convention in July, brief discussion of Perot following his announcement of his withdrawal in the same month, and coverage of Bush during the Republican convention in August. After that time, there was a steady crescendo of news interest through to the election. Perot was obviously little covered between his withdrawal and reentry in September.

A comparison of favorable to unfavorable news, averaged through the campaign (Table 2, bottom line),

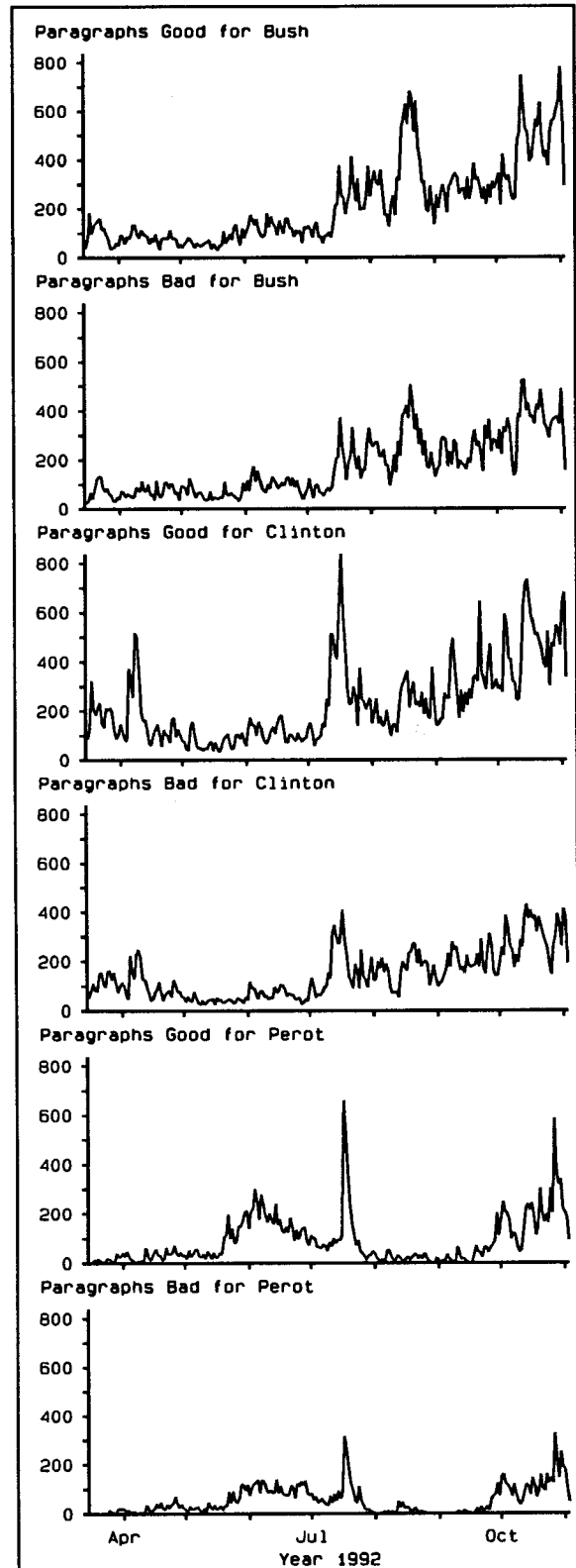shows that the positive messages were approximately

Figure 1: Paragraphs favorable and unfavorable to

Bush, Clinton and Perot.
the same for Bush and Clinton while Bush was confronted with significantly more negative news. For Perot, there was much less total news, and among that, most was favorable.

Predictions of candidate preference from the press

Candidate preference was forecast using a variant of the ideodynamic equations already described (Fan, 1988; Fan and Tims, 1989; Appendix A). This model extends the arguments used to derive the logistic function which describes the diffusion of innovations (Fan, 1988; 1993). The ideodynamic strategy is to compute a persuasive information share, in this case a media share, which is expected for theoretical reasons to correspond to public opinion. Then this media share, calculated only from media input, is compared to the results of opinion polls. This approach has the advantage that a media share can be computed every day to follow late breaking news regardless of whether a poll is taken. Also, computations can even be made during times when no polls are available as was the case when Perot was not a candidate.

The calculations began on April 24 after Bush, Clinton and Perot had all been clearly identified as the likely contenders in the general election. The media shares on this date were assigned to be the candidate preference percentages found by opinion polling. From this time forward, a media share was computed every 24 hours for each of the three candidates assuming that changes in share were driven by the persuasive forces shown in Fig. 1. The net persuasive force favoring a candidate was the favorable curve minus the unfavorable one for that candidate (Fig. 1) with the two curves being allowed to have different weights (Appendix A).

The result was a model with two parameters to estimate: the relative weight w of favorable to unfavorable information, and the persuasibility constant k describing the ability of messages represented in the news stories to persuade the public. Higher k values correspond to larger media share movement for a typical newspaper paragraph. These parameters were estimated by least squares minimization of a function which accounted for autocorrelation in the errors at different opinion polling dates.

The optimized k value was 0.0088 poll percent per paragraph per day with 95 percent confidence interval (0.0056,0.016). Value w corresponded to negative news having a weight 0.69 that of positive news with 95 percent confidence interval (0.50,0.82). The Root Mean Squared Deviation (RMSD) between the media share line and opinion poll values was 4.6 percent. The RMSD is one measure of forecasting power (Pindyck and Rubinfeld, 1981). The R^2 values was -0.03 for the

time trend for Bush, 0.83 for Clinton, and 0.70 for Perot. The negative value for Bush meant that the ideodynamic media share was slightly worse as a predictor of Bush opinion than simply taking the average of the Bush poll values. This was not surprising since the Bush numbers moved very little throughout the campaign making it difficult to improve on the average. The small loss in fit for the Bush trend improved those for Clinton and Perot.
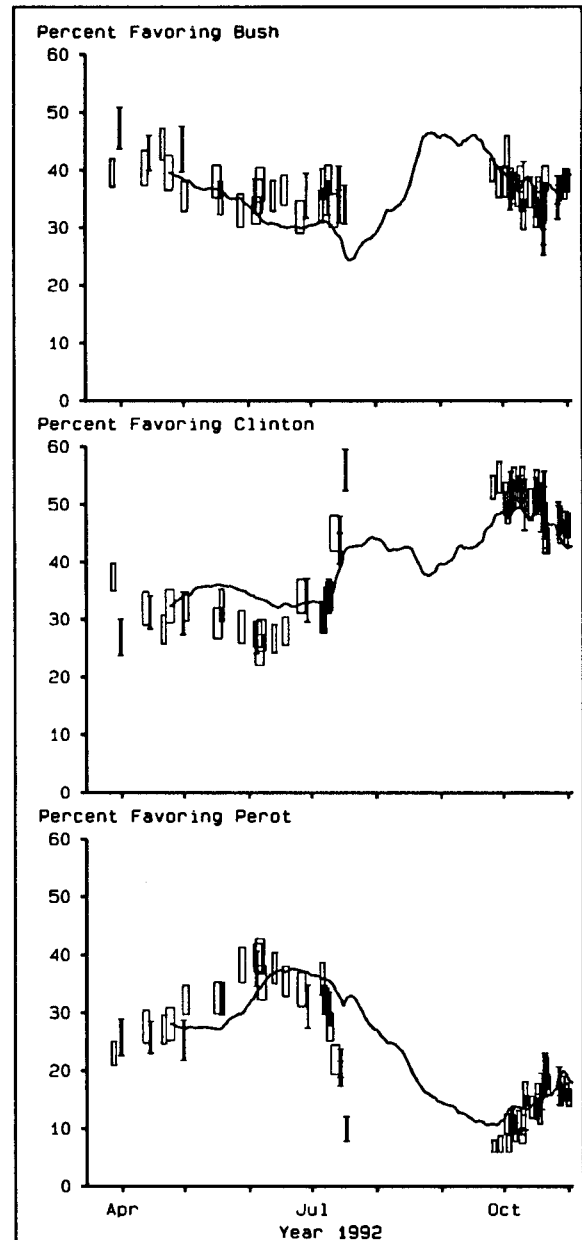

Figure 2: Opinion favoring Bush, Clinton and Perot. Line is media share computed from scores in Fig. 1; boxes are poll values (box left and right give poll dates; top and bottom give 95 percent confidence).

Graphical comparisons between the media share lines and actual poll data are shown in Fig. 2. Soon after Perot's withdrawal, there were no polls comparing the three candidates to each other. Nevertheless, the ideodynamic computations were continued through this time until Perot rejoined the race when, quite significantly, the computed values were again close to the poll numbers -- without changing the content analysis, the k value, weight w, or any other feature of the computation. In the interest of model parsimony, no corrections were made to the information structure following Perot's withdrawal so the divergence between the media share and poll trends just after Perot left the race in July was expected. On election day, the ideodynamic computations gave 39 percent to Bush, 43 percent to Clinton, and 18 percent to Perot. The corresponding percentages from the election were 38, 43 and 19 percent, only one percent different.

Besides showing high R^2 values and low RMSD values, there is another method to demonstrate that media share is a good predictor of opinion. This procedure is to perform utility regressions (Appendix A) to see if the media share, including the value at the present time, could reduce the error in opinion predictions based solely on past opinion measurements. The results (Eq. 6-8, Fig. 3) show that the media share gave significantly more information about current opinion than simply the past history of the poll series (F significance 0.048, 0.000021, and 0.0086 for opinion for Bush, Clinton, and Perot, respectively).

Having shown that the media share gave good forecasts of opinion, Granger causality tests were performed to determine if the media was driving or reflecting opinion or both. The results (Eq. 2-5, Fig. 3) showed that the media share was a Granger cause of opinion for both Clinton (F significance 0.033) and Perot (F significance 0.007). The reverse regression was also significant for opinion affecting the media share for Clinton (F significance 0.0016) and Perot (F significance 0.000026). With Bush opinion changing very little, causality was neither expected nor observed in either direction.

Besides its usefulness for predicting opinion, the media analysis in this paper also provides information about the types of messages sent to the public about the election. As an example, consider scores of all paragraphs both mentioning at least one of the candidates and discussing the individual campaign issues listed in Table 1. The data (Table 2) show, as expected, that both Bush and Perot had more favorable than unfavorable paragraphs in the context of character. The reverse was true for Clinton. Clinton was more discussed than the other two in the context of both disadvantaged subgroups and other subgroups like the

middle class. Therefore, Clinton succeeded in appealing to all segments of American society. While Bush had as much unfavorable as favorable discussion in the context of the economy, Clinton's coverage was significantly better for this issue.

Discussion

This paper is consistent with findings for the Bush-Dukakis race of 1988 that the press could be used to forecast candidate preference and electoral results (Fan and Tims, 1989). As expected, Granger methods showed the media having a causal effect on opinion and vice versa. Since the media share both matched opinion and was computed on the basis that opinion depended on the information presented in the press, it was not surprising to find the media being a Granger cause of candidate preference. The reverse finding of opinion affecting the media is quite consistent with the press reporting extensively on poll results.

Appendix A: Ideodynamic Modeling

Ideodynamic equations (see Fan, 1988 for rationale) were written for Bush, Clinton and Perot. Since the functions for all three candidates have the same form, only the one for Bush will be given here:

$$I_{B,t} = I_{B,t-1} + k(I_{C,t-1} + I_{P,t-1})F_{B,t} \qquad (1)$$
$$- kI_{B,t-1}(F_{C,t} + F_{P,t}) + \epsilon_t.$$

In this equation, I denotes the information or media share with subscripts B, C and P referring to Bush, Clinton and Perot, respectively. Subscript t is for time. Thus $I_{B,t}$ refers to the Bush media share. F, with the same subscripts B, C, and P refer to the persuasive force functions driving changes in the information share. $F_{B,t}$ describing net favorable Bush news is obtained by subtracting the bad news persuasive force function $F_{ConBush,t}$ (second curve in Fig. 1) multiplied, by constant weight w, from the good news function $F_{ProBush,t}$ (top curve) so $F_{B,t} = F_{ProBush,t} - w(F_{ConBush,t})$. Equivalent functions are constructed for Clinton and Perot.

Eq. 1 has the same form as Eq. 1 in Fan (1993) so its further statistical treatment is exactly analogous to that in Appendix A of that paper and will not be described here. Constants k and w, the only two parameters of the system, were estimated using standard nonlinear statistics. Significance for each constant is reported as the confidence interval computed at the optimal value for the other constant.

Since the media share is calculated using only media variables, standard Granger regressions (1980) were performed to see if the media share was a cause of opinion change (see Fan, 1993). Also utility regressions were performed including media share at 0 lags to show that the media was a significant predictor of public opinion (see Fan, 1993).

Table 2: Mentions of issues in paragraphs favorable and unfavorable to Bush, Clinton and Perot as well as all paragraphs mentioning at least one of these candidates.

Issue	Bush Paragraphs		Clinton Paragraphs		Perot Paragraphs		All Paragraphs
	Pro	Con	Pro	Con	Pro	Con	
Social Policy	2129.58	1524.23	1605.52	1151.82	527.74	292.61	15194.83
Education	943.57	751.69	1656.82	853.80	385.57	220.83	10265.37
Character	2370.88	1978.74	2562.87	2923.14	712.06	606.93	16537.04
Foreign	2868.18	2203.16	1787.77	1293.57	350.21	283.80	16885.37
Military	820.92	581.83	588.72	477.46	165.41	127.85	5760.43
Health	572.88	718.09	704.73	578.16	119.25	114.20	5917.04
Environment	506.16	454.85	477.49	414.88	60.00	15.50	3309.56
Relatives	1040.78	743.29	1092.83	935.18	267.90	256.86	9995.66
Miscellaneous	399.30	266.02	317.60	166.95	169.85	104.90	3049.42
Disadvantaged Groups	1163.46	1007.01	1857.55	1225.50	474.41	356.36	12014.08
Other Subgroups	1460.90	1327.73	2322.94	1162.78	414.63	294.26	14096.18
Abstract Ideas	2768.91	1916.05	2351.81	1499.15	823.94	565.10	17962.48
Economy	6061.18	5902.27	5692.39	4773.27	2050.20	1353.53	50971.03
All Paragraphs	26293.58	18404.82	26893.95	14879.68	11007.99	5885.84	

Granger causality: Media share i on opinion y for Clinton and Perot (insignificant for Bush)

$$y_{C,t} = 13.9 \quad +0.61y_{C,t-1} \quad +0.50y_{C,t-2} \quad +0.40y_{C,t-3} \quad -0.39i_{C,t-1} \quad -1.07i_{C,t-2} \quad +0.63i_{C,t-3} \tag{2}$$
$$(3.3) \quad (3.6) \quad (2.6) \quad (2.2) \quad (-0.64) \quad (-1.3) \quad (1.4)$$
$$F(2,46) = 3.2 \ (P < 0.033) \quad \bar{R}^2 = 0.90 \quad Q(21) = 23.9 \ (P < 0.30)$$

$$y_{P,t} = 2.0 \quad +0.76y_{P,t-1} \quad +0.37y_{P,t-2} \quad +0.15y_{P,t-3} \quad -0.27i_{P,t-1} \quad -0.24i_{P,t-2} \quad +0.13i_{P,t-2} \tag{3}$$
$$(2.0) \quad (5.6) \quad (2.2) \quad (0.92) \quad (-1.8) \quad (-1.3) \quad (1.0)$$
$$F(2,46) = 4.5 \ (P < 0.007) \quad \bar{R}^2 = 0.93 \quad Q(21) = 35.2 \ (P < 0.03)$$

Reverse Granger causality: Opinion y on media share i for Clinton and Perot (insignificant for Bush)

$$i_{C,t} = 5.2 \quad +0.86i_{C,t-1} \quad -0.29i_{C,t-2} \quad +0.05i_{C,t-3} \quad +0.13y_{C,t-1} \quad +0.01y_{C,t-2} \quad +0.11y_{C,t-3} \tag{4}$$
$$(4.2) \quad (4.9) \quad (-1.2) \quad (0.4) \quad (2.6) \quad (0.2) \quad (2.1)$$
$$F(2,46) = 8.3 \ (P < 0.00016) \quad \bar{R}^2 = 0.98 \quad Q(21) = 18.7 \ (P < 0.61)$$

$$i_{P,t} = 0.26 \quad +0.61i_{P,t-1} \quad -0.07i_{P,t-2} \quad +0.11i_{P,t-3} \quad +0.42y_{P,t-1} \quad -0.11y_{P,t-2} \quad +0.02y_{P,t-3} \tag{5}$$
$$(0.3) \quad (4.2) \quad (-0.38) \quad (0.90) \quad (3.3) \quad (-0.64) \quad (0.15)$$
$$F(2,46) = 10.3 \ (P < 0.000026) \quad \bar{R}^2 = 0.94 \quad Q(18) = 10.0 \ (P < 0.98)$$

Utility regression: Media share i improving predictions in opinion y for Bush, Clinton and Perot

$$y_{B,t} = 18.7 \quad +0.34y_{B,t-1} \quad +0.39i_{B,0} \quad -0.26i_{B,t-1} \tag{6}$$
$$(3.6) \quad (2.6) \quad (2.5) \quad (-1.6)$$
$$F(2,51) = 3.2 \ (P < 0.047) \quad \bar{R}^2 = 0.17 \quad Q(21) = 27.7 \ (P < 0.15)$$

$$y_{C,t} = 0.56 \quad +0.42y_{C,t-1} \quad +0.41y_{C,t-2} \quad +2.1i_{C,0} \quad -2.4i_{C,t-1} \quad +0.45i_{C,t-2} \tag{7}$$
$$(0.2) \quad (2.8) \quad (2.9) \quad (5.1) \quad (-4.1) \quad (1.2)$$
$$F(3,48) = 10.5 \ (P < 0.000021) \quad \bar{R}^2 = 0.93 \quad Q(21) = 14.0 \ (P < 0.87)$$

$$y_{P,t} = 1.9 \quad +0.71y_{P,t-1} \quad +0.48y_{P,t-2} \quad +0.17i_{P,0} \quad -0.40i_{P,t-1} \quad -0.037i_{P,t-2} \tag{8}$$
$$(1.9) \quad (4.7) \quad (2.9) \quad (1.0) \quad (-2.1) \quad (-0.3)$$
$$F(3,48) = 4.4 \ (P < 0.0086) \quad \bar{R}^2 = 0.93 \quad Q(21) = 19.8 \ (P < 0.53)$$

Figure 3: Regressions relating the press and candidate preference. Media share i is I in Eq. 1 less error term ϵ; polled opinion is y; the same subscripts apply as in Eq. 1; the t statistics are in (); the F test results are for all lags of the independent variable combined.

References

Abramowitz, A.I. 1988. PS 21:843-847.

Campbell, J.E. 1993. PS 26:165-166.

Campbell, J.E. and T.E. Mann. 1992. Brookings Rev. Fall:22-27.

Cook, R.D., and S. Weisberg. 1990. J. Amer. Statis. Assn. 85, 544-551.

Fair, R.C. 1988 Polit. Behav. 10:168-179.

Fan, D.P. 1988. Predictions of public opinion from the mass media. Westport, CT: Greenwood Press.

Fan, D.P. 1993. Predictions of Consumer Confidence/Sentiment from the press. Elsewhere in this volume.

Fan, D.P. and A.R. Tims. 1989. Internat. J. Public Opin. Res. 1:151-163.

Forsythe, R., F. Nelson, G.R. Neumann, and J. Wright. 1992. Amer. Econ. Rev. 82:1142-1161.

Granger, C.W.J. 1980. J. Econ. Dynam. Control 2:329-352.

Greene, J.P. 1993. PS 26:17-21.

Lewis-Beck, M.S., and T.W. Rice 1992. Forecasting Elections. Washington, DC: Congressional Quarterly Press.

Pyndyck, R.S., and D.L. Rubinfeld. 1981. Econometric models and economic forecasts, 2nd ed. New York: McGraw-Hill.