Electoral Studies xxx (2015) 1–5

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

# Forecasting the 2015 British general election: The Seats-Votes model

Paul Whiteley [a, *], Harold D. Clarke [a, b], David Sanders [a], Marianne C. Stewart [b]

[a] *Department of Government, University of Essex, UK*
[b] *School of Economic, Political and Policy Sciences, University of Texas at Dallas, USA*

A R T I C L E I N F O

*Article history:*
Available online xxx

*Keywords:*
Cube rule
Seat forecasts
ARIMA time series models

A B S T R A C T

This paper applies the Seats-Votes Model to the task of forecasting the outcome of the 2015 election in Britain in terms of the seats won by the three major parties. The model derives originally from the 'Law of Cubic Proportions' the first formal statistical election forecasting model to be developed in Britain. It is an aggregate model which utilises the seats won by the major parties in the previous general election together with vote intentions six months prior to the general election to forecast seats. The model was reasonably successful in forecasting the 2005 and 2010 general elections, but has to be modified to take into account the 'regime shift' which occurred when the Liberal Democrats went into coalition with the Conservatives in 2010.



This paper utilises the Seats-Votes model to forecast the outcome of the General Election in Britain in May 2015. This model has been used with some success in the past to forecast both the 2005 and 2010 general elections (Whiteley, 2005, 2008; Whiteley et al., 2011; Gibson and Lewis-Beck, 2011). It is derived from the so-called 'Law of Cubic Proportions' formalised by the statisticians Kendall and Stuart (1950) in an article which represents the starting point of contemporary election forecasting modelling in Britain.

The literature on election forecasting in Britain has grown tremendously in recent years and a variety of approaches have been used to predict electoral outcomes (Whiteley, 1979; Mughan, 1987; Norpoth, 2004; Sanders, 1991, 2005; Belanger et al., 2005; Lewis-Beck and Stegmaier, 2011; Murr, 2011; Fisher et al., 2011; Lebo and Norpoth, 2011). These models utilise different methodological approaches and can be classified in a variety of different ways, but two important types recur in the literature. There are popularity functions, which utilise time series analyses of monthly or quarterly poll data to capture the relationship between voting intentions, the economy and other variables, for the purpose of forecasting. This approach was introduced by Goodhart and Bhansali (1970) in a seminal paper on economic voting. Secondly, there are vote function models which utilise data on election results, an approach pioneered by Kendall and Stuart (1950). These models use aggregate data measured over time or alternatively at the constituency level to derive the forecasts (Johnston et al., 2006).

Each method has its advantages and disadvantages. Popularity functions are based on relatively large numbers of observations, particularly in the contemporary era with the presence of many opinion polls, and this increases the precision of model estimates (Duch and Stevenson, 2008). On the other hand this approach faces the problem of translating vote forecasts into seats, since winning a general election in Britain does not mean winning most votes, but rather the most seats in the House of Commons. This is not a trivial distinction either since in the twentieth century the party winning most votes did not win most seats on three different occasions. This happened in 1929, 1951 and again in the February 1974, so there is a clear advantage in modelling the number of seats at the outset rather than analysing voting intentions which subsequently have to be translated into seats.

The Seats-Votes model uses aggregate analysis combining seat shares from all eighteen general elections since the Second World War with poll data to forecast seats in the Commons. It does not face the same problem as popularity function models, but the sample size is small making it essential to conduct extensive diagnostic testing to ensure that the models are well-behaved. It also requires the analysis to focus on political parties that have been in existence since 1945, and so has little to say about the rise of new parties such as the Scottish National Party or UKIP. These considerations aside, it is a relatively simple model with a respectable track record, although as the discussion below shows it has to be modified to deal with the era of Coalition politics.

* Corresponding author.
*E-mail address:* whiteley@essex.ac.uk (P. Whiteley).

http://dx.doi.org/10.1016/j.electstud.2015.11.015
0261-3794/

## 1. The Seats-Votes model

The seats-votes model adapts 'Law of Cubic Proportions' or the 'Cube Rule' to forecast seats shares over time. According to the Kendall and Stuart the Cube Rule:

*'..states that the proportion of seats won by the victorious party varies as the cube of the proportion of votes cast for that party over the country as a whole.'* (Kendall and Stuart, 1950: 183).

Using their example of the 'White' and 'Black' parties then:

$$\frac{W}{B} \Rightarrow \frac{P_0^3}{Q_0^3} \quad (1)$$

where:

W is the 'White' party and B is the 'Black' party seat shares
$P_0$ is the White party vote share and $Q_0$ is the Black party vote share, with $P_0 = 1 - Q_0$, so that:

$$W \Rightarrow B \cdot (P_0)^3 \cdot (Q_0)^{-3} \quad (2)$$

When they applied this model to the task of forecasting seat shares in the 1950 general election in Britain using poll data collected three days before the election the results were extremely accurate. The forecasting errors were one seat for Labour, five seats for the Conservatives and four seats for other parties, with the Liberal forecast being spot on (Kendall and Stuart, 1950: 194).

The key weakness of the model, fully acknowledged by the authors, was that it really only works in a dominant two party system in which it is safe to ignore minor parties. This was certainly true in 1950 when the Conservatives and Labour together took 90 per cent of the vote and 98 per cent of the seats. But as the British party system evolved towards the multi-party system of today the forecasts got progressively less accurate. In the early 1970s Edward Tufte (1973) suggested that a '2.5 rule' should be used as an alternative and Laakso (1979) showed that this appeared to work quite well at that time. But as we enter a new context of a fragmented multi-party system this is no longer the case.

Accordingly, we make three modifications to the Cube Rule to adapt it for forecasting seats in the 2015 election. The first change is to estimate the exponents rather than assuming that they are 3.0, thereby removing one source of error. Secondly, we utilise seat shares won by each party in the last parliament rather than the seat shares won contemporaneously by the rival party, as in equation (2). This is designed to capture the incumbency effect, which is partly a matter of existing MPs having a personal vote (Cain et al., 1987), but also the fact that parties represented at Westminster generally have much better coverage in the media than their non-parliamentary rivals and therefore are much better known to the general public. Incumbency bestows several advantages on the existing parties which need to be taken into account in the modelling.

The third change is that we utilise poll shares six months rather than three days prior to the election in order to make the forecast. The six month lag has been identified as the most efficient compromise between having the longest lead time for the forecast with the highest goodness of fit of the model (Whiteley et al., 2011). It is clearly advantageous to have as long a lead time for the forecast as possible without this degrading its accuracy and the six months lag achieves this goal.

The theoretical forecasting model is given by the following expression:

$$S_{it} = \alpha(S_{it-1})^{\gamma i} \cdot \prod_{i=1}^{k} (P_{it-m})^{\beta i} \cdot \varepsilon_i$$

where

$S_{it}$ is the seat share of party i at the election at time t
$P_{it\text{-}m}$ is the vote share for party i out of k parties, in the polls m months prior to the election
$\alpha$, $\beta^j$, $\gamma^i$ are parameters to be estimated
$\varepsilon_i$ is an error term where $E(\varepsilon_i) = 0$, $var(\varepsilon_i) = \sigma^2$

The theoretical model includes all rival parties but in practice this cannot be estimated since it would be perfectly collinear, so the empirical model estimates future seats for a party from its past seats and also from poll data for the party and its main rival.

For example, the Labour seat model in log-linear form is:

$$\ln(LabS_t) = \ln\alpha + \beta_1 \ln(LabS_{t-1}) + \beta_2 \ln(LabP_{t-m}) + \beta_3 \ln(ConP_{t-m}) + \ln g\varepsilon$$

where:

$LabS_t$ is the number of Labour seats won at election t
$LabP_{t\text{-}m}$ is the Labour vote share in the polls m months prior to the election
$ConP_{t\text{-}m}$ is the Conservative vote share in the polls m months prior to the election

The Conservative seat share model has the same specification as the Labour model but with lagged Conservative seat shares as a predictor. In previous versions the Liberal Democrat model utilised lagged Liberal Democrat seat share along with Liberal Democrat and Conservative vote shares in the polls (Whiteley et al., 2011). However, soon after the Liberal Democrats entered the Coalition government in 2010 a major change occurred to their support.

Fig. 1 shows vote intentions for the Liberal Democrats using monthly data from the Continuous Monitoring Survey from the date of the general election of 2010 election up to February 2015.[1] After the party obtained 23 per cent of the vote in the 2010

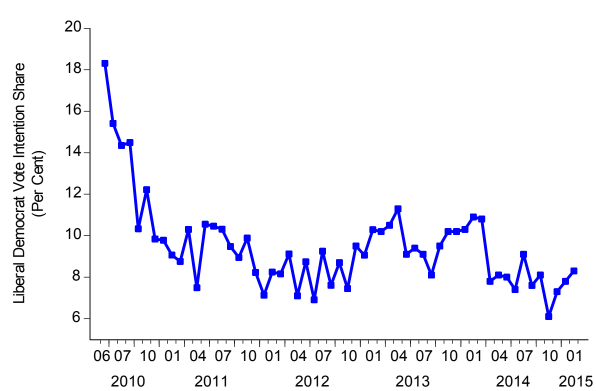


**Fig. 1.** Trend in Liberal Democrat voting intentions, June 2010 to February 2015.

[1] The Continuous Monitoring Survey of the BES ended in December 2012, and so the series is continued up to February 2015 using the same voting intention question in the Essex Continuous Monitoring Survey.