Week 5 – British Forecast Model
Lebo and Norpoth Paper

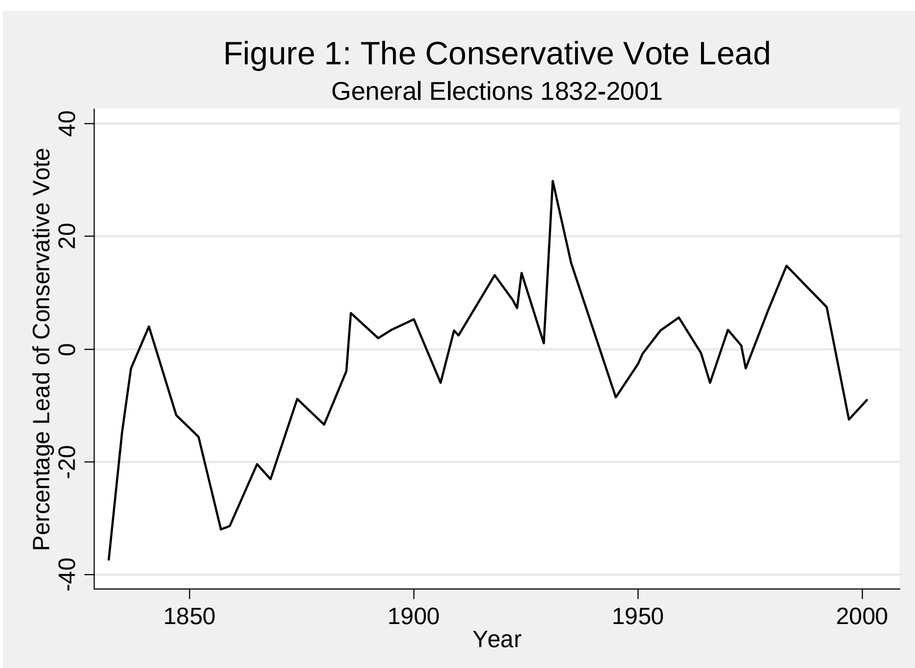
Steps of an election forecast.

First, choose an appropriate dependent variable.
The success of third parties can have a big effect on the ability to forecast so this needs to be considered.
Also, looking at the vote share of incumbents rather than of a single party changes the nature of the forecast considerably. By looking at the incumbent, the party whose vote share is measured will change over multiple elections.

Our choice is the Conservative vote lead over Labour - votcold. This is important since the outcome of post-war British elections has been much more defined by the Conservative-Labour lead than by the percentage of the vote won by either party. In some cases, the Liberal party has been very close to Labour in their vote share but this has never translated into very many seats.

Second, just as in ARIMA modeling, look for the univariate characteristics of the variable. In Helmut's work including this paper, the AR2 model captures the cyclical nature of elections.

If the AR1 parameter is positive and the AR2 is negative, a cycle is present. Looking at the figure below, the emergence of this cycle becomes very evident around the 1929 election.



In Stata, use the menu. Statistics → Time Series → Arima models.

Choose you dependent variable and the number of AR and MA parameters. Also choose the date range.

*Be sure that Stata knows your data are time series using the tsset command.

Differencing is a bad idea if elections are the unit of analysis. You will lose one observation and you will lose ability to capture long-term trends.

Third, find appropriate independent variables.

Common ones are economic variables. We use Prime Ministerial satisfaction.

What criteria should be used for a good independent variable?

Certainly statistical significance and theory. But also, the ability to collect the independent variable at a period sufficiently before the election.

So, last minute government approval ratings or vote intentions may work very well but they don't give you time to make a forecast and write a paper and ultimately don't impress many people.

In the 2004 U.S. presidential election, John Zogby waited until 4 pm on election day to finally post his forecast of the election. He said Kerry would win.

So, a long lead time is part of what can make a model particularly impressive and this is where Helmut's have been so successful.

Can start by graphing the residuals of the ARIMA model against potential independent variables.

This may alert you to problems in the construction of the independent variable. For example, PMsat by itself doesn't work because sometimes the PM is Labour and sometimes Conservative.

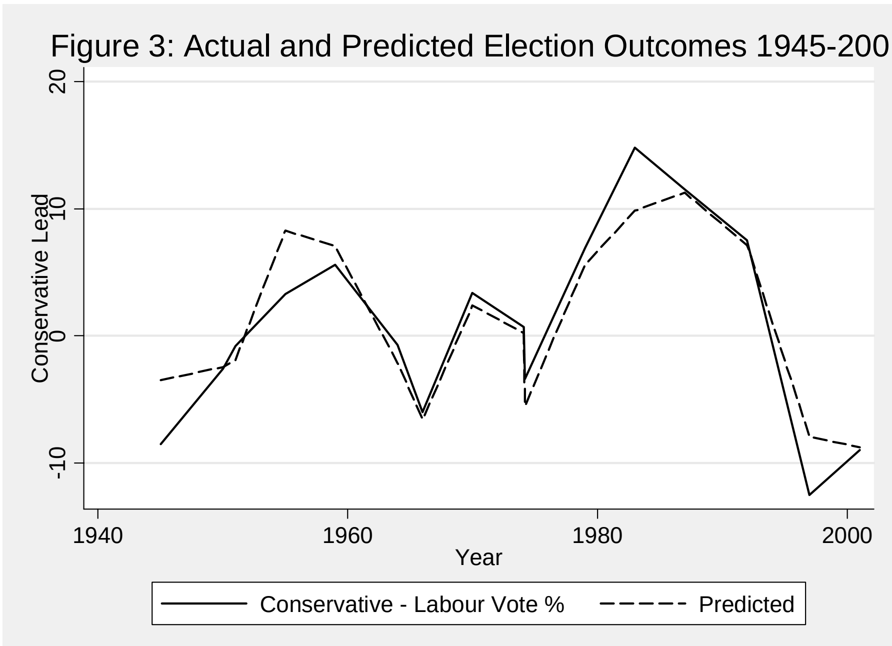
PMinv shows the diminishing marginal returns of having a very unpopular or popular PM. This lead to the choice of truncating the series and adjusting two values, giving us our independent variable, pminvt.

* Actually, the really painstaking part of this project was figuring out what polls we would use to determine PM approval for each of the post-war elections. Polls are not asked with regularity going back that far. So we came up with a set of decision rules that included using polls from 2 and 3 months before the election whenever possible and dividing the PM approval value by the share of the 2-party vote intention. This latter aspect means that, all else being equal, a given value for PM approval is *better* as the 2 party vote share goes down. For example, if PM approval is 40% and 70% of people will vote either Labour or Conservative, the PM has the support of well over half of the two-

party voters. If the two-party vote increases to 90% and PM approval stays the same, now the PM is in trouble.

Fourth, try the ARIMA model with the independent variable. Also, use robust standard errors.

We can look at our predicted and actual election results.



Looks pretty good!

Another way to judge is to see if the model gets the election winner correct.

```
. predict votehat
(option xb assumed; predicted values)
(27 missing values generated)

. list year votehat votcold pminvt if year >= 1945

     +-------------------------------------------+
     |   year     votehat    votcold      pminvt |
     |-------------------------------------------|
 28. |   1945   -3.506169       -8.5        -6.5 |
 29. |   1950   -2.495117       -2.6   -1.162788 |
 30. |   1951     -2.0729   -.799999   -6.497177 |
 31. |   1955    8.306355        3.3     6.52174 |
 32. |   1959    7.103558        5.6          15 |
     |-------------------------------------------|
 33. |   1964   -2.244889   -.699997   -.5494537 |
 34. |   1966   -6.660962         -6         -15 |
 35. |   1970    2.434148        3.4    -1.41243 |
 36. | 1974.1    .2941115    .700001   -2.046036 |
 37. | 1974.2   -5.453937       -3.4    -7.43243 |
     |-------------------------------------------|
 38. |   1979     5.66253        6.9    8.857143 |
 39. |   1983    9.072622       14.8    13.37048 |
 40. |   1987    11.09518       11.5    10.47198 |
```

```
 41. |   1992    7.497206        7.5    11.93549 |
 42. |   1997   -7.857802      -12.5   -8.181816 |
     |-------------------------------------------|
 43. |   2001   -9.319477         -9    -6.52174 |
 44. |   2005   -1.970205       -2.9    -5.68022 |
     +-------------------------------------------+
```

We are checking that we never make an error of having the wrong sign. Of course, this is arbitrary and silly by Gauss-Markov assumptions. We are essentially deciding that the sign of y-hat is more important that the size of the error. However, correctly predicting 16 of 16 elections sounds impressive.

Fifth, make some out-of-sample forecasts.

These can be done for the next sample point or for some in the series.

In Stata, we fill in the values of the independent variable that we have and leave blank the value of the observation we want to predict.

When we then use “predict” we get a value for the observation we skipped.

Using this method, we can construct “one-step-ahead” forecasts. Using only the pre-1979 data to predict 1979 and then the pre-1983 data to predict 1983 and so on. Excellent way of showing the abilities of the model. As the sample size increases, we may see our predictions getting better.

Our sixth step was to translate the votes into seats.

Again, we think of the lead the Conservatives have over Labour, here concentrating on seatcold.

An AR1 model works fine here and we add the vote percentage as a predictor.

Table 2: Actual and Predicted Conservative Seat Lead, 1945-2005

Year	Actual Seat Lead	Predicted Seat Lead	Error
1945	-180	-120.2	-59.8
1950	-16	-83.5	67.5
1951	26	-2.7	28.7
1955	68	72.5	-4.5
1959	107	91.8	15.2
1964	-13	-6.5	-6.5
1966	-110	-105.5	-4.5
1970	42	34.4	7.6
1974 F	-4	-6.6	2.6
1974 O	-42	-74.1	32.1
1979	70	109.8	-39.8
1983	188	196.9	-8.9
1987	147	138.9	8.1

1992	65	83.8	-18.8
1997	-254	-247.7	-6.3
2001	-247	-193.0	-54.0

The last step is the Monte Carlo simulation.

Each of the parameter estimates from the two steps of the forecasting model was used as a probability distribution from which draw random variables.

A Stata do-file ran this very quickly.

```
set obs 100000
set seed 61967
gen C1 = .5465+.7927*invnorm(uniform())
gen AR1 = .7585+.1271*invnorm(uniform())
replace AR1 = 1 if AR1 > 1
replace AR1 = -1 if AR1 < -1
gen AR2 = -.6797+.2102*invnorm(uniform())
replace AR2 = 1 if AR2 > 1
replace AR2 = -1 if AR2 < -1
gen PMC = .6211+.0608*invnorm(uniform())
gen PM = -3.42
gen vote = (1 - AR1 - AR2)*C1 + AR1*(-9) + AR2*(-12.5) + PMC*PM -
AR1*PMC*(-6.5) - AR2*PMC*(-9.3)
gen votecoef = 15.8916+.8879*invnorm(uniform())
gen C2 = -34.269+28.2725*invnorm(uniform())
gen seatAR = .7437+.1951*invnorm(uniform())
replace seatAR = 1 if seatAR > 1
replace seatAR = -1 if seatAR < -1
gen seats = (1 - seatAR)*C2 + seatAR*(-247 - (votecoef*-9)) +
votecoef*vote
```

This is just another way of dealing with the question of forecast error. We shouldn't think about a single forecast but rather look at it as the mean of a probability distribution.

Figure 3: Probabilities for 2005 Seat Distribution

Based on 100000 Simulations