Bayesian Forecasting of Election Results in Multiparty Systems

Emil Aas Stoltenberg

Department of Political Science
Faculty of Social Sciences
University of Oslo
May 2013

Bayesian Forecasting of Election Results in Multiparty Systems

Bayesian Forecasting of Election Results in Multiparty Systems

Emil Aas Stoltenberg

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

I present a Bayesian forecasting model particularly suited for multiparty systems. The method I develop systematically combines (i) information from a Multinomial logit regression model fitted on historical data and (ii) estimates of current party support produced by a Dynamic Linear Model for multinomial observations. I apply the method to the Norwegian multiparty system, and assess the performance of the model on past elections. As of present, the model is ready to be updated as the Norwegian parliamentary elections of 2013 draws closer. The current forecast for the upcoming election is that the four opposition parties will obtain a majority in parliament with a probability of 0.775.

Acknowledgements

I am grateful to my supervisor Bjørn Høyland who gave me the idea of something fun. It has never been boring to work with this thesis. For some great years at Blindern I thank Aksel Braanen Sterri, Alf Butenschön, Anders Jupskås, Magnus Rasmussen, Per Anders Langerød, Rune Busch, Øivind Bratberg and Øyvind Skorge. Particularly, I want thank Tore Wig for being my lunch mate for three years and never accepting a dull conversation. Rather silence. I managed to find the time to write this thesis because Lars Petter Berg called and woke me up each morning. Ylva Zachrisson Dahl have prevented this semester from being too monotonous. Annelise Lehmann and Nadine Lehmann Irgens have made Seilduksgata a second home for me. Without the support of Atle Aas and Camilla Stoltenberg things would be have been difficult. All possible errors in this thesis are my own.

Contents

List of Figures

List of Tables

1

Introduction

Predicting the outcome of political events is an integral part of the practice and study of politics. In the social sciences pure prediction models are only applied to a limited number of problems, of which predicting the outcome of elections is one. What these models for predicting the outcome of elections have in common is that they are developed for political systems with two candidates or two parties competing. Multiparty political systems, the most common form of political system among the democratic countries of the world, have only received scarce attention (Clark et al., 2013, 578). In this thesis my aim is to rectify this deficiency by constructing a prediction model particularly suited for the case of a multiparty political system. With this model I generate predictions for the Norwegian parliamentary elections of 2013.

By way of developing a prediction model for elections in multiparty systems I am in fact working with a model suited for a specific class of prediction problems with a wide range of potential applications: The class of prediction problems that consists of predicting the future realizations of an unobservable (latent) variable. In order to predict the future course of an unobservable variable it is an advantage to know something about its location today. Translated to the problem at hand, this means that precise estimates of the *current* level of support for all the political parties in a multiparty system is essential for predicting the *future* level of support for these parties. Therefore, in a first part I develop a model for tracking party support on a day-to-day basis particularly adapted to the case of a multiparty political system. Experience from countries with examples of successful prediction models shows that elections are predictable from political and economic fundamentals (Lock and Gelman, 2010). Consequently, in a second part I estimate a model and use these estimates to generate a prediction that is solely based on political and economic variables. Finally, I combine the information provided by the two models to produce a forecast for the Norwegian parliamentary election of 2013. The forecasts produced by this model can be continuously updated as election day draws closer. As of May 21. this year my prediction is that the four opposition parties win a majority of the seats in parliament with a probability of 0.775.[1]

1.1 Aim of the thesis

Prediction is at the core of all fields of science. The successive realizations of a prediction is a fundamental criterion by which to judge the validity of a scientific model, and it is by basing our decisions on the models that pass this test that we approach the world in a rational manner. As noted by Hempel, the same holds for explanations, they are only fully adequate if their explanans could have served as a basis for predicting the phenomenon under study (1948, 138). Carnap, one of the founders of the Vienna circle, also arrived at this definition for scientific explanations, and judged the explanations that lack predictive capability as not being explanations at all (2006, 565). In the social sciences this point has recently been reiterated by Schrodt (2010, 565), who attacks the notion that explanation does not imply prediction, a claim that according to him is widespread among social scientists. Although not concerned with explanation per se, this thesis develops and improves methodology that can be used for testing the predictive capability of explanations. This thesis is concerned with prediction problems of a particular kind, namely those that try to predict the future realization of an unobservable (latent) variable. Many social and political phenomena are of this kind, their true value is unknown to us until they take on one of a limited number of values. For clarity of exposition I divide these variables into three (not exhaustive nor exclusive) types. First, there are the phenomena that are only observable, for practical or for more substantive reasons, in a limited number of its states. By "states" I mean the conditions that a given phenomena can exist in. Second, there are the variables that are only measured at certain intervals of time, so that the values they take on in the intervening time is unknown. Third comes the variables whose current value are only determined with a certain delay. Examples of the first type of phenomena differ in whether their unobservability are due to practical or more substantial reasons. If one is looking at a wall with three windows and a person inside: the whereabouts of the person (the value of the location variable) is only revealed to us when she appears behind

[1]The seven parties currently represented in the Norwegian parliament are Arbeiderpartiet (Ap), Sosialistisk Venstreparti (Sv), Senterpartiet (Sp), Kristelig folkeparti (Krf), Venstre (V), Høyre (H) and Fremskrittspartiet (Frp). The three first of these have since 2005 been coalition partners in the Red-Green coalition. When I refer to the opposition, I mean the four latter parties on this list. Through the entire thesis I use the Norwegian names for these parties and their abbreviations interchangeably.

one of the windows. In the meantime we can only guess at her location based on which window we last saw her appear behind. Here it is the practical obstacles due to an apartment's limited number of windows that makes the location variable partly unobservable. In other cases the unobservability is less due to practicalities. Consider the stereotypical study of the civil war literature that has as its dependent variable a dichotomous measure taking the value one if civil war is observed and zero otherwise. If, as one can argue, civil war is not an either-or phenomenon and the observed outcomes are just the ends of a continuum where the actual variable of interest is, perhaps, political instability, then the observable dichotomous variable is in fact generated by another variable that civil war is an easily observable expression of. As this first type of latent variables is presented here, in both examples there is in fact a latent variable y^* ranging from a to b (where a and b might be minus and plus infinity) that generates the observed values, the y's. In the case of civil war, the units with y^* larger than some threshold are observed as $y = 1$ while those below the threshold are observed as $y = 0$ (Long, 1997, 40). In the windows example the observation y can only take on as many values as there are windows, while y^*, the unobservable position in between the windows is continuous and possibly infinite.

The second type of latent variables are best thought of as variables with missing values at most points in time. Sometimes missing data points can be caused by poor statistical reporting, a common problem when working with aggregated national level data. In other situations, the lack of observations is inherent to the phenomenon under study. The latter is the case if measurement follows a pattern or is associated with certain dates, but the variable evolves in between measurements. Support for poltical parties, which is only really measured on election day, is a case in point. In cases like this the variable of interest is a time series $\{y_t\}$ for which we, for example, only have observations at $t = 1, 101, 201, \ldots$ and so on, where t might be days, years etc..[2]

Finally, I place the variables that are only determined with a certain delay in the last category. These are often composite variables that demand extensive data collection to actually measure. A country's gross domestic product (GDP) is a pertinent example of such a variable. Normally, precise measures of a country's GDP are determined with a delay of several months. To know the current level of activity in the economy is information that policymakers and investors need to make decisions, and economists have termed the expression "nowcasting" to the task of estimating the size of GDP in real-time (Giannone et al., 2008).

[2] In the case of political elections one objection can be reaised to this exposé. One can argue that intending to do something, vote for a given party, is something qualitatively different from actually doing it.

For these three types of unobservable variables one thing is clear: If one seeks to make forecasts of future values (or realizations) of a latent variable, a good estimate of its current value is advantageous. The point is that a good estimate of its unobservable value is likely to provide for better forecasts than only knowledge of its observable value (i.e. zero or one), or its last observed value (i.e its value last month). To continue with one of the examples above. Given that one wants to make statements about the chances of civil war occuring at future points in time, one probably wants to know something about the current value of the latent variable. It is an advantage, but certainly not optimal to only know whether the country in question presently is a zero or a one, that is whether or not the country suffers a civil war. In order to make good forecasts one has to know something about the risk of civil war occurring (Hegre et al., 2012, 3). And *risk* is, in effect, the name that we put on the unobservable variable that generates the observed outcomes, war and peace. To summarize, common for this general kind of prediction problems is that one is dealing with two unknowns: The future realization of the latent variable is clearly unknown, and the present value of the latent variable is also, for various reasons, unknown.

In this thesis I develop a model that attempts to handle one instance of this general class of prediciton problems: namely the problem of forecasting the outcome of political elections. Election forecasting is, for two reasons, a good field for the development of such models. First, the level of support for a given party by the next election is clearly unknown, and the level of support for a party today is also for practical reasons an unobservable quantity. Since we have little reason to believe that the support for political parties is invariant between to elections, only to change on election day, support for political parties is a variable with missing values at most points in time. It belongs to the second type of latent variables I outlined above. Second, as with many latent variables, I do actually have measurements of the variable of interest, but the measurements (the political polls) are inaccurate and errorprone.

Models for forecasting the outcome of elections are not novel and several have been utilized with success. In the next chapter I provide an overview of the literature on election forecasting and place the models developed in this literature in four different categories, depending on the theoretical approach or the model applied. What the contributions in these four categories have in common is that none are particularly adapted to forecasting the outcome of elections in multiparty systems, they are all developed for political systems with two presidential candidates or two political parties/blocs competing for power. Thus, when attempting to construct a model fit for the dual task of of making inferences on the latent value of party support in between elections and forecasting the outcome of elections in multiparty systems, one is faced with challenges that relate to theoretical as well as statistical issues. In this thesis I attempt to do something with the latter. To this end I adopt the following strategy. First, I assess how a model that have been used to track the evolution of voter sentiment in presidential and two-party/bloc systems works when applied to the Norwegian multiparty system. Second, I develop another model for the same task that is based on other distributional assumptions that are more sound when tracking the support for parties in a multiparty system. Both of these models are estimated with political polling data exclusively. Third, I draw on arguments put forward in the literature I review below to build a model for forecasting the outcome of elections in multiparty systems. Finally, the information provided by these two models is combined to produce predictions for the election outcome of the seven main political parties in Norway, and these predictions are in turn translated to actual seats in parliament. In all of the steps outlined above I rely on Bayesian statistical methods, a branch of statistics that I show, lends itself naturally to the problem at hand. The methods I use for forecasting elections are applicable in all multiparty political systems, not only the Norwegian, and in the conclusion I outline possible extensions and ameliorations of the models in further research. When it comes to the application in this thesis, the Norwegian parliamentary elections of 2013, I find that the result of this election will most probably be an alternation in government. The parties of the Red-Green coalition, Arbeiderpartiet, Sosialistisk Venstreparti and Senterpartiet, are predicted to receive 30.3%, 6.6% and 5.8% ($[28.8\%, 31.8\%]$, $[5.8\%, 7.4\%]$ and $[5.1\%, 6.6\%]$ are the 95% confidence intervals) of the votes respectively and consequently lose seven seats in parliament with a total of 79 seats. The opposition parties of Kristelig folkeparti, Venstre, Høyre and Fremskrittspartiet, on the other hand, are predicted to obtain 6.8%, 4.7%, 25.1% and 16.3% ($[6\%, 7.6\%]$, $[4.1\%, 5.5\%]$, $[23.7\%, 26.6\%]$ and $[15.1\%, 17.6\%]$ are the 95% confidence intervals) of the votes respectively and receive a total of 87 seats.[3] In addition, the category of parties I call "others" have a mean prediction of 3 seats. As of May 21. the predicted probability of the opposition passing the majority mark of 85 seats is 0.775.

1.2 Outline

Election forecasting is a developed art in a handful of democratic countries. In Chapter 2 I present some of the most important contributions in this field, with special attention

[3]To be precise, the uncertainty estimates are not confidence intervals, but highest density regions (HDRs). HDRs will be defined in Section 3.3.

paid to their relevance for the problem at hand. The central part of this thesis is the development of a model for forecasting elections in multiparty systems. Two important ingredients of this model are political polls and Bayesian methods. Therefore, before I proceed to the main part of the thesis, I provide a general introduction to both topics in Chapter 3. The forecast model of this thesis consists of two parts and I present them separately. In Chapter 4 I present the Dynamic Linear Models (DLMs) that I use to make inferences about the latent level of current party support. In Section 4.1 I introduce the DLM used by Jackman (2005) and Beck et al. (2006) to track party support and presidential approval in a two-bloc and a presidential system. I elaborate on the potential problems with applying this model to a multiparty system such as the Norwegian. Thereafter, in Section 4.2, the second part of the DLM chapter, I develop a model that is theoretically sound for modeling the distribution of support among the parties in a multiparty system. With these two models I use polling data to track party support in Norway from 1997 until present and estimate the possible bias of individual polling institutes. Finally, the two DLMs are compared and their estimates are presented in a series of graphs. In Chapter 5 the *forecasting* part of the forecast model is developed. I present two methods for predicting election outcomes. The first is a simple method for extrapolating the time series estimated by the DLM, while the second method combines predictions generated on the basis political and economic variables and combine these with estimates from the DLM. Since both methods produce predictions of the national votes shares of individual parties, the chapter begins with a section on the Norwegian electoral system and on my strategy for getting from estimates of a party's national vote share to its actual number of seats in parliament. Both methods are tested out-of-sample on previous elections. In Section 5.5, I present my forecast for the Norwegian parliamentary election of September 2013. Finally, in Chapter 6 I conclude with a discussion of possible extensions and improvements of the model of this thesis.

2

State of the art and research gaps

With minor expections Norwegian elections have not been subject to attempts at forecasting using rigorous methods. Certainly, political pundits as well as political scientists make statements about the outcome of the next election, but seldom if ever, do these predictions come with an estimate of uncertainty attached to it. Furthermore, the pieces of information that these forecasts are based upon (polls combined with general knowledge about Norwegian history and politics) are seldom, if ever, combined in a manner consistent with the laws of probability. Arnesen (2012a) provides the first systematic attempt at constructing models for forecasting the outcome of Norwegian parliamentary elections. The situation is different in other advanced democracies, such as the United States, the United Kingdom, and France, where election forecasting is quite common. In this section I provide a review of the approaches utilized in election forecasting in these three countries, and discuss if and how these models are tranferable to the Norwegian case. Broadly defined, when it comes to the methodology utilized, the models can be grouped into four categories: (i) economic vote models; (ii) electoral cycles models; (iii) models usings prediction markets; and (iv) models that use political polls as their primary source of data.

2.1 Economic voting

Most of the models developed for the US, the UK, and France combine economic variabels such as GDP growth, unemployment figures and inflation with an incumbency dummy and some measure of governmental approval. These forecast models draw their theoretical underpinnings from the field of *economic voting*, a field that according to Lewis-Beck and Paldam (2000, 113) "mixes economics and political science [...] by the means of econometrics". Studies in this tradition rests on the responsibility hypothesis, according to which voters hold the government responsible for economic events (Lewis-Beck and Paldam, 2000, 114).[1] In simple terms, if economic times are bad the voters are expected to turn their back on the incumbents. Foucault and Nadeau (2012) is a recent example of an economic vote model, developed for foreasting the French presidential election of May 2012. In accordance with the responsibility hypothesis Foucault and Nadeau state that the now former president Nicolas Sarkozy was likely to be held accountable for the poor performance of the French economy (2012, 218). And indeed, the model of Foucault and Nadeau did sucessfully predict the defeat of Sarkozy, and the victory of the current president François Hollande. The model of Foucault and Nadeau is illustrative of the two main assumptions of the economic vote literature, which they state clearly. The first assumption is that the electoral outcomes can be satisfactory explained by a limited number of economic and political variables. Second, that the values taken by these variables several months in advance of the elections often are more useful for predicting elections than information picked closer to election day (Foucault and Nadeau, 2012, 219). The latter assumption can be empirically investigated, but the contributions that systematically investigate the optimal lag-structure are rare or non-existent (I have not found any). For example, in their forecast model for the French presidential elections Nadeau et al. (2010, 12) choose a six months lag on the independent variables, a choice that they base "partly on theory and partly on empirics", with no further justification. Generally, the important point here is that since the forecasts based on economic vote models are made several months in advance of the election, information of interest that appears between the forecast date and election day is not incorporated into the models. That is, in most instances the forecasts are not systematically updated as election day nears. Below, I present one exception to this general pattern, and in Chapter 5.3 I develop an economic vote model that can be continously updated as election day nears.

A good defense of economic vote models is found in Bartels and Zaller (2001). They examine the Al Gore vs George Bush 2000 US presidential election that promted many analysts and political scientists to cast doubt on the basic premise of economic vote models, that economic and political factors play a systematic and largely predictable role in shaping the presidential election outcomes (Bartels and Zaller, 2001, 9). The reason for this was that GDP had been growing at a steady rate in the year leading up to the 2000 election, and that the incumbent president Bill Clinton (a Democrat as Gore) was highly popular. Given these two facts, Gore was expected to win comfortably, and his

[1] Lewis-Beck and Paldam (2000) and Nannestad and Paldam (1994) provide good reviews of the literature in this tradition.

failure to do so was in turn imputed on Gore's personality and his efforts during the campaign. Bartels and Zaller (2001) conduct a meta-analysis of 48 regression models from the economic vote literature to examine whether the 2000 US election really refutes the fundamental premises of the theory. They conclude that, rather than refuting the theory, the presidential election of 2000 was largerly in line with what the theory predicts. First, they find that even though the GDP had been growing, the real disposable income (RDI) per capita of 2000 was below the post-war average. Importantly, by the meta-analysis they find the rate of RDI per capita growth to be a more accurate predictor of presidential election outcomes than GDP growth (Bartels and Zaller, 2001, 10). Second, the fact that Clinton had served not one, but two terms, caused a certain fatigue with the incumbent party that counted negatively for Gore. These two variables both belong to what is called political and economic fundamentals in the economic vote literature, and taken into account Gore did in fact as well as one would expect (Bartels and Zaller, 2001, 10). To summarize, the point that Bartels and Zaller convey is that appeals to election-specific explanations of the 2000 election are misplaced, and that the economic vote theory fare well in face of the outcome of the 2000 US elections (Bartels and Zaller, 2001, 18).

An important limitation with the economic vote models derived from the responsability hypothesis is that it is not evident what this kind of models should look like for multiparty systems with changing coalition partners in government. Nannestad and Paldam (1994, 213) go as far as stating that such models only work in two party/bloc systems. This causes no problems for economic vote models built to forecast presidential elections in the US and France, as well as parliamentary elections in the UK, but application of the economic voting scheme on the Norwegian political system is not straightforward. In his article on economic voting in Norway Arnesen (2012b) draws on a branch of the economic vote theory that does not relate solely to the performance of the government, but emphasizes how parties at different ends of the political spectrum maintain issue-ownership over different economic domains, and thereby benefit from varying economic conditions (Carlsen, 2000; Hibbs, 1977; Petrocik, 1996; Swank, 1993). According to the *clientele hypothesis* of this theory, growing unemployment and low economic growth favours the left, while a prospering economy and a low unemployment is to the advantage of the parties of the right (Arnesen, 2012b, 4). To test this hypothesis on the case of Norway Arnesen argues that, for analytical purposes, it is fair to pool the Norwegian political parties in two blocs, the left and the right (2012b, 7). With this division at hand Arnesen runs regression models that cover the national and the local elections from the first post-war election in 1945 until the last election of 2009, and his findings are in line with the clientele hypothesis: the aggregate vote for the parties on the left of the ideological spectrum increases when citizens fear for their jobs (2012b, 18). Pooling the parties can indeed be a good strategy for analytical purposes, for the purpose of predicting the outcome of elections there are, however, some weaknesses with this approach. First, problems occur if there exists parties that cannot be neatly placed in one bloc or the other. In Norway, Senterpartiet (Sp) who made a switch to the left bloc before the election of 2005 is a case in point. To tackle this issue Arnesen (2012b) excludes this party from the analysis. But, given that the Red-Green coalition, which includes Sp, is aiming for re-election in 2013 the exlusion of Sp makes the model unapt for forecasting. Second, the model is asymmetric in the sense that one minus the left bloc share does not equal right bloc share. This is due to the fact that one has seen centrist coalitions in Norway (in 1972-1973 and in 1997-2000), and in the event of a Red-Green defeat in the 2013 election, it is not clear which parties will form the new coalition. Third, and most importantly, as I show in Section 5.1, the conversion from national vote shares to seats in parliament is not one-to-one in the Norwegian electoral system. Therefore, precise forecast estimates of the support for each individual party is needed in order for a forecast to be a good forecast. This means that an economic vote model must say something about the effects of political and economic conditions on each individual party (that it is worth considering). A general statement about what will happen to "the left" given some economic indicators is not sufficient.

2.2 Electoral cycles

Inherent in the concept of democracy is alternation. As Norpoth (1991) remarks, "as long as people have chosen political leaders through some form of election, it has been noted, almost like a law of politics, that popularity diminishes with time in office." Nannestad and Paldam (1994) estimate that it costs the average government 2 percent of the vote to rule. Due to this cost-of-rule effect one should expect to observe some form of electoral cycles where power is passed from one side of politics to the other in a more or less regular fashion. Some forecast models estimate and include this cost-of-rule effect. An original attempt is Lebo and Norpoth (2007, 72) who remarks that "the swing of the electoral pendulum is as British as ale and kidney pie". Even if this is so, they do face the complicating factor of the period and amplitude of the electoral cycles being irregular. This excludes the use of a simple sine-function to describe the swings of the electoral pendulum. Instead, Lebo and Norpoth (2007) rely on a second-order autoregressive model originally developed to track the irregular fluctuations of sunspot observations. In addition to the two pendulum coefficients their model only includes one other variable: the approval rating of the prime minister. With this parsimonious model Lebo and Norpoth were able to predict that the outcome of the 2010 British election would be a Hung Parliament. Another model in the same genre is the time-for-change model developed by Abramowitz (2008) for presidential elections in the US. This model includes the length of time that the incumbent president's party has controlled the White House and a dummy variable indicating whether a party has controlled the White House for one term or for two or more terms. The first of these Abramowitz (2008, 692) calls the time-for-change factor, while the latter is intended to capture the strength of the time-for-change sentiment in the electorate (Abramowitz, 2008, 693). With this model Abramowitz predicted that Barack Obama would receive 54.3% of the major party vote and that John McCain would receive 45.7% (the actual figures were 53.6% and 46.4%).

2.3 Prediction markets

Prediciton markets are, to my knowledge, only experimented with once in Norway. Arnesen (2011b) conducted a small scale prediction market experiment prior to the election of 2009. Prediction markets are internet based betting markets where the purpose (for the researcher) is to use the information content of the market values to make predictions about future events (Arnesen, 2011a, 45). The idea is twofold. First, contrary to political polls, that provide estimates of the current political preferences among potential voters, the information in the prices in a prediction market provides an estimate of the outcome on election day. Second, given the financial incentive the participants in the market have to make accurate forecasts, there is considerabe incitement for digging deeper for relevant information, and not least, not to let oneself be blinded by what one wishes the outcome to be. In another article where he studies the *Iowa Electronic Market* (a prediction market in the US) Arnesen (2011a) corroborates these two ideas. Considering the 2004 and 2008 presidential campaigns the variability of the market predictions are much less pronounced than that of the polls. Furthermore, the market predictions lie closer to the actual election result during the whole period under study (Arnesen, 2011a, 53). The findings from the 2009 experiment in Norway are more mixed. Possible reasons for this is the limited number of participants in the market, multiparty politics being more difficult to forecast, and not least, that gambling is illegal in Norway. Lacking the financial incentive it is harder to make the case for the rational behaviour of the participants in the market. Consequently, the results of the 2009 experiment are inconclusive: they do not show that prediction markets are superior to other methods of predicting the outcome of elections in a multiparty system (Arnesen, 2011b).

2.4 Poll based methods

The last branch of the literature that I consider consists of those studies that base their analysis and forecasts, solely or primarily, on political polling data. Some of these contributions focus less on explicit forecasting, and concentrate instead on locating the level of current support for political alternatives. In the terms used in the previous section, this means that focus has been more on determining the level of the latent state today, than forecasting its realization on election day. Jackman (2005) is an example of such a study, where a Dynamic Linear Model (DLM) is used to track the latent state of support for the two blocs during the months leading up to the Australian election of 2004 (the DLM will be introduced in Chapter 4). His study is conducted after the election and its goal is to show how one effectively can take advantage of the information conveyed by polls to measure the bias of each individual polling house and determine the effects of events occuring during the election campaign. Jackman (2005, 514) finds that the variability in the estimates provided by the different polling houses is much larger than what can be explained by random sampling. In addition, he notes that the largest house effects are associated with the mode of interview. In the same domain is the study by Beck et al. (2006) of the Bush presidency. The question that the Beck et al. article attempts to answer is: how much did the Katrina debacle (the hurricane that hit New Orleans in August 2005) hurt Bush's approval (2006, 1). In order to give a precise answer to this question Beck et al. need a method that can separate out the consequences of Katrina from the long term decline in Bush's approval, and single out the true level of the latent variable, approval of the president, from noisy measurements (political polls) (2006, 2). To handle this dual challenge Beck et al. use a DLM that they feed polling data and some economic variables thought to influence the approval of the president. The work is preliminary, and even with the amount of data they have at hand they find it hard to single out the effects of Katrina. This is especially so because the hurricane was followed by a quick sucession of presidential missteps (Beck et al., 2006, 23).

A more explicit poll based attempt at constructing a forecast model for the US is Lock and Gelman (2010). Since the US presidential elections are decided in swing states, they argue that one should look at state polls. But state polls are noisy, so one needs a method to detect the information that these polls actually do contain. The observation that even with wide national swings in the support for a candidate the spatial distribution of support remains fairly stable, leads Lock and Gelman to construct a Bayesian model that integrates prior data (the 2004 election results) and local level poll data to arrive at estimates of the position of each state relative to the national popular support for each candidate (2010, 338). In this manner they are able to determine how much information a local level poll carries, and thereby use local level polling results to predict the election outcome in each individual state. The model of Lock and Gelman performs well in forecasting previous presidential elections, but in their 2010 article Lock and Gelman do not apply the model to the 2012 election.

A forecast model for the US that combines insights from three of the branches of the literature on forecasting elections is Linzer (2012). Linzer introduces a Bayesian forecasting model that unifies the regression-based historical forecasting approach (as in the economic vote and time-for-change models) with the poll-tracking capabilities made feasible by Bayesian models such as those used by Jackman (2005) and Beck et al. (2006). In fact, what Linzer attempts to resolve is the problem I presented above in connection to the models of Foucault and Nadeau (2012) and Abramowitz (2008), namely that since the forecasts based on economic vote models are made several months in advance of election day, these models do not exploit the information that appears in the time between the day the forecast is made and election day. Linzer (2012) recognizes that structural models that predict election outcomes from economic and political fundamentals such as the level of economic growth, changes in unemployment and whether the incumbent is running for re-election etc., often provide for accurate forecasts. The deficiency that Linzer seeks to rectify is that these structural models contain no mechanism for updating predictions once new information becomes available closer to election day. What Linzer does is that he uses Bayesian methods to continously update the forecast generated by a structural model with local level polling data. In this sense, the structural model produces a prior forecast that is in turn revised by combining it with his estimates of the current latent level of support for the two candidates. Linzer's model (see his blog `votamatic.org`) correctly predicted the Barack Obama victory in the 2012 US presidential election.

In this brief review I have covered the four main branches of the literature on election forecasting, with an eye on how theory and modelling strategies can be adopted from these contributions to the case of a multiparty system. A general feature of this literature is the penury of models particularly adapted to the multiparty case, and I have pointed at some of the problems associated with direct application of models built for presidential and two-bloc/party political systems to a party system such as the Norwegian. In the approach I adopt to start alleviate this deficiency of the literature there are two prime ingredients: political polls and Bayesian methods. Before I proceed to the actual modelling a general discussion of the problem with political polls, and a short introduction to Bayesian statistics, is due. These two things are the topic of the next chapter.

3

Key Ingredients: Political polls and Bayesian methods

The fundamental ingredients of this thesis are political polls and Bayesian methodology. In the first part of this chapter I discuss the problems associated with political polls. In a second part I provide a brief introduction to Bayesian statistics and inference, and show why Bayesian methods are particularly apt for the problem of making inferences on a latent variable for which there are only noisy measurements.

3.1 The problem with polls

The most common way of making inferences about the support that a political party enjoys, is to look at the latest poll. In this section I show why this is a problematic strategy. There are primarily three issues that limit the usefulness of polls for social scientific purposes (Jackman, 2005, 500). First, imprecision due to sampling error. Second, the polling institutes use different methods of interviewing, as well as different weighting schemes that can potentially induce systematic biases in the estimates. Finally, polls with rather small samples are not capable of capturing the fine grained day-to-day variations in voter sentiment. In order to answer questions concerning the effect of a particular event, precise estimates of these day-to-day swings are necessary. In the following I will elaborate further on the limits of political polling.

Polling institutes almost always report margins of error. These are most often 95% confidence intervals around a point estimate $\hat{\alpha} \in [0, 1]$ of α, with the confidence interval given by

$$\hat{\alpha} \pm 1.96\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{n}} \tag{3.1.1}$$

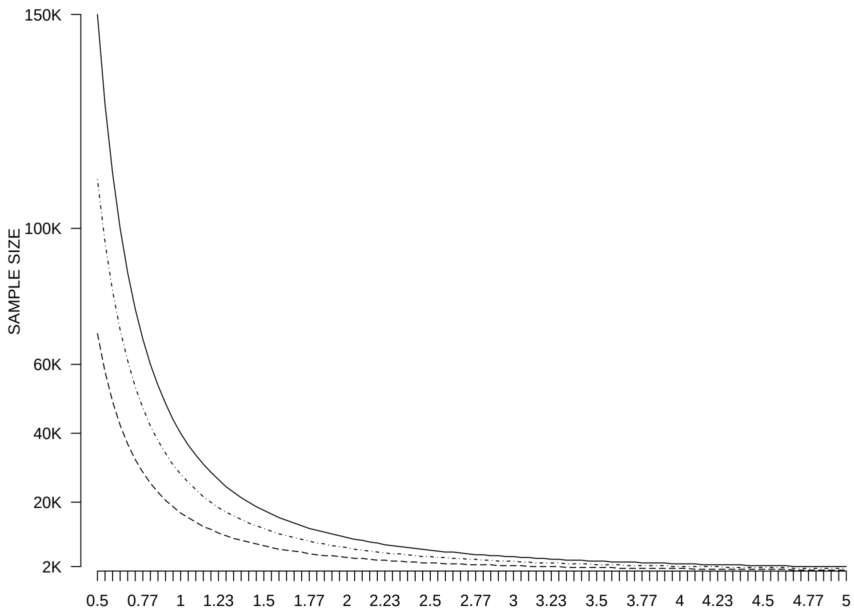


Figure 3.1: Required sample size. The y-axis shows the sample size required to detect a given percentage point change (x-axis) in voter sentiment. Assuming a baseline level of 50%. The solid line is 95% CI, dashed lines are 90 and 80% CIs.

Where $z_{0.05/2} = 1.96$ in Equation (3.1.1) comes from the fact that a statistic such as $\hat{\alpha}$ computed with a large sample follows the normal distribution (Devore and Berk, 2007, 293). If a random variable follows a normal distribution then we expect the mean of 95% of such samples from the same population to lie within 1.96 standard deviations. From Equation (3.1.1) we also see that the statistical precision is a function of $\sqrt{n}$, the second derivative of this function $-(4n^{3/2})^{-1}$ highlights the fact that an increase in sample size produces diminishing marginal returns in statistical precision. The additional cost for another survey respondent, on the other hand, remains more or less constant. Since the polling institutes have limited time to conduct their polls and want to make money, it is natural that the sample sizes remain rather small and the precision limited. Almost all the polls used in the analysis in this thesis have sample sizes of between 800 and 1000 respondents.[1] These sample sizes are too small to detect small but potentially significant changes in voter sentiment (Jackman, 2005, 501). To illustrate the limits of these sample sizes Figure 3.1 graphs the sample sizes necessary to detect various percentage point changes in voter sentiment. By visual inspection it can be seen from the plot that a sample

[1]For example: average sample sizes between September 2009 and January 2013: Synovate 948 respondents, Gallup 966.4 respondents, and Opinion, Nielsen and Sentio 1000 respondents.

size of about 60 000 respondents is needed in order to have a 95% chance of detecting a 1 percent change in voter sentiment. Even if with a lower level of statistical significance, represented by the dashed lines, the researcher will need about around 20 000 respondents to detect a 1% change. The point is that with the sample sizes that the polling institutes normally use, we are not able to detect changes in the interval whithin the support for most parties fluctuate. In order to be able to detect these normal variations larger samples are clearly necessary.

A typical remedy for insufficient samples is to pool the polls. The most basic way to pool polls is by taking the average of the estimates. A slightly more sophisticated way of pooling the polls is to take a precision weighted average (Jackman, 2005, 503). With a precision weighted estimate one takes into account that the estimates are based on different sample sizes, and as the name says, one puts more emphasis on those estimates based on larger samples. If the polling houses Opinion (O) and Gallup (G) provides estimates $\hat{\alpha}_O$ and $\hat{\alpha}_G$, then the precision weighted estimate is

$$\hat{\alpha}_{OG} = \frac{p_O \hat{\alpha}_O + p_G \hat{\alpha}_G}{p_O + p_G}$$

where $p_O = 1/\sqrt{\text{Var}(\hat{\alpha}_O)}$ and equivalently for Gallup. The standard deviation for the pooled estimate is then $\sqrt{1/(p_O + p_G)}$, which is clearly smaller that the standard deviation of any of the two estimates individually.

3.2 House effects

Pooling polls will always result in tighter confidence intervals, but the pooling rests on a critical assumption: that the polls are unbiased. Beck et al. (2006) and Jackman (2005) show that in the case of the US and Australia this is not the case. The polls are subject to bias, and the bias is often specific to each particular polling organization. This is known as "house effects", where "house" refers to the polling organization. Differences in the mode of interviewing, the wording of the questions, the time of the day, the sampling procedures, and the different weighting procedures utilized, all have potential to induce house specific biases in the estimates. The important point is that pooling several biased estimates does not in general produce an unbiased estimate.

As an explorative example of the possibility of house effects in the case of Norway, consider the estimates for the support for Fremskrittspartiet (Frp) provided by five polling intitutes from 2009 to the present.[2] The typical survey in this period employs a sample size of $n = 1036$ respondents, and the average estimate for Frp is about 20%. Under the assumption that the survey houses employ the same unbiased random sampling procedures half of the poll results should lie within plus minus 0.67 standard deviations σ, where $\sigma = \sqrt{.2(1 - .2)/n} \approx 0.01$. With an estimate of 20% half of the polls with $n = 1036$ should lie between 19.3% and 20.7%, which gives an average inter-quartile range a little below two percentage points. The expected inter-quartile range of about two percentage points holds roughly for all the polls results for Frp. Inspection of the poll results suggests more dispersion than what is expected under simple random sampling. In 42 of the weeks in the sample five or more polls are available (which for the sake of this example I accept as a reasonable number to try to compute an inter-quartile range). In those 42 weeks the inter-quartile ranges of the polls range from 0.3 to 6.6 percentage points with a mean of 2.43 percentage points, which is larger than what one would expect under random sampling assuming, given that the level of support for Frp does not vary much within the same week. In the more formal analysis in Section 4.4 this suspicion of rather large house effects in the estimates for Frp is confirmed.

Pooling polls can alleviate one issue of polling, the lack of precision. The problem is that this strategy assumes that house effects are non-existent. This is in general a risky strategy since the chances of the two biases cancelling each other out is rather small, and if the biases run in the same direction the bias will be exaggerated in the pooled estimate, and one will be falsely more confident about a flawed estimate. The conclusion is that in order to pool the polls we must have estimates of the bias in each of the polls (Jackman, 2005, 505). This is a challenge, because in order to say something about the bias of each poll we must know the population quantity that the polls try to estimate, but the whole point of polling is that this population quantity is unknown. In Chapter 4 I describe how the house effects are estimated in the framwork of a DLM. The next section gives a brief introduction to Bayesian statistics, with examples relating to polling data.

3.3 Bayesian analysis: A short intro

Bayesian estimation and inference relies of Bayes Theorem.[3] In a general sense, Bayes Theorem tells us how to rationally update our beliefs in light of data. In political science these beliefs are most commonly probability statements about parameters, hypotheses

[2] The exploration of possible house effects is inspired by Beck et al. (2006).

[3] It is often written "Bayes' Theorem" with an apostrophe, because it looks nicer and is common in the literature I drop the apostrophe.

and models (Jackman, 2009). Bayes Theorem describes how one's probability statements prior to observing data should be revised to updated probability statements after having observed data. The updated knowledge is then a combination of one's prior beliefs and the parameters most likely to have generated the observed data. Usually one says that *prior* beliefs become *posterior* beliefs through the act of observing data, that is

$$\text{priors} \rightarrow \text{data} \rightarrow \text{posterior}$$

Bayes Theorem itself follows directly from the rules of conditional probability. If $A_1, ..., A_k$ are k mutually exlusive events ($A_i \cap A_j = \varnothing$ for all $i \neq j$), where the union of these $A_1 \cup ... \cup A_k$ make up the whole sample space S, and all of these k events have a probability greater than zero of occurring, that is $P(A_i) > 0$ for $i = 1, ..., k$. Then, for any other event B in S with $P(B) > 0$

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

As an example assume that a woman is interested in the chances of being pregnant after a single sexual encounter. She buys a test where the text on the package says that the test correctly classifies 89% of the pregnant woman, but that the test also classifies 15% of the non-pregnant woman as pregnant. The woman tests positive and remembers that among her friends one in ten sexual encounters have resulted in pregnancy. By Bayes Theorem she reasons that

$$P(\text{preg}|T+) = \frac{P(T+|\text{preg})P(\text{preg})}{P(T+|\text{preg})P(\text{preg}) + P(T+|\text{not preg})P(\text{not preg})}$$

$$= \frac{0.89 \times 0.1}{0.89 \times 0.1 + 0.15 \times (1 - 0.1)} = 0.397$$

the probability of being pregnant, given the positive test, is 0.397. Perhaps this result does not satisfy the woman and she decides to take the test one more time. This time her prior estimate $P(\text{preg}) = 0.397$ equals the posterior estimate she obtained the first time she applied Bayes Theorem. If she tests positive again, the probability of her being pregnant is 0.796. A subsequent positive test would yield a probability of 0.956. This process of continuously updating our beliefs in face of new data highlights an important aspect of Bayesian statistics: we do not regard the world anew every time we attempt to answer a given hypothesis, rather we accumulate information such that the posterior distribution becomes more and more precise (and more precise than the prior and the likelihood separately) (Lynch, 2007, 49).

Even though the woman reasons in accordance with the laws of probability, some objections can be raised. Most evidently, the woman seems a bit too crude in her assessment of the underlying probability of becoming pregnant after a single sexual encounter. As her estimate of 1/10 is based on a rather limited sample of friends, she can not be certain about it. A more rational approach would be if she could say something about the uncertainty in her estimate, in other words if she could ascribe a probability distribution to her estimate. This would lead her to Bayes Theorem applied to probability distributions. Expressed in terms of probability distributions Bayes Theorem reads

$$f(\theta|\text{data}) = \frac{f(\text{data}|\theta)f(\theta)}{f(\text{data})} \tag{3.3.1}$$

where, when the parameters are continous,

$$f(\text{data}) = \int f(\text{data}|\theta)f(\theta)\text{d}\theta$$

with summation for discrete parameters. $f(\text{data}|\theta) = L(\theta)$ is the probability of the data given the parameters: the likelihood function. The distribution that is obtained by multiplying the prior and the likelihood is called the posterior distribution. The integral above is the normalizing constant that ensures that the posterior integrates to one, and it is often most convenient to drop this one when doing calculations (DeGroot and Schervish, 2012, 391). From this the classic statement of Bayes Theorem follows, namely that the posterior is proportional to the likelihood times the prior

$$f(\theta|\text{data}) \propto f(\text{data}|\theta)f(\theta) \tag{3.3.2}$$

To illustrate the use of Bayes Theorem with probability distributions I will consider a poll published by the polling institute Opinion in December 2009. This poll gave the Red-Green coalition 38.5% of the vote intentions and had a sample of 1000 respondents (I'll call the proportion of Red-Green votes for rg). Knowing that the coalition received 47.8 of the actual votes on the election day in September 2009, had their support really decreased by almost 10 percentage points? We might be interested in asserting the probability that their support was below 40 percent, that is $P(\text{rg} < .4)$. In order to answer this question I will apply Bayes Theorem to obtain a posterior distribution for rg. From Equation 3.3.2 we see that $f(\text{poll}|\text{rg})$ and $f(\text{rg})$ need to be specified. The former can be viewed as a binomial distribution with 385 "successes" (votes for the coalition) and $1000 - 385$

"failures". Thus,

$$f(\text{poll}|\text{rg}) = \binom{1000}{385}\text{rg}^{385}(1-\text{rg})^{615} \propto \text{rg}^{385}(1-\text{rg})^{615}$$

To fully specify the Bayesian model the prior distribution $f(\text{rg})$ must be specified. Since rg is a proportion it is natural to choose a distribution that is only defined on the interval $[0, 1]$. The beta distribution is such a distribution. We then get

$$f(\text{rg}|a, b) \propto \text{rg}^{a-1}(1-\text{rg})^{b-1}$$

Specification of the shape parameters a and b reflects my confidence in the prior, and is easy to operationalize because the Beta distribution becomes narrower the larger the values for a and b. Therefore a and b can be set to reflect the number of pseudo-observations I have made. By this I mean that my confidence in the prior, determined by the values of a and b, is as if I had made actual observations. Let's assume that a quick revision of Norwegian political history reveals that no government has ever lost as much as 24% of its voters between two elections, I am therefore fairly confident that this can't be the case barely two months after the election. I decide that my confidence is as if I had sampled 1 000 respondents and obtained a proportion $\text{rg} = 0.478$ equal to the election result. I then set $a = 478$ and $b = 522$. With Binomial likelihood and a Beta prior it is a matter of adding the exponents together to obtaint the posterior.

$$f(\text{rg}|\text{poll}) \propto \text{rg}^{385}(1-\text{rg})^{615}\text{rg}^{477}(1-\text{rg})^{521} = \text{rg}^{862}(1-\text{rg})^{1136}$$

Which is a Beta distribution with $a = 863$ and $b = 1137$. The expectation of the Beta distribution is $E[X] = \frac{a}{a+b}$ so the posterior estimate of the support for the Red-Green coalition is $863/(863 + 1137) = 0.4315$ or 43.2%. With the posterior distribution one can calculate all quantities of interest, for example the probability that rg is less than or equal to 40

$$P(\text{rg} \leq .4) = \frac{\Gamma(863+1137)}{\Gamma(863)\Gamma(1137)}\int_0^{.4} \text{rg}^{862}(1-\text{rg})^{1136}d\text{rg} = 0.00211$$

which means that $P(\text{rg} \leq 0.385)$ is even smaller and I can conclude that it is highly unlikely that the estimate provided by Opinion in December 2009 was particularly good. A quantity, calculated from the posterior distribution, that there will be much talk of in this thesis is the *highest probability density region* (HDR). For example a 95% HDR is the region of values that contains 95% of the posterior probability and also has the characteristic that the density within the region is never lower than that outside (Gelman et al., 2004, 38). [4] Below I obtain the HDRs by finding the relevant percentiles of the posterior distribution, and often an HDR will be equivalent to the classical confidence interval (CI), and I use the two names interchangeably even though I never obtain the HDRs (or CIs) analytically in this thesis.

Bayes Theorem, as I have discussed it so far, is an undisputed mathematical fact of probability theory. Where Bayesian statistics differs from classical frequentist statistics is in allowing the prior probability distribution $f(\theta)$ in Equation 3.3.1 to be subjectively specified, as I did in the polling example above. But it is exactly by allowing for subjective priors that the Bayesian approach lends itself naturally to the problem that this thesis aims to tackle, that of locating the level of a latent variable. Since political polls are noisy and can potentially be plagued by bias one needs a way of weighting the information conveyed by the polls. In the context of using political polls to determine the level of support for a given party, the specification of subjectice priors economize the use of data. We know a whole lot about Norwegian politics, and this knowledge should not be wasted when observing polls. In fact, very few people with an interest for politics forget what they know about a political party or a political system when they observe a poll result. A poll result is always interpreted as more or less likely to be a good estimate, given other things we know. Commentators might say that a given poll result indicates that a party is struggling, but probably not as much as the poll result suggests, and so on. Bayes Theorem is the formula for how one should combine these two pieces of information to new (posterior) beliefs about the state of the world. In the polling example above Bayes Theorem was used to determine the level of support for the Red-Green coalition. Until observing the poll published by Opinion I believed that the coalition was enjoying the support of 47.8% of the electorate. After having observed the poll which gave the coalition 38.5% of the vote intentions, my updated belief was that 43.2% of the electorate supported the coalition. In the next chapter I introduce the Dynamic Linear Model, which is an extension of the type of model used in this example to the case where party support is not static, but evolving over time.

[4] More precisely, an $100(1-\alpha)\%$ HDR for a parameter θ is a region A in the probability space for the parameters where $P(\theta \in A) = 1 - \alpha$ and $P(\theta_1) \geq P(\theta_2)$ for all $\theta_1 \in A$ and $\theta_2 \notin A$ (Jackman, 2009, 26)

4

The Dynamic Linear Model

In this chapter I introduce the two Dynamic Linear Models (DLMs) I use to track the distribution of support among the Norwegian political parties. In a first part the model most commonly used for this purpose in presidential and two-bloc/party systems, the Gamma-Normal model, is presented, and I show why this model is problematic to use in the case of a multiparty system such as the Norwegian. Therefore, in a second part I develop a model that is theoretically sound when tracking party support in a multiparty system. This is the Dirichlet-Multinomial model. I use the two models to estimate party support for the three previous parliamentary periods, as well as the current. The results are displayed graphically and the two models are compared. Lastly, I present the estimated house effects and discuss their potential use in forecasting of election outcomes. For the sake of forecasting, the DLMs provide me with a estimates that can be used to determine the effects of political events and economic conditions on party support, as well as a clear description of how the support for individual parties is evolving over time. All this is valuable information when trying to predict the future course of party support, and ultimately the outcome of elections.

Dynamic Linear Models (DLMs) are Bayesian models well suited for modelling the dynamics of a latent quantity for which there only are imprecise measurements. As discussed in Section 1.1, support for political parties is such a quantity. The actual distribution of support is only observable on election day, while in between elections political polls provide us with imprecise measurements of what the distribution and level of support might be. This is a setting that lends itself naturally to modelling with a DLM. In effect, a DLM tackles the three problems elaborated on in Section 3.1 on the problems with political polls (Jackman, 2005, 508). First, with a DLM I will be able to use all the polling data I have available, and thereby take advantage of all the information there is and increase the precision of my estimates. In other words, the DLM provides a solution to the potential problems caused by pooling the polls. Second, the DLM, as I formulate it below, makes estimation of the house effects possible. Given that these estimates tell a convincing story, I can adjust for these in subsequent analysis and when updating an election forecast. The third point is of more general interest, and is not that important for this thesis. Nevertheless, it is worth mentioning that the DLM produces estimates of the support for each party at each point in time (weeks and months in my analysis) so that questions concerning the effect of a given event can be meaningfully answered. In addition, the DLM produces an estimate of the variance of the latent state. Generally, DLMs are methods for modelling of a time series $\{y_t\} = y_1, y_2, ..., y_t$ that may be a scalar or a vector quantity. In the Gamma-Normal model I present below, the time series is a scalar quantity. This means that the time series takes the form

$$0.32, 0.29, 0.28, \ldots$$

In the Dirichlet-Multinomial model I develop for multiparty systems I will work with a time series that is a series of vector observations. A time series of vector observations looks like this[1]

$$\begin{bmatrix} 320 \\ \vdots \\ 50 \end{bmatrix}, \begin{bmatrix} 290 \\ \vdots \\ 60 \end{bmatrix}, \begin{bmatrix} 280 \\ \vdots \\ 40 \end{bmatrix} \cdots$$

A DLM consists of three components (West and Harrison, 1997, 102). First, an *observation* equation describing how the observed time series $\{y_t\}$ is generated by an unobserved times series $\{\alpha_t\}$ of latent states. Second, a *transition* equation describing how this latent state evolves over time. And third, the *initial information* (the prior as discussed in Section 3.3) through which I specify my prior knowledge about $\{\alpha_t\}$. In the two sections that follow I introduce two DLMs that are based on different distributional assumptions. The first is the Gamma-Normal DLM. This model is the most conventional DLM and the one used by Jackman (2005) and Beck et al. (2006) in their papers tracking support for the two political blocs in Australia, and for president Bush in the US. As will become clear, applying the Gamma-Normal model to a multiparty system is theoretically problematic, as one breaks many of the assumptions that underpins this model in so doing. Therefore,

[1]The proportions of the scalar time series and the elements in the vectors are meant to represent the data I will be working with in the two models. A scalar observation might as well be an integer, and the elements of a vector real numbers.

in Section 4.2 I will develop and fit a Dirichlet-Multinomial model to the same data. This model is theoretically sound when modelling party support in multiparty systems.

4.1 The Gamma-Normal model

When using a Gamma-Normal model and polling data to track the support for individual political parties in a multiparty system such as the Norwegian I must make some rather strong assumptions. First, I must regard the data generating process as n independent Bernoulli trials. This means that I regard the questioning of a survey respondent as an experiment with two possible outcomes: the survey respondent either has the intention to vote for a given party or not. Consequently the observed y_t will follow a Binomial distribution. This is a strong assumption, because the response of one survey respondent obviously does not have to fall into one of two categories, but rather one of as many categories as there are political parties (plus a category for others and those who have no intention to vote). Second, a consequence of regarding the data generating process as independent Bernoulli trials is that I am not dealing with one time series $\{\alpha_t\}$, but several independent time series'. This is also problematic because a party's share of support is not independent of the share of support enjoyed by another party, but in the Gamma-Normal framework they are modelled as if they were. This means that the model can generate absurd results, such as the sum of support for all the parties exceeding one.

Nevertheless, the advantages of these assumptions comes from the fact that, under some conditions, the Binomial distribution can be well approximated by the Normal distribution (Devore and Berk, 2007, 184).[2] Thus, these assumptions allow me to work with a Normal distribution that is intuitive and easy to interpret, where specification of models that include house effects is straightforward. Furthermore, under these assumptions I can use a transition equation that is intuitive and produces a posterior at t that is directly applicable as a prior at $t + 1$. Most importantly, the Gamma-Normal model is the workhorse model for the problem at hand, applications using the Dirichlet-Multinomial model to track party support in multiparty systems are rare or non-existent (I have not

[2]Recall that a Binomial random variable $X = \{\text{number of successes among } n \text{ Bernoulli trials}\}$. With pmf $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$, and $E[X] = np$, $\text{Var}(X) = np(1-p)$. When, as here, one wants to make inferences about a population proportion p, the sample number of successes is divided by the sample size, X/n, and $E[X/n] = p$, $\text{Var}(X/n) = p(1-p)/n$. For the normal approximation to the binomial to be tenable, one must have $np \geq 10$ and $n(1-p) \geq 10$, if this is not the case the Binomial distribution will be too skewed for the symmetric normal curve (Devore and Berk, 2007, 185-186). Translated to the problem at hand, the requirements of $np \geq 10$ and $n(1-p) \geq 10$ means that the number x of survey respondents who said the would vote for a given party cannot be too small, and that the sample size n of the survey cannot be too small either.

found any). These advantages will become clearer as I go along.

Under these assumptions the observed data (the polls) are regarded as generated by a Normal process. Let $i = 1, ..., n$ index the polls published on a given point in time $t = 1, ..., T$, and let n_{ti} be the sample size of poll i at time t. Then $y_{ti} \in [0, 1]$ is the observed poll result. The observational model is then

$$y_{ti} = \mu_{ti} + \nu_{ti}, \quad \nu_{ti} \sim N[0, \sigma^2_{ti}] \tag{4.1.1}$$

Where the observational error variance is σ^2_{ti} is estimated from the data and is due to the Normal approximation to the Binomial.

$$\sigma^2_{ti} = \frac{y_{ti}(1 - y_{ti})}{n_{ti}} \tag{4.1.2}$$

The expectation μ_{ti} is modeled as a linear combination of the actual but unobservable support for a given party α_t (the latent state), and an effect δ_j specific to each particular polling institute $j = 1, ..., J$

$$\mu_{ti} = \alpha_t + \delta_j \tag{4.1.3}$$

Ultimately, it is the latent state α_t I want to make inferences about. The transition model I use is a random walk (West and Harrison, 1997, 53), this means that I assume that on average the level of support for a given party is equal to yesterday's level of support for the party (Jackman, 2009, 479). This yields the following transition equation

$$\alpha_t = \alpha_{t-1} + \epsilon_t \quad \epsilon_t \sim N[0, \omega^2] \tag{4.1.4}$$

When the model is estimated the sampling algortihm (briefly described in Appendix B.3) runs through the time series as many times as I specify. For each iteration the random walk is intialized by a value sampled from a uniform distribution. In other words, the time series of the latent state starts at a different place for each iteration

$$\alpha_1 \sim \text{Unif}(l, u)$$

where l and u are party specific points set to bracket the range of plausible election results for the party in question. When it comes to the variance ω^2 I follow what is common in the Bayesian literature (West and Harrison, 1997, 53) and work with the precision $\phi = 1/\omega^2$ instead of the variance directly. The precison ϕ is unknown and I ascribe it a Gamma prior density,

$$f(\phi) = \text{Gamma}[a, b] \tag{4.1.5}$$

where a and b are the shape and rate parameters ($1/b$ is the scale parameter). In Appendix B I provide a more thorough description of the Gamma-Normal model. I discuss the specification of the priors, derive the conditional posterior distributions and show how the model is implemented in JAGS, a software for Bayesian analysis. In the next section I present the Dirichlet-Multinomial model, before I go ahead and discuss the data and the results of the two models in Section 4.3.

4.2 The Dirichlet-Multinomial model

As I showed above in Section 4.1 the use of a Gamma-Normal normal model in the case of a multiparty system rests on some rather strong assumptions. It was primarily three issues that make it problematic to use a model where the underlying assumption is that the data generating process is a binomial process. First, when a survey respondent in a multiparty system is asked about his or her vote intentions it is clearly wrong to regard this as a Bernoulli experiment. Remember that a Bernoulli experiment is what you perform when you flip a coin, and that a coin has two sides. Questioning of a survey respondent in a political system such as the Norwegian, on the other hand, is more like rolling a biased dice. This means that since each survey respondent can fall into as many categories as there are political parties (plus a category for those who can't answer and for those who have no intention to vote) it is more correct to regard the data as being generated by a multinomial process. Second, since the Binomial distribution is not well approximated by the Normal distribution when support for a party is close to its limiting values of 0 (or 1), the approximation might not fit that well for smaller parties in the analysis (Devore and Berk, 2007, 186). Since there exist parties relevant for the analysis (they might be represented in parliament) that enjoy very little support among the electorate, this is a problem that one has to face when doing this kind of analysis in all multiparty systems. And finally, when modelling support for the parties in a multiparty system one needs a model that is restricted to providing estimates whose sum is always equal to one. The model I now present tackles all these problems, but as I will show, it is not as fit for modelling a quantity that evolves over time as the model above.

I view the questioning of each of the n_{ti} participants in a given survey as a multinomial experiment with $k = 1, \ldots, 8$ possible outcomes. Either, the respondent intends to vote for one of the seven parties currently represented in parliament, if this is not the case, the respondent is put in an eighth category. Consequently, this eighth category consists of those who intend to vote for an eighth party, are not able to answer, or have no intention to vote. Furtermore, I assume that the probability of each of the outcomes is constant across the n_{ti} survey respondents. This means that the probability of interviewing a survey respondent that intends to vote for Venstre is equal across all the interviews (Lid Hjort, 2012, 15). Summing up the number of respondents in each category one obtains a count vector $\mathbf{y}_t = (y_{t,1}, \ldots, y_{t,8})$ where the elements are the number of respondents in each category. With the data I have at hand these count vectors are obtained by multiplying the percentage share in each of the eight categories by the sample size of the poll, and then rounding off to the nearest integer. Furtermore, since I estimate the models on a weekly or a monthly basis, the count vectors appearing in the same week/month are simply added together. This results in an observed time series $\mathbf{y}_1, \ldots, \mathbf{y}_T$ of vector quantities. The observational model is then

$$(y_{t1}, \ldots, y_{t8}) \sim \text{Multi}(n_t, \alpha_{t1}, \ldots, \alpha_{t8}) \tag{4.2.1}$$

where the probability mass function of the Multinomial distribution is given by

$$p(y_{t1}, \ldots, y_{t8}) = \frac{n_t!}{y_{t8}! \cdots y_{t8}!} \alpha_{t1}^{y_{t1}} \cdots \alpha_{t8}^{y_{t8}} \tag{4.2.2}$$

In fact, since the variance of the Multinomial distribution is a function of α_{tk} and n_t Equation 4.2.1 fully specifies the observational model. As in the previous model, it is the parameters α_{tk} giving the latent state of support for the seven parties (plus one) I want to make inferences about. In order to fully specify a Bayesian model I need a prior over these. The conjugate prior distribution for the Multinomial is a multivariate generalization of the Beta distribution (encountered in Section 3.3) known as the Dirichlet distribution (Gelman et al., 2004, 83).[3] A Dirichlet prior over $(\alpha_{t1}, \ldots, \alpha_{t8})$ ensures that the α_{tk} parameters sum to one as wanted. The probability density function of the Dirichlet distribution is given by

$$\pi(\alpha_{t1}, \ldots, \alpha_{t8}) = \frac{\Gamma(b_{t1} + \cdots + b_{t8})}{\Gamma(b_{t1}) \cdots \Gamma(b_{t8})} \alpha_{t1}^{b_{t1}-1} \cdots \alpha_{t8}^{b_{t8}-1} \tag{4.2.3}$$

[3] A prior is conjugate for a likeliood if the posterior distribution belongs to the same class of distributions as the prior (Hoff, 2009, 49). Often this means that the prior and the posterior are the same distributions. In the Beta-binomial model in the example in Section 3.3 the prior and the posterior were both Beta distributions. In the Dirichlet-Multinomial model the prior and the posterior are both Dirichlet.

By Bayes Theorem this gives a posterior distribution for the α_{tk}'s that is also Dirichlet with parameters $b_{tk} + y_{tk}$. Which, as in the Beta-binomial model, is simply the sum of the exponents

$$\pi(\alpha_{t1}, \ldots, \alpha_{t8} | y_{t1}, \ldots, y_{t8}) \sim \text{Diri}(b_{t1} + y_{t1}, \ldots, b_{t8} + y_{t8}) \tag{4.2.4}$$

For each iteration the parameters of the Dirichlet prior are drawn from a uniform distribution

$$b_{tk} \sim \text{Unif}(l, u) \tag{4.2.5}$$

where l and u are set to bracket the range of plausible election results for party $k = 1, \ldots, 8$. To implement this model as a DLM is a little more challenging because the model does not lend itself as easily to a dynamic setting as the Gamma-Normal model. The challenge consists of deciding how to deal with the fact that the samples from the Dirichlet posterior are proportions $\alpha_{tk} \in [0, 1]$ while the parameters used to specify the Dirichlet prior, the b_{tk}'s, are integers. How I deal with this issue and how the Dirichlet-Multinomial model is implemented as a DLM is shown in Appendix C together with the Python program I have written to run the model.

4.3 Real-time tracking of party support

The DLM of ultimate interest in this thesis is the model that tracks the evolution of party support from the election of 2009 until present on a weekly basis. As new polls are published, the model is continuously updated. Before I go on to discuss the results of this model, I will take a closer look at the Gamma-Normal and Dirichlet-Multinomial models estimated for the three parliamentary periods of 1997-2001, 2001-2005 and 2005-2009, and presented in a series of graphs in Figures A.1, A.2 and A.3 in Appendix A. The data used to estimate the Gamma-Normal and the Dirichlet-Multinomial models for the three previous parliamentary periods are polling data collected by Professor Aardal.[4] For the periods 2001-2005 and 2005-2009 I have data from five polling institutes and from four institutes for the 1997-2001 period.[5] Compared to the data I have for the period from 2009 to the present, there are two weaknesses with these polling data: the data are averaged for each of the polling institutes on a monthly basis, and no sample sizes are reported. To deal with the latter I have simply imputed the average sample sizes for the

[4]The data was kindly provided to me by Aardal via my supervisor.

[5]These are Synovate, Gallup, Opinion, Nielsen and Sentio. Minus Sentio for the oldest parliamentary period.

five institutes for the period from 2009 to the present. Since the number of polls conducted by a institute a given month is unknown to me, I have, as a conservative measure, imputed the sample size of one poll only. This measure assures that the sample sizes are unlikely to be exaggerated. Due to the fact that the data are monthly averages, the actual fields dates (the days during which the interviewing took place) are unknown. Since a monthly average is necessarily calculated at the end of each month, I have experimented with model specifications where I date the polls to the last day of each month, and estimate the models on a weekly basis. This means that the model is only fed data on every fourth point in time. Such sparse data availability does not need to be a problem, as a DLM is well suited to handling missing data points since inference for α_t can proceed with the transition model serving as the posterior (Jackman, 2009, 474). For the Gamma-Normal model this sparsity of data does, however, result in an estimated time series $\{\alpha_t\}$ that is overly noisy, and causes the confidence intervals to jump up and down. I have therefore chosen to estimate the Gamma-Normal model on a monthly basis for the three previous parliamentary periods, and in order to make the models comparable the Dirichlet-Multinomial model is also estimated on a monthly basis.

To reiterate, the unknown parameters estimated in the Gamma-Normal model are for each time series 47 α_t parameters giving the monthly levels of support for the party in question for each month between two elections (12×4 minus the month the election took place), the five or four house specific effects δ_j, and the precision parameter $\phi = 1/\omega^2$ characterizing the week-to-week volatility of α_t. Inference for these parameters proceed by using the Gibbs sampler in JAGS to sample repeatedly from the posterior densities of the model parameters. For the Dirichlet-Multinomial model the unknown parameters are only the 47 α_t parameters, and inference for these parameters is obtained by sampling from the posterior Dirichlet distribution for each point in time. In Appendices B and C I provide more details on how this is done.

The results of these two models estimated for the three previous parliamentary periods are presented in a series of graphs in Figures A.1, A.2 and A.3 in Appendix A. The graphs show the estimated latent levels of support for the seven parties currently represented in parliament. Each plotted point is the mean of the values sampled from the posterior density of α_t. The shaded regions are the 2.5 and 97.5 percentiles of the posterior distribution for each α_t (the 95% HDRs). All the models for the previous parliamentary periods are fit subject to the constraint that on the last day α_T has to equal the actual election outcome. Visual inspection of the graphs reveals that the most striking thing is the resemblance of the estimated times series. For the Gamma-Normal model this is reassuring, because it indicates that even though some fundamental assumptions are broken when using a Gamma-Normal model on the multiparty Norwegian system, the results are sensible. A more formal inspection of the time series that I conducted in **R** shows that the sum of the estimated means α_t never exceeds one. There are, however, individual samples from the posterior distribution that do. The resemblance of the two models does not indicate that the Gamma-Normal model will work for all multiparty systems, neither does it mean that it is a good model for the Norwegian system. It only means that the results seem plausible for the Norwegian case for these three parliamentary periods. There are probably two reasons for this. First, that none of the seven parties at no point in time during the three parliamentary periods experienced poll results that were so close to zero that the Normal approximation to the Binomial breaks down. Second, the relative agreement between the polls and the fact that they all have sample sizes of about the same size, hinders the sum of the estimated means from exceeding one when the model is estimated at a monthly basis.

Of more substantive interest the graphs show clearly that the Norwegian party system has been moving towards a system with three dominating parties, Ap, H and Frp, and that Ap has become less dominating relative to the two latter. Another interesting feature of the graphs is that the time series of H and Frp are close to being mirror images of each other, when Frp goes up H dips down and opposite. Another pertinent feature of the three graphs is the volatility of the estimated time series of the three big parties relative to the four smaller parties. For Frp the volatility of the time series are particularly pronounced. According to Jupskås (2011) the reason for this volatility is the protest motives of the Frp-voters and their lack of strong party identification. This visual impression is confirmed by the estimated standard deviation $\sqrt{\omega^2}$ of the three time series for Frp, which at 0.017, 0.017 and 0.021 are much higher than the estimated standard deviations for the time series of the other parties (except Ap in 1997-2001 which has $\sqrt{\omega^2} = 0.02$). In addition, few others parties seem to experience the same sudden spikes in support as Frp. My prime focus here will therefore be in this party. Jupskås (2011) suggests that these spikes are caused by particular events and issues that dominated the public's attention for a certain period of time. Chronologically, the apparent immense rise in support for Frp in September 2000 was most probably caused by a prolonged media focus on gasoline duties and taxes in general, Norway was marked by what could resemble an anti-tax protest movement (Jupskås, 2011). In the months leading up to the election of 2001 it was, however, the other party of the right, Høyre (H) that managed to take advantage of the discontent when Frp became plagued by internal antagonism that weakened the party. In Figure A.3 we see that the prime loser of the general discontent was Ap, whose veritable downturn coincides with the start of the media focus on gas prices and taxes. The sudden spike in support for Frp around october 2002, visible in Figure A.2 was most probably the product of an unpopular budget and the attention that asylum seekers were given in the media. In March and April 2006 the publication of caricatures of the Profet Muhammed marked the political debate and received a lot of attention in the media. Figure A.1 indicates that this worked to the benefit of Frp. The sudden dip in the support for Frp in October 2007 is often explained as a consequence of that climate change had been put to the forefront of the public's attention, with Al Gore receiving the Nobel Peace Prize for his efforts against climate change the same year. Frp is not known for being a climate friendly party. In August 2008 many cultural personalitites, often associated with an urban leftist elite residing in Oslo, went toghether in a public attack on the culture policy of Frp. This attack lead many voters to rally around Frp (Jupskås, 2011). Another feature that is particularly visible in Figure A.3 and A.1 is the spikes in support for H and the dips in support for Frp in the middle of the parliamentary period. This is the period when the local elections take place, and the spikes and dips of H and Frp are often seen as a consequence of many right leaning voters opting for H in the local elections (Jupskås, 2011). In the time series plots in Figure A.1, A.2 and A.3 I have indicated the events mentioned in this short discussion.

In Figure 4.1 I graph the Gamma-Normal and Dirichlet-Multinomial models for the current parliamentary period. The first poll in this period dates from November 5. 2009, while the most recent poll was published on May 11. 2013. This makes for a total of 406 polls. Without exception the data used to update the model are drawn from the site `pollofpolls.no`. On this site, which was created in 2009, one can find all the political polls published in Norway, including meta data such as sample sizes, field dates and method of interviewing. There are eight polling institutes that conduct national polls. These are Gallup, InFact, Ipsos, Norfakta, Norstat, Opinion, Respons and Sentio. I estimate both the Gamma-Normal and the Dirichlet-Multinomial model on a weekly basis, where a poll is labeled as having been conducted in the middle of its field period, i.e. if the interviewing for a poll took place from May 2. to May 8. 2013, I date the poll to May 5., which is week 190 in the period.

To get a sense of where the DLM estimates are relative to the individual polls I have overlayed the point estimates of the respective polling houses for Ap and Frp. From these time series plots the house effects are clearly visible. For example, we see that when it comes to Fremskrittspartiet (Frp) the point estimates of the polling house Sentio consistently lies above the estimated line, and at many points in time far outside the 95% highest density region of the posterior density. The polling house Respons, on the other hand, seems to underestimate Frp. These results are more formally presented in Figure 4.2 in Appendix A, where the exact estimates of the house effects are plotted for each of the parties with the corresponding confidence intervals.

When it comes to the point estimates of the Gamma-Normal and the Dirichlet-Multinomial models they are less alike in Figure 4.1 than in the graphs for the three preceeding parliamentary periods. The reason for this is the higher frequency of polling data, and that differences in the specification of the priors between the two models make them react differently to new observations. The Dirichlet-Multinomial model is less sensible to rapid swings in the estimates provided by the polling houses, that is, the Dirichlet-Multinomial model produces a smoother time-series than the Gamma-Normal. The clearest example of this is how the estimated time series react to the polls published in the wake of the terrorist attacks of July 2011, a period during which Ap experienced some poll results far above what the party had seen prior to the attacks. Some of the polls showed support for Ap above 40 percent.[6] It is clear that the Gamma-Normal model is much more sensitive to these poll results than the Dirichlet-Multinomial model. The priors of the two models are different in the sense that in the Dirichlet-Multinomial model my confidence in the prior is less a function of yesterday than in the Gamma-Normal model, this means that if the Gamma-Normal model traverses a period with no or little polling, it becomes very sensitive to the appereance of polls. And normally, July is a month with very little polling. In the Dirichlet-Multinomial model, on the other hand, the prior weights more heavy and is not as prone to being dominated by the likelihood, even though it has not been fed data for a period of time. Obviously, we have little reason to believe that support for the political parties fluctuates with the frequency that the Gamma-Normal model suggests, which means that the smoother time series of the Dirichlet-Multinomial model seems more plausible. In addition, it is as I have shown, the theoretically correct model. For the forecast functions that I develop in the next chapter I will therefore use the estimates provided by the Dirichlet-Multinomial DLM.

[6]Ipsos 31. July, 41.7%; Gallup 1. August, 40.5%; and Norfakta 6. August, 40.4%

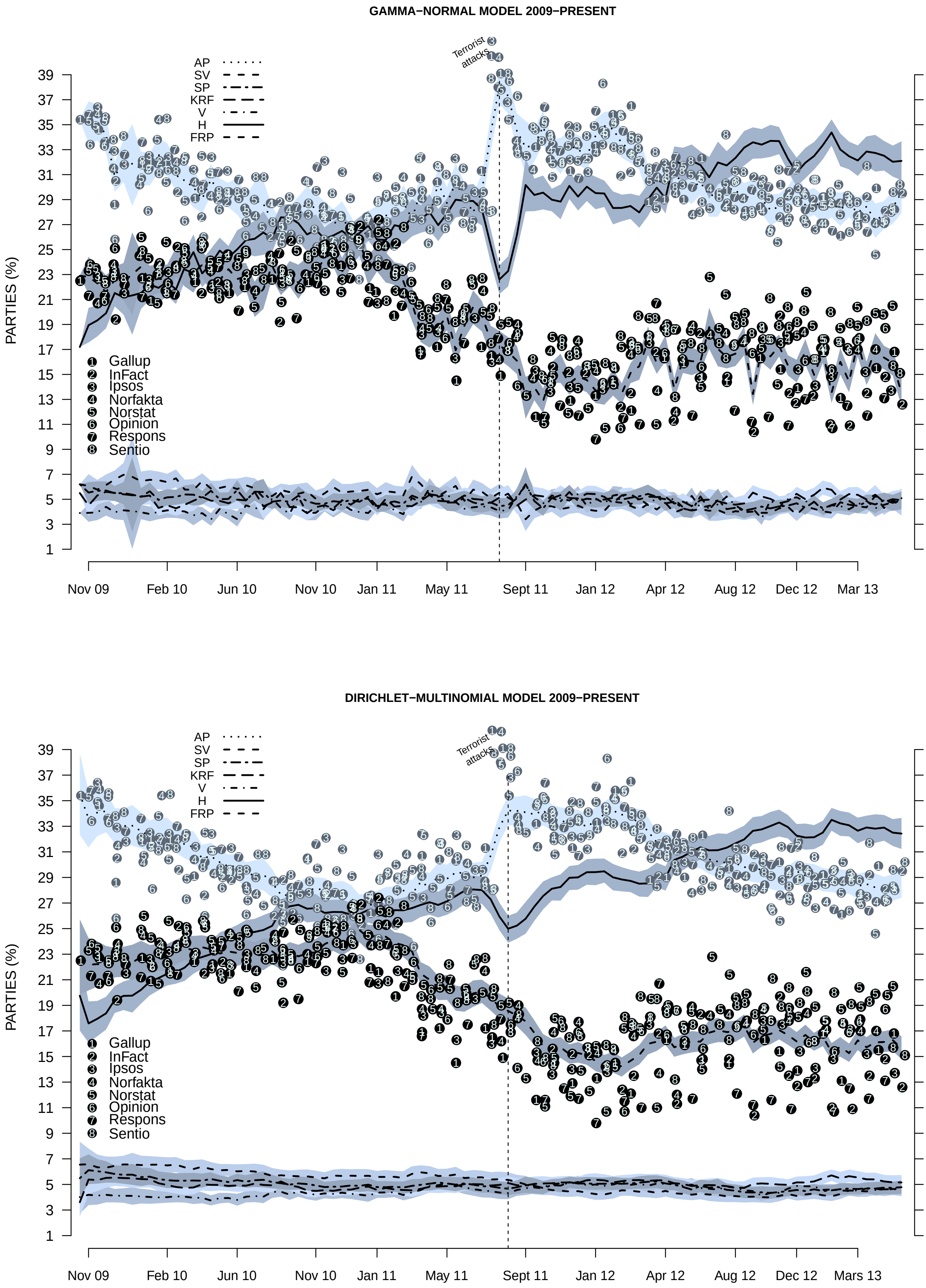


Figure 4.1: Weekly tracking of party support. Shaded areas are 95% HDRs. Individual polls are represented with plotted point at their respective point estimates for Ap and Frp.

4.4 Estimated house effects

In Figures 4.2, A.4, A.5 and A.6 the point estimates and confidence intervals for the house effects are plotted for the present and the three preceeding parliamentary periods. It is important to note that the selection of polling houses is not the same in the model for the current period as in the three preceeding. For the sake of comparison, I will start by discussing the parliamentary periods of 1997-2001, 2001-2005 and 2005-2009. The question of substantive interest here is whether the bias (or lack thereof) of a given polling house points in the same direction and has about the same magnitude for all three periods (in addition to having HDRs that don't overlap zero).

Visual inspection of the figures reveals that there are only a few polling houses for which the answer to this question is unequivocally affirmative. The polling house Opinion consistently underestimates the support for Arbeiderpartiet (Ap). The magnitude of this negative bias ranges from 0.4 percentage points in the 2001-2005 period to almost 2 percentage points in the 2005-2009 parliamentary period. Furthermore, we see that Gallup's estimates for Kristelig folkeparti (Krf) consistently lies about half a percentage point below the estimated line for the three last periods, while it is still negative but cannot be distinguished from zero in the current period. For the other parties and polling houses the biases are more variable, for some periods pointing in a negative direction, and in a positive direction for other periods, with point estimates that are distinguishable from zero in one period, but not in the others, and so on. An interesting pattern should, however, be commented upon. That is the variability of biases when it comes to the polling houses' estimates for Fremskrittspartiet (Frp) and Høyre (H). Even though it is hard to discern a pattern (because there aren't any) as to which houses are biased in which direction, my estimates of the house effects could indicate that the polling houses are having a hard time finding their preferred weighting procedure when it comes to these two parties. An example of a polling house that my estimates indicate as being more often on the mark than not, is Nielsen. The estimates for Nielsen are not distinguishable from zero for any of the parties during the 2001-2005 period. During the period that followed Nielsen did, however, perform poorly in estimating Frp and H, but so did most other houses as well. One thing is clear from these figures, that is that the polling houses are more often off in their estimates for the three big parties Ap, Frp and H, than for the four smaller parties. The estimated house effects for the present parliamentary period are shown in the Figure 4.2, while the remaining three Figures are found in Appendix A.

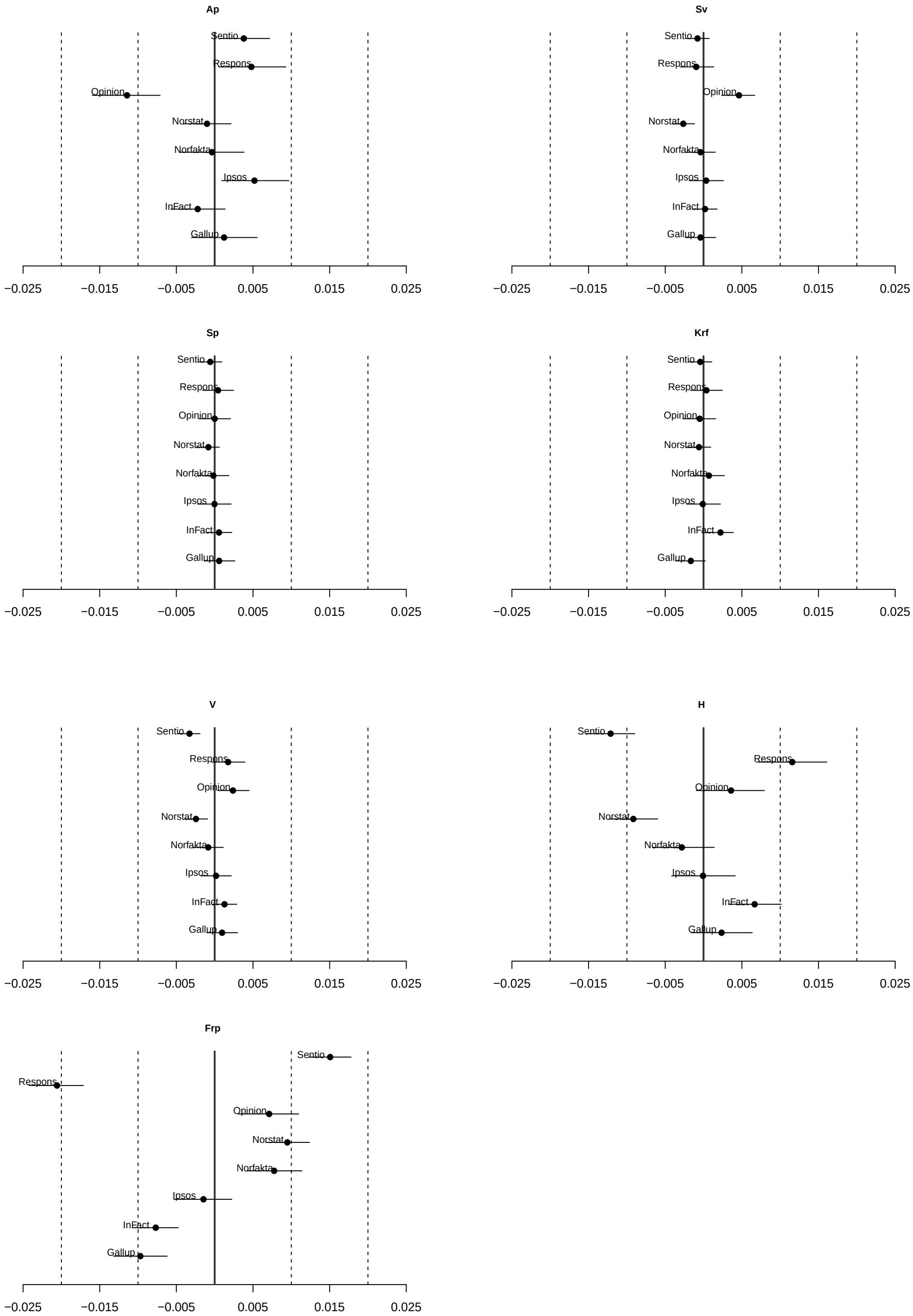


Figure 4.2: House effects 2009-present. The effects are constrained to sum to zero.

4.5 Summary

In this chapter I have presented the two DLMs and compared the estimates that these two models produce for four parliamentary periods. I have explained why the Dirichlet-Multinomial DLM is theoretically more sound, and shown how it produces a smoother and more plausible time series. The results of the DLMs are interesting in their own right because they can be used to answer questions of substantive interest in political science. Particularly, with the Dirichlet-Multinomial DLM I have shown how one can track party support in multiparty systems. In the next chapter I look at the problem of forecasting the outcome of elections. Forecasting elections is in many ways extrapolating the time series of the DLMs into the future, and therefore the development of the time series up until today provides essential information for forecasting their future path. I develop two models for forecasting elections that incorporate the output of the Dirichlet-Multinomial DLM in two different ways, and I use the benefit of hindsight to evaluate their performance.

5

Forecasting the distribution of seats

This chapter proceeds in three steps leading up to the forecasts for the election of 2013. First, since I am forecasting elections in a multiparty system where it is ultimately the number of seats in parliament that counts, I provide an introduction to the Norwegian electoral system in Section 5.1. Second, in Section 5.2 I present my strategy for translating the national vote shares of the parties to the actual distribution of seats in parliament. Third, the methods I develop for predicting elections are presented in Section 5.3 and their performance is evaluated in Section 5.4. Finally, in Section 5.5 I present my forecast for the election of 2013.

5.1 The electoral system

Below I develop two methods for predicting the outcome of elections in multiparty systems. What these two methods have in common is that they provide estimates of the national vote shares of each of the seven parties currently in parliament (plus an eighth category). This is because they are based on estimates from the DLMs, and these estimates are, as we saw, produced by national level polling data. The number of seats a party obtains in parliament is only partly based on the national vote share of the party, I must therefore find a way of translating the national level forecasts to actual seats in parliament. Before I consider how to develop such a method, I explain how the Norwegian electoral system works.

The Norwegian electoral system is a proportional system at the level of the counties, where each of the counties are represented according to a weighting scheme involving the number of inhabitants and the size (square meters) of the county. Norway consists of 19 counties, and these 19 counties are also the electoral districts. The parliament (Stortinget)

	H	A	Frp	Sv	Sp	Krf
Votes	81 140	80 862	39 851	26 295	12 187	11 229
1.4	57 957 (1)	57 758 (2)	28 465 (3)	18 782 (6)	8 705	8 020
3	27 046 (4)	26 954 (5)	13 283 (9)	8 765	4 062	3 743
5	16 228 (7)	16 172 (8)	7 970			
7	11 591 (10)	11 551 (11)				
9	9 015	8 984				
Total seats	4	4	2	1	0	0

Table 5.1: Example of modified version of Sainte-Laguës method

consists of 169 members of parliament, of whom 150 are distributed at the level of the county, while the remaining 19 are so called equalization mandates (In Norwegian: utjevningsmandater), where each county has one equalization mandate. The delegation of seats in parliament proceeds as follows. First, a modified version of Sainte-Laguës method is used in each of the 19 counties. This method consists for calculating successive quotients

$$\frac{v}{2s+1}$$

where v is the total number of votes received by a party from the county, and s is the number of seats delegated to the party so far in the process. The modified version used in Norway consists of setting the first denominator to 1.4 thereby giving an advantage to bigger parties. Table 5.1 shows a hypothetical example with eleven seats being delegated We see that Høyre received 81 140 votes in this county, and that their first quotient is the largest. The ranking of the quotients is shown in parentheses. The second largest quotient belongs to Ap who gets the second mandate, and so on until the eleven mandates are distributed. When this first part of the process is done, 150 mandates have been distributed. The second step is then the calculation of the equalization mandates. Only the parties who have received 4% or more of the national votes compete for these. The modified version of Sainte-Laguës method is used, but this time considering the entire nation as one electoral district. In this second stage the number of mandates distributed equals 169 minus the mandates obtained from the counties by the parties below the 4% threshold. Each party is accorded as many equalization mandates as the difference between what they would have obtained considering the nation as one electoral district and what they have obtained from the counties. If one finds that some of the differences are negative, i.e. that a party has obtained more mandates from the provinces than it would have obtained nationally, all the county-mandates of the party or parties in question are subtracted from the initial mandates (the 169 minus the mandates won by the parties below 4%), and Sainte-Laguës method is applied once over with the remaining mandates and the parties with positive differences in the first iteration of Saint-Laguës method. This process continues until all the differences are positive, and these positive differences are the number of equalization mandates the parties get.[1] The third step consists of deciding from which counties the parties get their equlization mandates. To decide this one ranks the following quotients

$$\frac{\text{votes}}{(\text{mandates} \times 2 + 1)\text{avg.votes}}$$

where the nominator is the number of votes the party has received in the given county, while in the denominator one finds the number of mandates that the party has obtained in the county and the average number of votes that is behind each mandate in the county. Finally, the first equalization mandate is delegated to the party and the county with the largest quotient, the second to the party and county with the second largest quotient and so on. One stops considering the quotients of a party when the party has been delegated the number of equalization mandates that it is due.

5.2 From votes to seats

For each of the three forecast functions developed below I employ the same strategy for mapping votes to seats. I simply use Saint-Laguës method. But in order to do this I need an estimate of the actual number of votes that each party receives in each of the 19 counties. My approach is based on studying the county-wise deviance from the national share of votes for each party in the 2009 election. I assume that this deviance changes very slowly over time, so that the national vote share of a party is a good predictor of its share of the votes in a given county. For example, in 2009 Sosialistisk Venstreparti (Sv) obtained 6.2% of the votes nationally and 5.7% percent of the votes in the county of Akershus. The deviance, as I calculate it here, is then $5.7/6.2 = 0.92$. With a forecast for Sv at 6.4% nationally I predict that Sv receives $6.4 \times 0.92 = 5.9$ percent of the votes in Akershus. Furthermore, in 2009 there were 375 622 people eligible to vote in Akershus, of whom 302 021 turned out on election day. I assume that the turnout will be about the same in 2013 in each of the counties. The forecast is then that $0.059 \times 375\ 622 = 22\ 162$ voters cast their ballots for Sv in Akershus.[2] This procedure is applied to all the counties

[1] I thank Professor Aanund Hylland for clearing up this point.

[2] This last part superfluous as it does not change the ordering of the quotients to multiply them by the turnout. Nevertheless, it only costs me one line of coding and it is nice to look at something close to

and all the parties. With these figures I can proceed algorithmically with Sainte-Laguës method to obtain the predicted distribution of seats in parliament and the number of equalization mandates obtained by each party. Sainte-Laguës method is a deterministic method, so the uncertainty associated with the number of seats obtained by a given party results from providing Sainte-Laguës method with several samples from the distributions produced by the forecast functions. In Appendix F I have attached the Python-program that applies Sainte-Laguës method to calculate the distribution of seats in parliament.

5.3 The forecast functions

Even though they are often spoken of as equivalent, a poll is not the same as a forecast (Lock and Gelman, 2010, 347). Neither are the estimates produced by the DLMs forecasts, they are real-time tracking of party support. The utility of the estimates from the DLMs are that they provide precise and (almost) continuous estimates of the level of support for the different political parties, which can be crucial information when making a forecast. In this section I develop two different forecast functions and use the benefit of hindsight to test the performance of these functions on past elections. Since it is in May that I will make my first forecast for the election of 2013, all of the ex post forecasts are conducted in May 2001, 2005 and 2009. The first function I develop is based solely on the output from the DLM, while the second combines the output from the DLM with the estimated effects of political and economic variables that affect the parties differently.

The most common way of forecasting election results is by extrapolating the distribution of support among the political parties today and into the future (see Hanretty (2013) for a recent example from Italy). For example, if my estimate for Sp today is at seven percent, my forecast is that Sp will receive seven percent of the votes on election day. This is not necessarily a bad strategy, and intutively it should be a better strategy the closer one is to election day. I use this linear forecast as the benchmark for my two other forecast functions. If they are any good they should be able to outperform the simple linear extrapolation of a time series.

The first forecast function is based purely on the output of the Dirichlet-Multinomial DLM and is founded on a very simple idea: namely that there is an underlying trend in the support for the parties that it is possible to capture by fitting a second order polynomial

---

the actual numbers instead of very small numbers.

through the points estimated by the DLM up until the date the forecast is made.

$$\alpha_t = \gamma_0 + \gamma_1 t + \gamma_3 t^2 \tag{5.3.1}$$

Considering the time series plots in Appendix A we see that many of the time series start trending downward or upward several months prior to election day, and that this trend often continuous until the election. The second order polynomial in Equation 5.3.1 is fit through the estimated points from the beginning of the parliamentary period until the day the forecast is made, and the coefficients γ_0, γ_1 and γ_2 are in turn used to extrapolate α_t forward to election day. A problem with this approach is that I am not guaranteed that the eight α_t's sum to one, as they must. But since they are all generated by the points of a Dirichlet distribution, I assume that they individually are Beta distributed, as the Beta distribution is the marginal density of the Dirichlet (Albert and Denis, 2012). With the estimates of the eight α_T's at hand, I view these as the means of eight Beta distributed variables and draw 10 000 samples from the eight Beta distributions. Finally, I divide each draw by the sum of the eight Beta distribution. The resulting vector of means then follows the Dirichlet distribution (Albert and Denis, 2012). In Figure 5.1 below and Figure A.7 and A.8 in Appendix A, the forecast distributions produced by extrapolating the quadratic function four months ahead in time are presented, with the vertical lines indicating the actual election outcome. Below I evalute the performance of this function and compare it to the performance of the second method of forecasting that I describe next.

The second forecast function I construct is inspired by the work of Linzer (2012). It consists of two pieces: the estimated effects of political and economic variables on the support for different parties and the latest output of the Dirichlet-Multinomial DLM. In a first stage, I use the estimated effects of political and economic variables on the support for the different parties to produce a prior forecast distribution for each party. Then, in a second stage the prior forecast distributions are multiplied by the distributions of the latest estimates of the latent level of support for each of the parties, as supplied by the Dirichlet-Multinomial DLM. The idea is that the prior forecasts that I estimate in the first stage give the election outcome one would expect if one considered political and economic variables only: a "natural" outcome given the state of the economy, plus possible incumbency effects. However, with the output of the Dirichlet-Multinomial DLM I have estimates of where party support is today, and I use this information to discipline the pure economic forecast. The product of the two, the posterior, gives the final forecast.

This forecast function draws its theoretical underpinnings from the clientele hypothesis of the economic voting literature (Carlsen, 2000; Hibbs, 1977; Swank, 1993), that I discussed in Section 2.1. According to this hypothesis we should expect varying economic times to have different effects on political parties on opposite ends of the political spectrum. High unemployment rates and low economic growth makes people fear for their jobs and flock around the parties on the left. An anecdotal example that accords well with this hypothesis is the re-election of the Red-Green coalition in 2009. Many analysts attributed the victory of this center-left coalition to the fear of the possible economic consequences of the financial crises that had struck Europe one year prior to the election. In more prosperous economic times, on the other hand, the hypothesis postulates that the parties of the right should gain support. In his study of the US, Canada, the UK and Australia Carlsen (2000) finds evidence that supports this hypothesis, namely that higher levels of unemployment hurts the parties of the right. The findings of Arnesen in the study discussed in Section 2.1 (where he divides the Norwegian political parties into two blocs) also point in the same direction (2012b, 13). By a regression analysis of the national and local elections in Norway, with political and economic independent variables Arnesen identifies a pattern consistent with the clientele hypothesis: ”when unemployment goes up, so does the popularity of the left. Conversly, when the economy prospers, the voters tend to support non-left parties to a higher degree than when growth is slower” (2012b, 13).

Most of the studies investigating this hypothesis has been conducted on presidential or two-bloc/party political systems. This means that the dependent variable either has been a proportion far from its limiting values of 0 and 1, in which case a normal model works fine, or dichotomous indicator (i.e. left = 0, right = 1), in which case a generalized linear model with a logit link is a common option. Since I am working with a multiparty system I want to have estimates of the effects of political and economic variables for all seven political parties that I am considering, I therefore use a model for multinomial responses with a logistic link, a Multinomial logit regression model. To estimate these effects I use as my dependent variable the election results of the parliamentary elections from 1973 to 2009, which makes for 10 observations. I start in 1973 because this is the first election where all the seven parties currently represented in parliament ran for election. Instead of using the proportions of the votes obtained by each party in these elections, I take the proportions and multiply them by an arbitrary high number to obtain a count vector that follows the multinomial distribution.

$$\mathbf{y}_t \sim \text{Multi}(n_t, \alpha_{t1}, \alpha_{t2}, \dots, \alpha_{t8}) \tag{5.3.2}$$

where the α_{tk}'s represents the probability of each of the $k = 1, \ldots, 8$ categories for the $t = 1, \ldots, 10$ observations. I follow Gelman et al. (2004) and parametrize the model in terms of the logarithm of the ratio of the probability of each category relative to a baseline category. The categories are the seven parties plus an eighth category, where I have set Ap as the baseline. I label the baseline as category $k = 1$, the link is therefore

$$\ln\left(\frac{\alpha_{tk}}{\alpha_{t1}}\right) = \eta_{tk} = (\mathbf{X}\boldsymbol{\beta}_k)_t \tag{5.3.3}$$

where $\boldsymbol{\beta}_k$ is the vector of parameters for the k'th party. The distribution of the data is then

$$p(\mathbf{y}|\boldsymbol{\beta}) \propto \prod_{t=1}^{10}\left(\frac{e^{\eta_{t1}}}{\sum_{k=1}^{8} e^{\eta_{tk}}}\right)^{y_{t1}} \cdots \left(\frac{e^{\eta_{t8}}}{\sum_{k=1}^{8} e^{\eta_{tk}}}\right)^{y_{t8}} \tag{5.3.4}$$

where $\boldsymbol{\beta}_1$ is constrained to zero. This is the Multinomial logit model that allows me to estimate the effects of structural variables on each of the seven parties. For a complete Bayesian specification of the model I need to set the priors over the $\boldsymbol{\beta}_k$'s. I ascribe normal priors to these

$$\boldsymbol{\beta}_k \sim N[\mathbf{b}_k, \boldsymbol{\Sigma}_k] \tag{5.3.5}$$

where the contents of $\mathbf{b}_k$ and $\boldsymbol{\Sigma}_k$ are specified in Appendix E. Here it suffices to say that the priors are consistent with the clientele hypothesis and the findings of Arnesen (2012b). In accordance with the economic vote literature and the clientele hypothesis the independent variables are annual unemployment figures measured in percentage of the work-eligible population in the election year, GDP growth and inflation (consumer price index growth in percent) in the election year. These are the same variables that Arnesen (2012b) uses in his study.[3] In addition I have included a dummy variable, called "left", indicating whether or not Ap is in government. Even though this variable is not perfect, due to Sp changing from the centre-right to the centre-left, it is meant to capture the colour of the government and possible incumbency effects.

As discussed in Section 2.4 Linzer (2012) uses the forecasts generated by structural (economic vote) models and updates these with the information conveyed by polls as elections day approaches. The model I construct here is based on the same idea. To reiterate, in a first stage I insert the current values of the independent variables in the Multinomial logit model to obtain prior forecasts for each of the parties. The distributions of the prior forecasts are produced by sampling repeatedly from the posterior distributions of the coefficients of the Multinomial logit model. Then, in a second stage the forecasts

[3]The data is from Arnesens web-appendix (2012b), found at `http://folk.uib.no/st03889/`

generated by the Multinomial logit model are updated using the output from the Dirichlet-Multinomial DLM. This means that the Multinomial logit model generates a prior, this prior is multiplied by the likelihood of the estimated latent levels of support today, and the resulting posterior distribution is the final forecast. In the figures where I show the results of this method of forecasting, all the three distributions are displayed.

5.4 Evaluating the forecast functions

In Figure 5.1, A.7 and A.8 I graph the forecast distributions of the Multinomial logit forecast function and the quadratic function.[4] In addition, the plots contain the prior and the likelihood of the Multinomial logit forecast function. Since the likelihood is the output of the Dirichlet-Multinomial DLM the day the forecast is made, the mean of the likelihood is equivalent to a linear forecast, the benchmark type of forecast. By visual inspection of the plots it seems that the ex post forecasts for the elections of 2009, 2005 and 2001 do not provide an unequivocal answer as to which method of forecast is to prefer. When it comes the Multinomial logit forecast function there is one thing that is important to consider when putting the model to the test of out-of-sample ex post forecasts: how much weight should the prior have relative to the likelihood. It is tempting to let the weighting be decided by what makes for the best ex post forecasts.[5] But this would be to overfit the model to each individual election and I would have no guidance as to which weights are the best for future elections. As can be seen from the plots I have decided to give the likelihood and the prior about equal weight, this is reflected in the density of the prior and the likelihood having about equal spread. I will now consider how the two models perform when making ex post forecasts in May the three last election years. In this evaluation there are two things to consider, the first is the extent to which the point estimates for the percentage share of the votes are on or off the mark for each party. The second, and for the purpose of predicting the colour of the next government more important question, is whether the model gets the distribution of seats right, and if the forecasts get the correct majority.

The 2009 *election* led to the re-election of the Red-Green coalition, who obtained 86 seats in parliament. Parliament is made up of a total of 169 seats, so 85 seats are needed to constitute a majority. Would the forecast functions have predicted the re-election of the

[4] By the Multinomial logit forecast function I refer to the forecast function that combines information from the Multinomial logit regression model and the Dirichlet-Multinomial DLM.

[5] The weight accorded to the likelihood is decided by the sum of the parameters of the Dirichlet distribution, while the sum of the elements of the count vector determines the relative weight of the Multinomial prior.

	seats	HDR	actual		seats	HDR	actual
Ap	49	[45,53]	64	Krf	13	[10,16]	10
Sv	21	[17,24]	11	V	6	[1,9]	2
Sp	8	[6,10]	11	H	29	[26,33]	30
				Frp	35	[32,39]	41
Total	78	[73,83]	86		84	[78,88]	83

Table 5.2: Multinomial logit forecast function. Ex post seat predictions for the 2009 election. 90% HDRs in parantheses.

	seats	HDR	actual		seats	HDR	actual
Ap	61	[55,68]	64	Krf	9	[7,11]	10
Sv	12	[10,16]	11	V	9	[8,11]	2
Sp	7	[1,9]	11	H	23	[19,27]	30
				Frp	47	[41,51]	41
Total	80	[75,87]	86		88	[81,93]	83

Table 5.3: Quadratic function. Ex post seat predictions for the 2009 election. 90% HDRs in parantheses.

coalition four months prior to the election of 2009? The largest party of the Red-Green coalition is Ap, and from Figure 5.1 below we see that the quadratic function taps the upward trend of Ap that starts during the summer of 2008 to outperform a linear forecast. At 33.6% the quadratic function is not far below the actual result of 34.5% for this party. The Multinomial logit forecast function, on the other, underpredicts Ap with the mean of the posterior at 26.4%. From the plot we see that this underprediction is caused by the prior that puts Ap just above 20%. The unemployment rate of 2009 was at 2.7% a little below the average unemployment rate of the nine election years the Multinomial logit regression model is estimated on. A relatively high unemployment rate of 5.5%, would have produced a prediction for Ap at 30%. For Sv and Krf we see that the quadratic function does a better job than the Multinomial logit model, while for the four remaining parties Sp, V, H and Frp the Multinomial logit model is closer. In Table 5.4 we see that the Multinomial logit forecast function would in May 2009 not have predicted a Red-Green re-election four months later, underpredicting the total number of seats by eight with only 2% of the samples falling above the majority mark. For the opposition parties of Krf, V, H and Frp, on the other hand, the Multinomial logit model is at 84 just one seat off, but has 40% of the samples above the majority mark for the opposition parties. Neither is the

	seats	HDR	actual		seats	HDR	actual
Ap	57	[53,59]	61	Krf	15	[13,17]	11
Sv	24	[21,26]	15	V	1	[0,2]	10
Sp	13	[11,16]	11	H	30	[28,32]	23
				Frp	28	[25,31]	38
Total	94	[90,98]	87		74	[70,77]	82

Table 5.4: Multinomial logit forecast function. Ex post seat predictions for the 2005 election. 90% HDRs in parantheses.

	seats	HDR	actual		seats	HDR	actual
Ap	58	[53,63]	61	Krf	15	[12,18]	11
Sv	23	[20,27]	15	V	2	[0,7]	10
Sp	10	[8,13]	11	H	35	[32,39]	23
				Frp	24	[20,28]	38
Total	92	[86,97]	87		77	[71,82]	82

Table 5.5: Quadratic function. Ex post seat predictions for the 2005 election. 90% HDRs in parantheses.

quadratic function capable of predicting the re-election of the coalition four months prior to the election. Instead it overpredicts the total number of seats of the oppostion by six, with 74% of the samples falling above the majority mark of 85 seats for the oppostion. This means that the qudratic forecast function had in May 2009 ascribed a 0.74 probability to an alternation in government (assuming that the opposition had figured things out between them), while the Multinomial logit forecast function had ascribed a proabability of 0.40 to the same event. The ex post predictions for *the* 2005 *election* are graphed in Figure A.7 in Appendix A and the results for the distribution of seats in parliament are presented in Table 5.4 and 5.5.[6] Again it is not easy to discern a clear pattern as to which function does the best job. Both forecast functions miss badly on H and Frp due to the sudden leap of Frp right before the election, and the corresponding dip for H, but perform decent when it comes to predicting the total number of seats that these two parties obtained. The actual number was 61, with the quadratic and Multinomial logit predicitons at 59 and 58 seats respectively. When it comes to the three partners of the Red-Green coalition the quadratic function predicts that these would obtain a total of 91

[6] In the tables presented in the following pages, round-off errors will in some instances prevent the predicted number of seats for the parties of the Red-Green coalition or the opposition to add up to their predicted total.

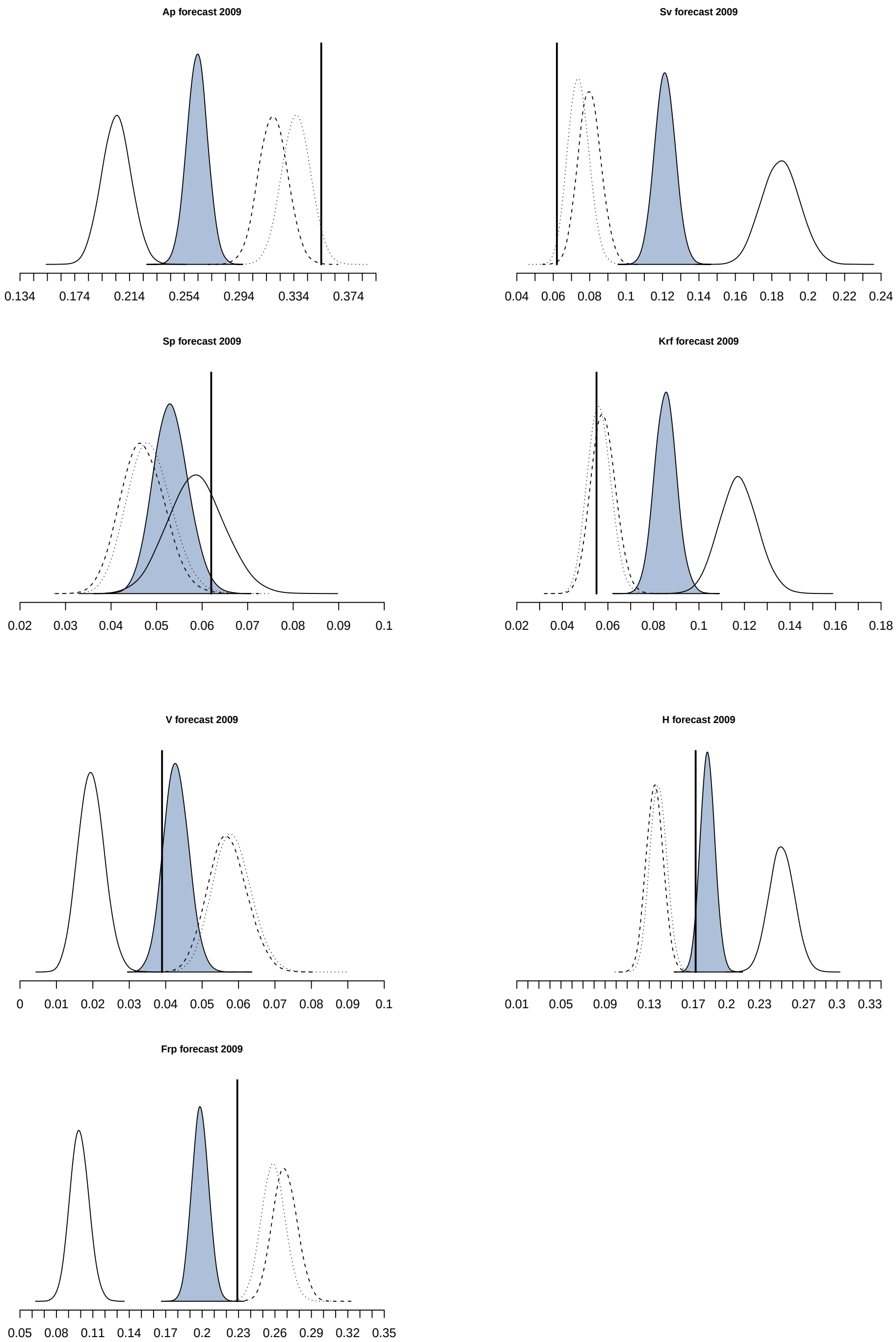


Figure 5.1: Forecasts for the 2009 election. The filled density is the posterior of the Multinomial logit forecast model, the solid line represents the likelihood (the mean of which is equivalent to a linear forecast) and the dashed line is the prior. The line consisting of small points is the forecast distribution produced by the quadratic function.

seats contra the 87 they actually did obtain. Furhermore, 93% of the samples fall above the majority mark of 85 seats for these three parties. This means that the quadratic function ascribed a probability of 0.93 to a Red-Green coalition victory four months prior to the election of 2005. Considering the predictions for the individual parties produced by the Multinomial logit forecast function, things are not much better. Sv is predicted at five percentage points above its actual election result of 8.8%. As with the 2009 predictions, H is overpredicted and Frp underpredicted compared to the actual election results of the two parties. Nevertheless, when the predictions for the Red-Green coalition and the opposition are compared, the performance of the Multinomial logit forecast is much better. The two major parties of the oppostion, H and Frp, are predicted at a total of 58 seats, three short of the actual result. Because of overprediction for Sv, the Red-Green coaliton is predicted as a sure bet four months in advance of the election of 2005, with 99% of the samples falling above the majoriy mark of 85.

The 2001 *election* was for Ap, with 24.3% of the votes, the worst election in the post-war period. In Figure A.3 in Appendix A the decline of Ap is visible from the end of the summer of 1998 and continues more or less unabrupted until election day in 2001. The trend is only disrupted by the inauguration of the Ap cabinet about a year prior to the election. Given such a pronounced trend it is not surprising that the quadratic function taps this four months prior to the election, and outperform the benchmark linear forecast, as shown in Figure A.8. The Multinomial logit forecast function performs badly for Ap, Sv and Krf, but is right on the mark for the remaining parties. For H and Frp it is interesting to note that the Multinomial logit forecasts outperform both the quadratic function and a linear forecast. A reason for this could be the very noisy time series of the two parties. In both cases it is the prior that drags the posterior for the two parties in the right direction. This result is a an indication of the advantage associated with having a forecast that incorporates information that is unafflicted by the volatility of peoples party preferences in between elections. The 2001 election was the last election before a reform of the electoral system that increased the number of equalization mandates to nineteen and changed the system for distributing these. I have not programmed the old electoral system, and therefore I do not translate the forecasts for the 2001 election to seats in parliament.

There is no clear winner among the two forecast functions. Nevertheless, some points are worth putting forward. First, the quadratic function seems to perform poorly with overly volatile time series, and in these instances it is advantageous to have a forecast that is not based on pure data mining. Second, the Multinomial logit regression model does in most instances produce prior forecasts (those based purely on the political and economic variables) that are sensible, given the state of the economy, but sometimes they are a bit exaggerated compared to the election results. In the next section I perform the forecasts for the election of 2013, and two considerations drive my choice of model for this forecast. The first is that the terrorist attacks during the summer of 2011 caused a sudden jump in the time series of Ap, and dips in the time series of Frp and, yet to a lesser degree, for H. These jumps and dips will disturb the quadratic function from tapping onto any possible underlying trend in the time series. The second is simply that I find the Multinomial logit forecast function more solid and satisfying from a theoretical point of view, and that I see this model as one that can be further developed and applied to other multiparty systems. In addition, the Multinomial logit forecast function is more prone to continous updating as election day approaches. To forecast the 2013 election I will therefore use the Multinomial logit forecast function.

5.5 Predictions

The forecasts for the election of 2013 are made with the Multinomial logit forecast function. In Figure 5.2 the forecast produced by the Multinomial logit forecast model is presented by graphing the posterior distribution that I obtain by applying Bayes Theorem to the output of the Dirichlet-Multinomial DLM with the prior provided by the Multinomial logit regression model. The prior distributions are represented by the solid lines and the likelihoods given by the DLM are graphed with the dashed lines. Finally, the posteriors are the filled densities. Before I go ahead and present the forecast for the election of 2013 there are some features of the plots in Figure 5.2 that it is worth commenting upon. Considering the plots for Ap, Sv and Sp, the three coalition partners of the Red-Green coalition, we see that the priors (solid graphs) are to the right of the posteriors. This means that the prior predicts these parties at a higher level than the posterior, this latter being dragged down by the likelihood. The reason for the priors being at such a high level for the parties of the Red-Green coalition has to do with the current economic conditions, which are the data generating the prior forecast. The estimated coefficients of the Multinomial logit regression model are found in Table E.1 in Appendix E. The unemployment rate is currently at 3.4%, which is 23% higher than the average unemployment rate of the last ten election years (which had an average unemployment rate of

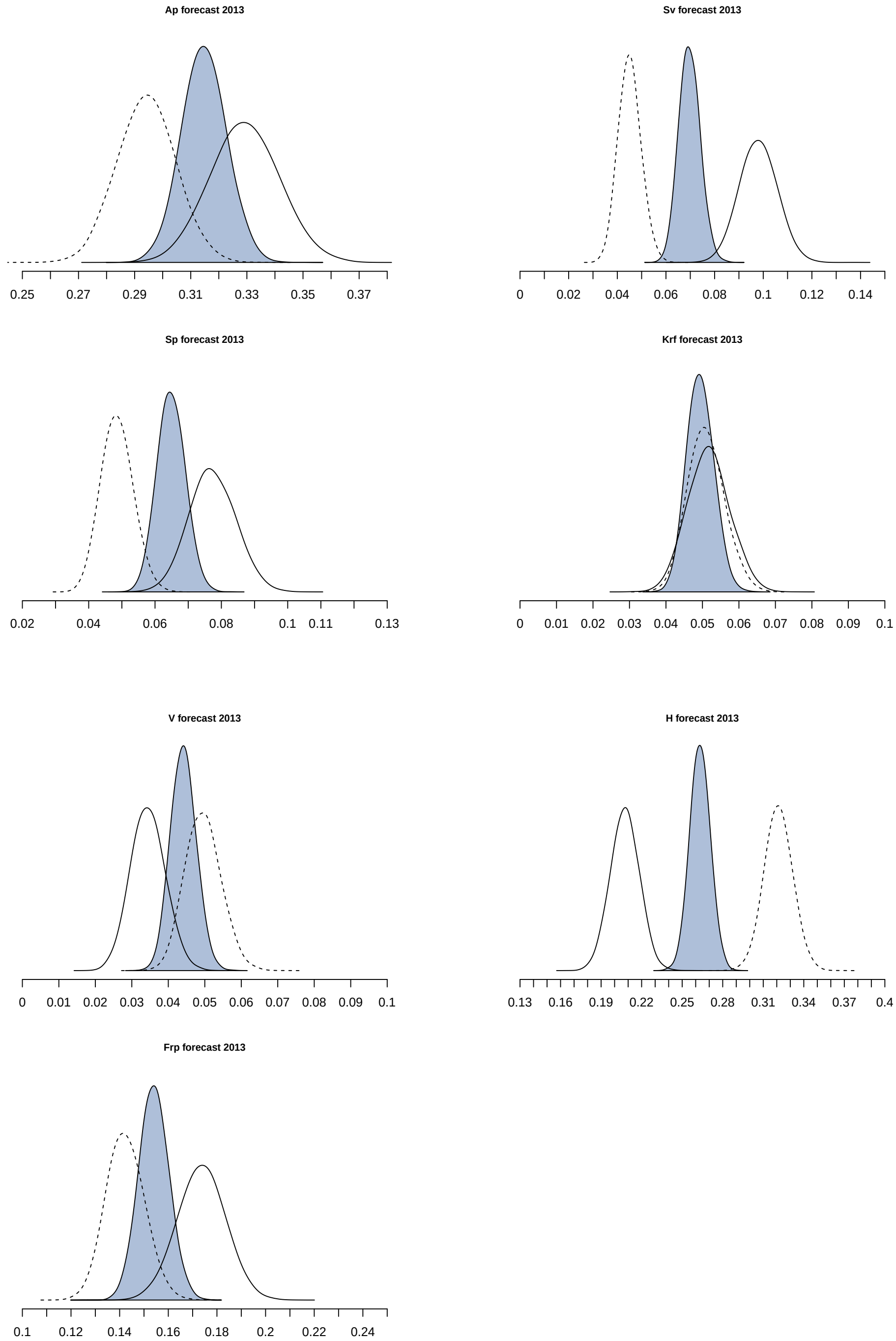


Figure 5.2: Forecasts for the election of 2013. National vote shares. The posterior distribution in grey, the likelihood in dashed lines, and the prior represented by the solid line. The prior is produced by the Multinomial logit model, and the likelihood is the output from the Dirichlet-Multinomial DLM.

	seats	HDR		seats	HDR
Ap	56	[52,60]	Krf	16	[13,18]
Sv	14	[12,17]	V	4	[1,8]
Sp	13	[10,14]	H	28	[24,31]
			Frp	36	[33,39]
Total	83	[79,87]		83	[78,88]

Table 5.6: Prior based on economic and political variables only. Predicted distribution of seats in parliament after the election of 2013. 90% HDRs in parantheses.

	seats	HDR		seats	HDR
Ap	54	[48,59]	Krf	7	[1,11]
Sv	6	[1,10]	V	7	[2,11]
Sp	6	[1,11]	H	56	[49,63]
			Frp	25	[20,30]
Total	67	[60,74]		96	[88,104]

Table 5.7: Likelihood (Linear extrapolation). Predicted distribution of seats in parliament after the 2013 election. This prediction is generated by the inferences made on polling data only. 90% HDRs in parantheses.

2.76%).[7] As discussed above, in Section 5.3, a relatively high unemployment rate is to the benefit of the parties to the left. The figures for CPI growth and GDP growth push the priors in the same direction. CPI growth is currently at 1.5% while GDP growth is at 2.6%, this means that the former is almost at one quarter of the average CPI growth in the last ten election years, and the latter at one third of the average for the last ten election years. The relatively slow economic growth of present is also to the benefit of the left. The exception to this rule is Frp, who even though being a party of the right, seems to benefit from less prosperous economic times. The daily tracking of the latent state as produced by the Dirichlet-Multinomial DLM, on the other hand, shows that the opinion is not pointing in the direction of a Red-Green re-election, as reflected by the dashed graphs in Figure 5.2 being to the left of the priors (and the posteriors) for all three coalition partners. There are certainly other ways of combinig the information of the linear prediction and the prediction generated by the Multinomial logit regression model than the one I have chosen. One would be to combine the predicted distribution of seats of the prior and the likelihood separately, instead of first combining the predicted

[7]The data used to produce the prior forecasts are from `http://www.ssb.no/nasjonalregnskap-og-konjunkturer/nokkeltall/konjunkturer-statistikk-analyser-og-prognoser/`

	mean	HDR		mean	HDR
Ap	0.303	[0.288,0.318]	Krf	0.068	[0.060,0.076]
Sv	0.066	[0.058,0.074]	V	0.047	[0.041,0.055]
Sp	0.058	[0.051,0.066]	H	0.251	[0.237,0.266]
			Frp	0.163	[0.151,0.176]
Total	0.427	[0.411,0.443]		0.530	[0.513,0.546]

Table 5.8: Predicted national vote shares after the election of 2013. 95% HDRs in parantheses.

	seats	HDR		seats	HDR
Ap	56	[53,59]	Krf	8	[7,10]
Sv	12	[10,13]	V	7	[2,9]
Sp	11	[9,12]	H	44	[41,47]
			Frp	27	[25,29]
Total	79	[76,82]		87	[82,91]

Table 5.9: Predicted distribution of seats in parliament after the election of 2013. 90% HDRs in parantheses.

percentage shares, and then translating the result to seats in parliament. Since, as I have shown, the conversion from percentage shares to seats is not a one-to-one mapping, the two approaches do not produce identical results. In Table 5.6 and 5.7 I show the predictions for the distribution of seats in parliament generated by the likelihood (the linear extrapolation of the current level of party support) and the prior separately. Remember that the prior prediction is based solely on political and economic variables, so based one these variables alone the prediction is a tie between the Red-Green coalition and the four opposition parties, with three seats going to the category "others" (with 90% HDR of $[0, 8]$). The share of samples that fall above the majority mark for the Red-Green coalition and the opposition are 31% and 40% respectively, thus giving the opposition a slight advantage. To sum up this short discussion, if political polls are nothing but noise and the economy is all that matters, then the election of 2013 seems like a close race. If, on the other hand, we disregard the economy completely and look exclusively at the latest output of the Dirichlet-Multinomial DLM, the opposition gets an overwhelming majority with 96 seats. I believe that it is in combining the information of these two models that the best forecast is obtained, and I will now present the results of this averaging of the two models.

All the figures presented in Table 5.8 and 5.9 are based on 100 000 samples from the posterior distribution produced by the Multinomial logit forecast model, and then supplying Saint-Laguës method with a subset of these samples.[8] These numbers are obtained by multiplying the two distributions I get from the Dirichlet-Multinomial DLM and the Multinomial logit regression model. In addition to the seven parties reported in the tables the category "others" is predicted to obtain 4.32% of the votes and 3.7 seats in parliament with corresponding 90% HDRs at $[0.036, 0.050]$ and $[0, 7]$ respectively. The Red-Green coalition is predicted to obtain 79 seats, seven seats less compared to what they enjoy currently. Furthermore, the majority mark of 85 seats is outside the 90% highest density region for the coalition. The opposition parties of Krf, V, H and Frp are predicted to receive 87 seats in parliament, with 77.5% of the samples falling above the majority mark of 85 seats. This means that *the prediction as of May* 21*. is: with a probability of* 0.775 *the four opposition parties will win a majority of the seats in parliament.*

[8] The point forecasts presented in Figure 5.2 are based on all 100 000 samples, because the Python-code I have written to calculate Saint-Laguës method have yet to be optimized, I supplied it with a subset of 1 000 random samples.

6

Conclusion and plans for further research

"Elections are predictable". These words belong to the statistician and political scientist Andrew Gelman (1997), and with the experience of successful predictions of election outcomes from other countries, we have little reason to believe that they are not. As we have seen, most of the countries where prediction has been successful have party systems different from the Norwegian, they are presidential systems or systems with two parties or two major blocs. But, looking at the electoral systems of the world, these are the systems that are particular and strange, most democratic countries are multiparty systems (Clark et al., 2013, 587). Clearly, more modesty is demanded when attempting to predict election outcomes in systems with many parties and changing alliances, but at the same time the constellation and the relative size of parties in a country are slowly changing characteristics of a political system. And when it comes to the voters, besides having more options, which makes them seem more volatile, we have no reason to believe that the behaviour of voters in multiparty systems are less predictable.

In this thesis I have worked with two models particularly suited for the task of tracking and forecasting party support in multiparty systems, and combined the information they provide to produce forecasts for the Norwegian election of 2013. The Dirichlet-Multinomial DLM and the Multinomial logit regression model are, from a statistical perspective, the theoretically sound models for the task, and I am convinced that the model I am aiming for, the correct model so to say, is an extension of the combination of these two models. By the combination of the two, I mean the manner in which I have combined information from the two models to arrive at a posterior distribution for my forecasts. If I consider the Dirichlet-Multinomial DLM and the Multinomial logit model of this thesis together, the former seems to me as the most solid, even though more work is demanded to fully explore how to best apply this model in a dynamic setting. When it comes to the latter model, I see primarily three deficiencies. First, a good model is not a cure for a lack of data, and estimation of such a large model as the Multinomial logit demands more observations than the ten elections I have estimated it on. An interesting question in this regard is whether the model could be estimated with data points from several different countries, with similar party structures. For example, are the post-materialist left parties (such as Sv in Norway) across multiparty systems similar enough to use data from several countries to determine the effect of incumbency, or unemployment, on the support for these parties? Since elections are rare events, this is a question that is worth exploring in further research. Second, massive amounts of good data are not a substitute for solid theory. The focus of this thesis has not been on social scientific theory, but on the development and novel applications of statistical models. Therefore, lack of solid theory is a weakness of the analyses I conducted in Section 5.5, where the theoretical justification for the model specification bordered at simple hand-waving. I must add that despite this deficiency, the results produced by the Multinomial logit regression model makes sense. In future applications more solid theoretical underpinnings of the model are essential. I think that a promising avenue of research is in applying theory on the voting behaviour of individuals in order to model voting as a two-stage decision process (Steenbergen et al., 2011). In such a process the voters, in a first stage, use broad-based heuristics such as class and partisanship to narrow down the options to a few viable parties they consider voting for. In the second stage more aggregated variables could be used to predict the party that the voters finally opt for. In such a model I could use a Multinomial logit regression model with different variables for the different groups of voters delimited in the first stage. This would be in accordance with the intuition I have that different groups of voters are moved by varying types of issues.

Despite these objections to the prime forecast model of this thesis, the model combining the Dirichlet-Multinomial DLM and the Multinomial logit regression model, I believe that the forecasts it produces are the best forecasts of the Norwegian election of 2013 as of May this year. I would put my money on them.

Bibliography

Abramowitz, A. I. (2008). Forecasting the 2008 presidential election with the time-for-change model. *Political Science & Politics*.

Albert, I. and Denis, J.-B. (2012). Dirichlet and Multinomial distributions: properties and uses in jags. Technical report, Unité Mathématiques et Informatique Appliquées, INRA `www.inra.fr/miaj/public/nosdoc/rap2012-5.pdf`.

Arnesen, S. (2011a). How prediction markets help us understand events' impact on the vote in us presidential elections. *The Journal of Prediction Markets*, 5:42–63.

Arnesen, S. (2011b). Informasjon, motivasjon, prediksjon. eit forsøk med prediksjons-marknad før stortingsvalet 2009. *Tidsskrift for Samfunnsforskning*, 52:99–121.

Arnesen, S. (2012a). *Leaping into the unknown. Comparing, testing, and applying methods of predicting elections.* PhD thesis, University of Bergen.

Arnesen, S. (2012b). Out of work and out of office: The golden mix of left bloc support in Norway.

Bartels, L. M. and Zaller, J. (2001). Presidential Vote Models: A Recount. *Political Science & Politics*, pages 9–20.

Beck, N., Jackman, S., and Rosenthal, H. (2006). Presidential approval: the case of George W. Bush. `http://jackman.stanford.edu/mcmc/index.php`.

Bonnet, C. and Wagner, P. (2006). *L'âge d'or de l'empirisme logique. Vienne-Berlin-Prague 1929-1936.* Gallimard.

Carlsen, F. (2000). Unemployment, inflation and government popularity — are there partisan effects? *Electoral Studies*, 19:141–150.

Clark, W. R., Golder, M., and Golder, S. N. (2013). *Principles of Comparative Politics.* Sage CQ Press.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics. Fourth Ed.* Addison-Wesley.

Devore, J. L. and Berk, K. N. (2007). *Modern Mathematical Statistics with Applications.* Duxbury.

Foucault, M. and Nadeau, R. (2012). Forecasting the 2012 French Presidential Election. *PS: Political Science and Politics.*

Gelman, A. (1997). Bayesian Analysis of Election Surveys and Forecasts: Learning from the State of Valencia Experience `http://www.stat.columbia.edu/~gelman/research/published/bernardo.pdf`.

Gelman, A., Carlin, J. B., and Stern, Hal S. Rubin, D. B. (2004). *Bayesian Data Analysis.* Chapman & HALL/CRC.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55:665–676.

Hanretty, C. (2013). The 2013 Italian Election. A poll-based forecast `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2214605`.

Hegre, H., Karlsen, J., Mokleiv Nygård, H., Urdal, H., and Håvard., S. (2012). Predicting Armed Conflict, 2010-2050. *International Studies Quarterly*, pages 1–21.

Hempel, C. G. and Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15:135–175.

Hibbs, D. A. (1977). Political parties and macroeconomic policy. *American Political Science Review*, 71:1467–1487.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods.* Springer.

Jackman, S. (2005). Pooling the polls over an election campaign. *Australian Journal of Political Science*, 40:4:499–517.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences.* Wiley.

Jupskås, A. R. (2011). Oppslutningen om høyresiden: endrede styrkeforhold? `http://www.sv.uio.no/isv/forskning/aktuelt/blogg/ta-politika/2011/frp-hoyre.html`.

Lebo, M. and Norpoth, H. (2011). Yes, Prime Minister: The key to forecasting British elections. *Electoral Studies*, 30:258–263.

Lebo, M. L. and Norpoth, H. (2007). The PM and the Pendulum: Dynamic Forecasting of British Elections. *British Journal of Political Science*, 37:71–87.

Lewis-Beck, M. S. and Paldam, M. (2000). Economic voting: an introduction. *Electoral Studies*, 19:113–121.

Lid Hjort, N. (2012). STK 4020: Bayesian Statistics. Course Notes and Exercises. Technical report, University of Oslo.

Linzer, D. A. (2012). Dynamic Bayesian Forecasting of Presidential Elections in the States.

Lock, K. and Gelman, A. (2010). Bayesian combination of state polls and election forecasts. *Political Analysis*, 18:337–348.

Long, S. J. (1997). *Regression Models for Categorical and Limited Dependent Variables*. SAGE.

Lynch, S. M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer.

Nadeau, R., Lewis-Beck, M. S., and Bélanger, E. (2010). Electoral forecasting in France: A multi-equation solution. *International Journal of Forecasting*, 26:11–18.

Nannestad, P. and Paldam, M. (1994). The vp-function: A survey of the literature on vote and popularity functions after 25 years. *Public Choice*, 79:213–245.

Norpoth, H. (1991). The Popularity of the Thatcher Government: A Matter of War and Economy. In *Economics and Politics*. Michigan University Press.

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

Plummer, M. (2011). JAGS Version 3.1.0 user manual. Technical report.

Schrodt, P. A. (2010). Seven Deadly Sins of Contemporary Quantitative Political Analysis `http://polmeth.wustl.edu/media/Paper/Schrodt7SinsAPSA10.pdf`.

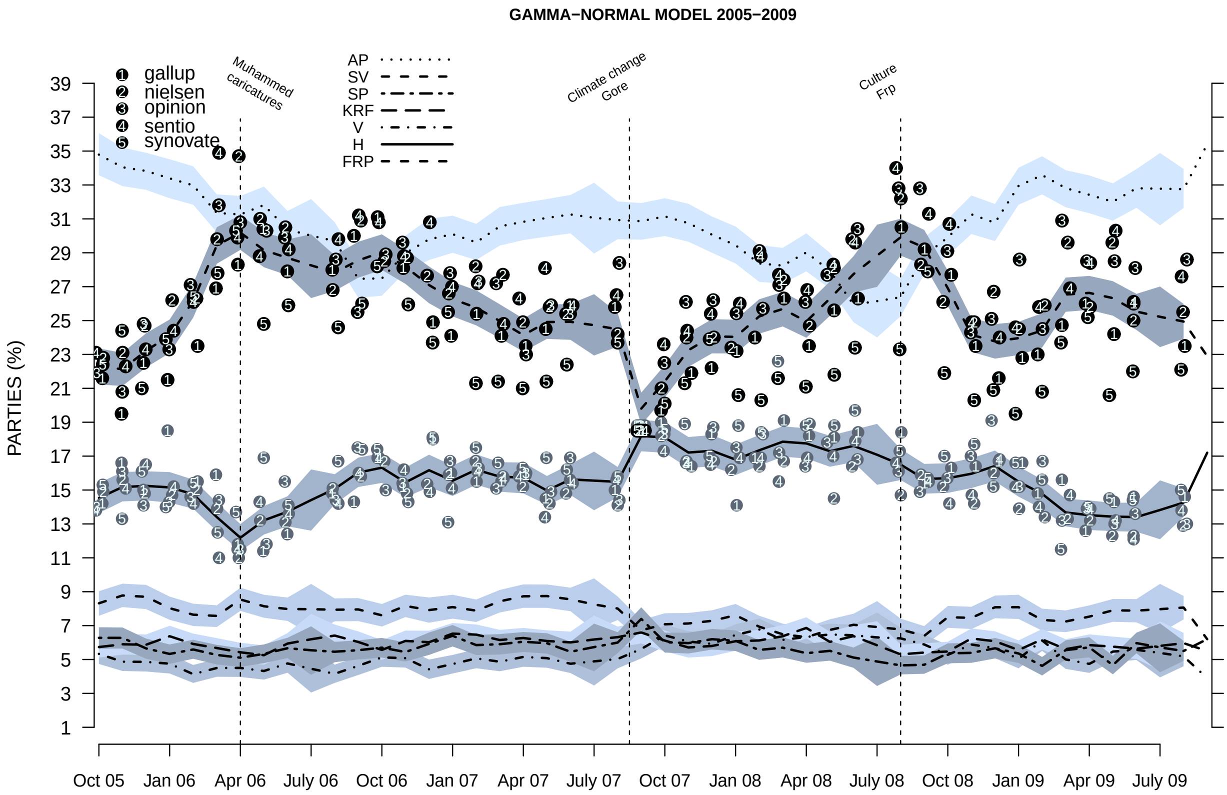
Steenbergen, M. R., Hangartner, D., and E. de Vries, C. (2011). Choice under Complexity: A Heuristic-systematic Model of Electoral Behaviour `http://www.catherinedevries.eu/ChoiceUnderComplexity.pdf`.

Swank, O. (1993). Popularity functions based on the partisan theory. *Public Choice*, 75:339–356.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* Springer.

Appendix A

Figures



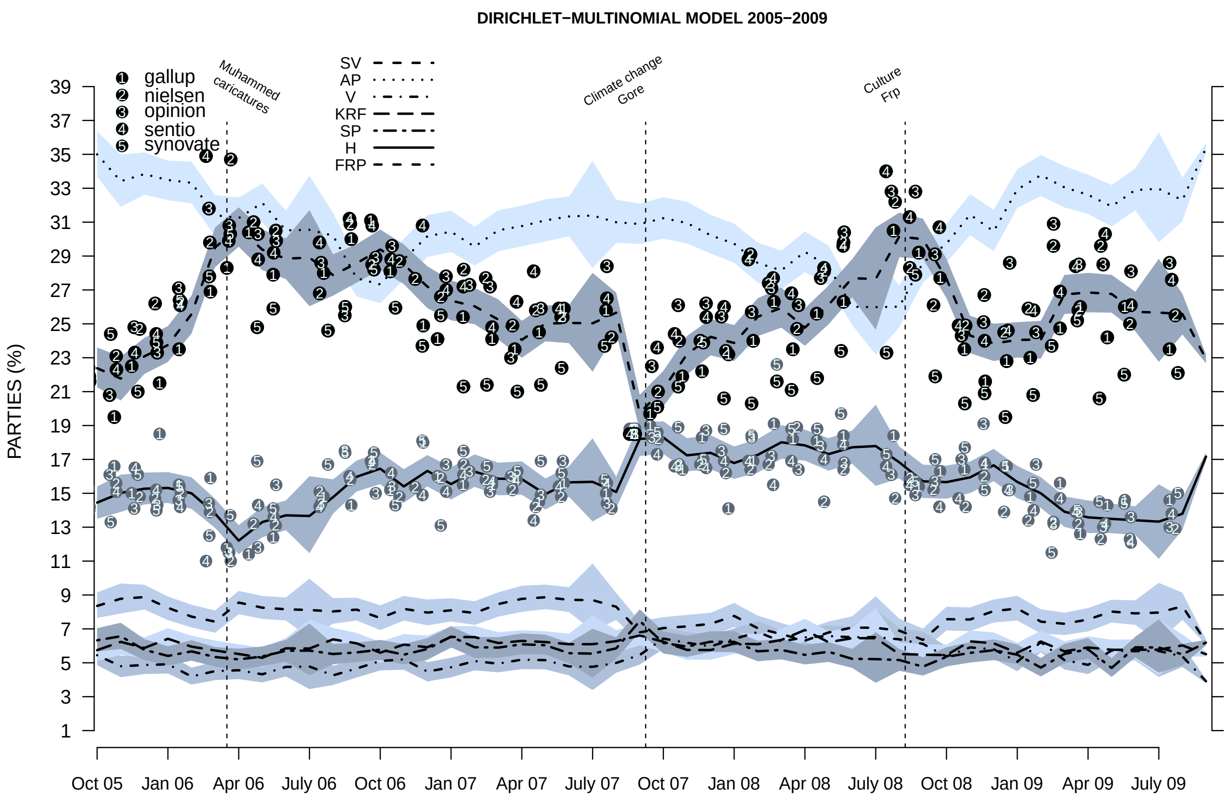
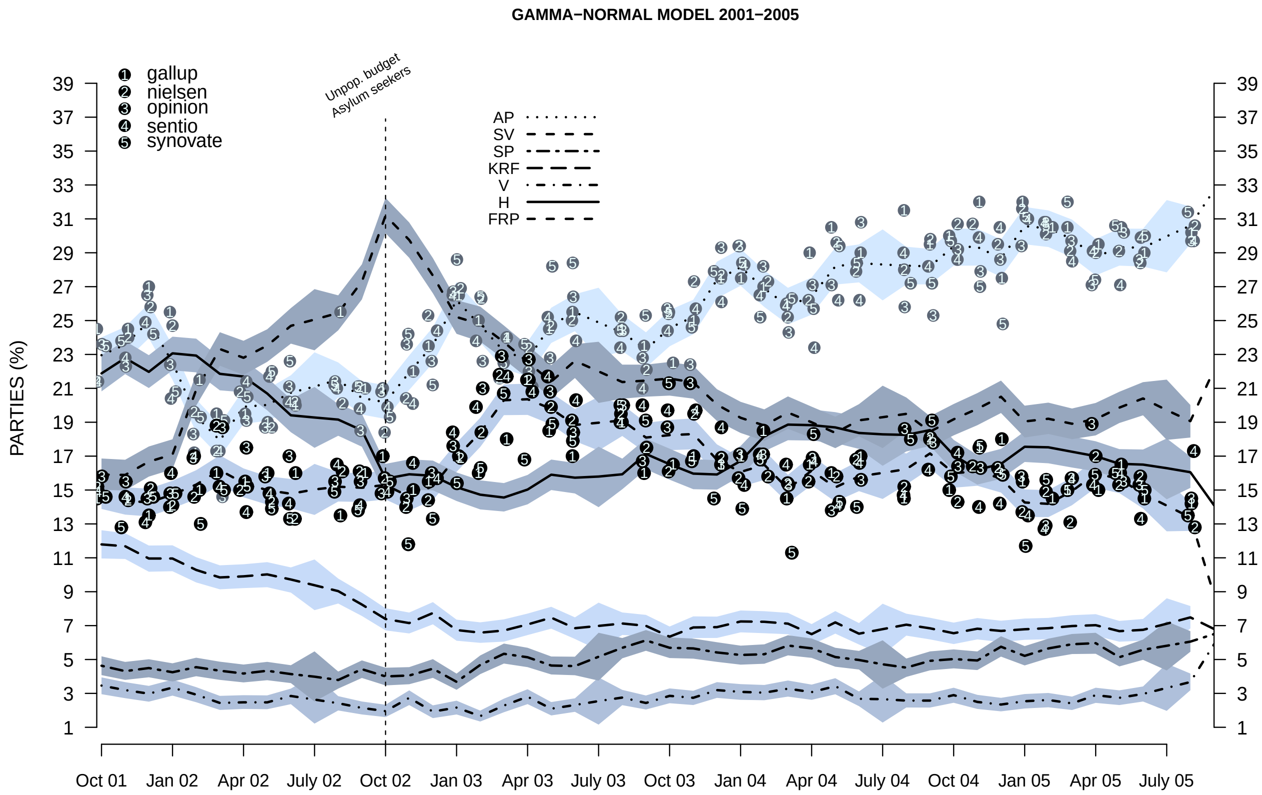


Figure A.1: All parties 2005-2009. Weekly estimates. 95% HDRs. Poll results for H and Frp are represented with a plotted point at their respective point estimates.



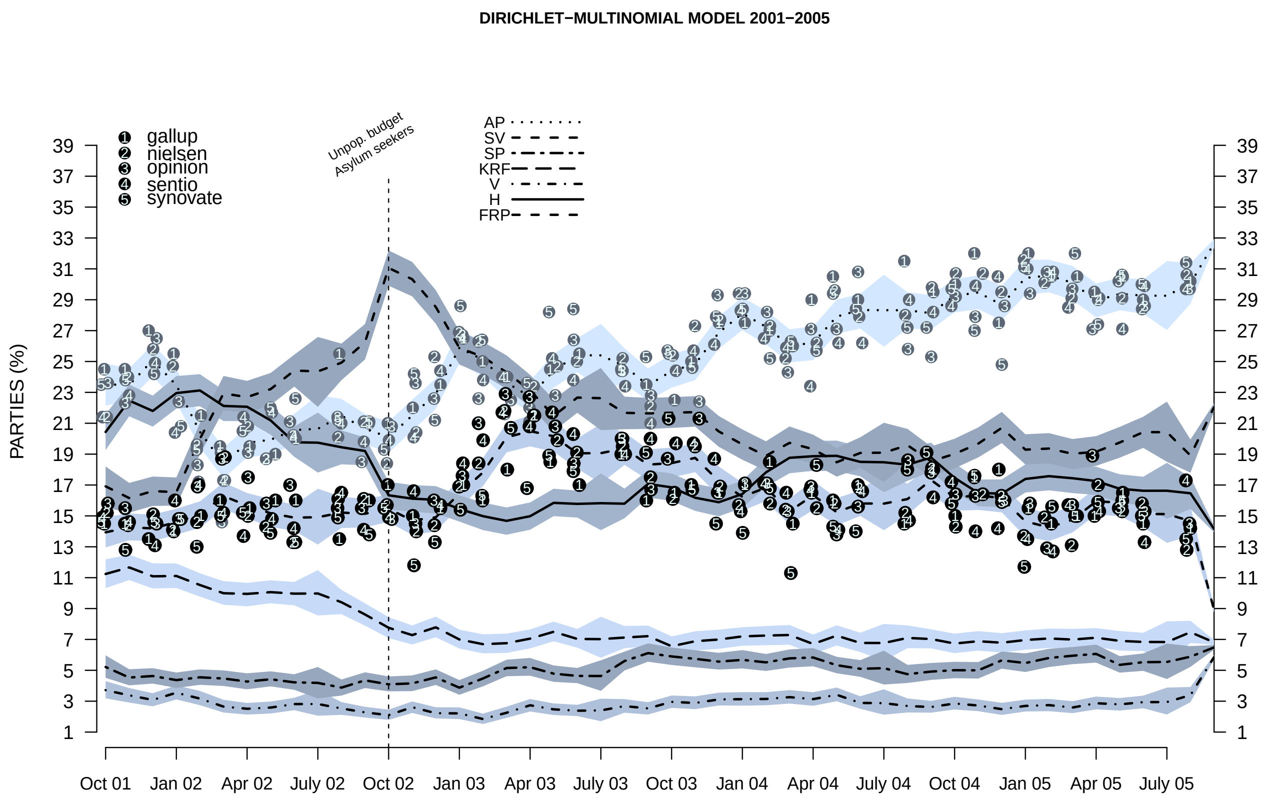
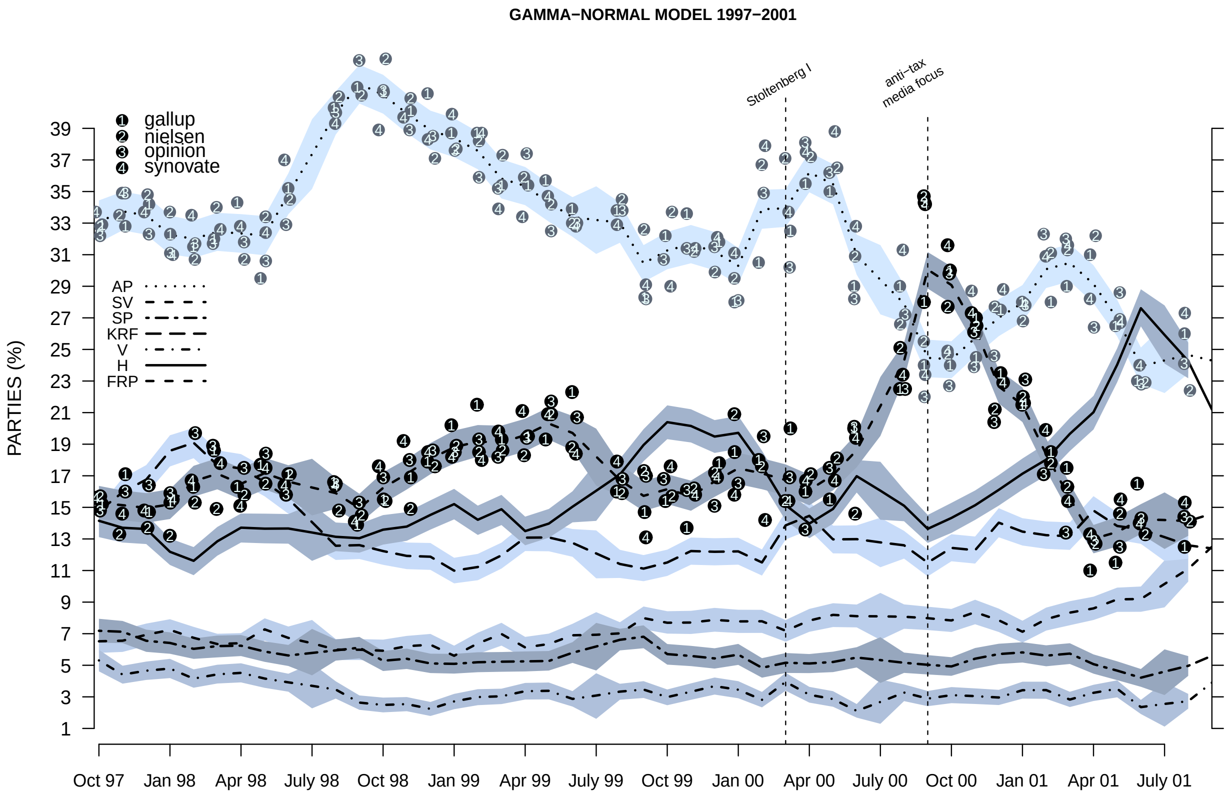


Figure A.2: All parties 2001-2005. Weekly estimates. 95% HDRs. Poll results for Ap and Sv are represented with a plotted point at their respective point estimates.



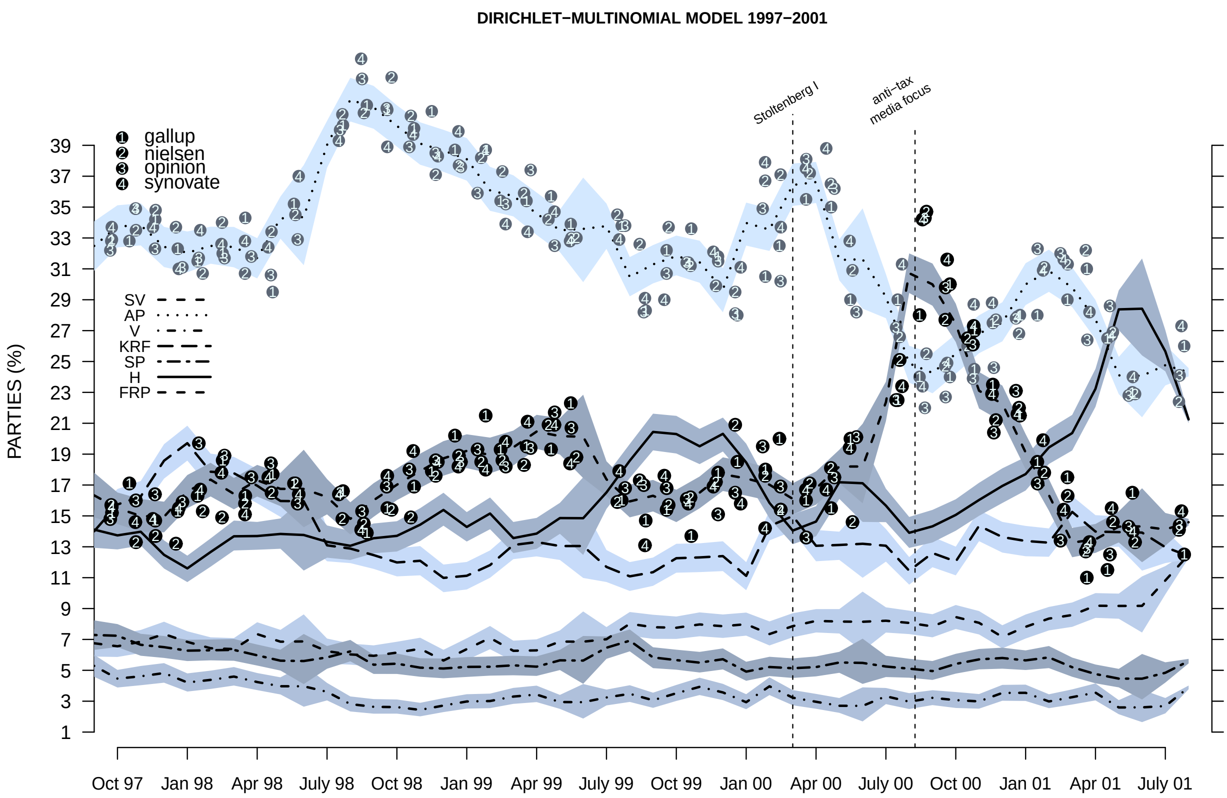


Figure A.3: All parties 1997-2001. Weekly estimates. 95% HDRs. Poll results for A and Frp are represented with a plotted point at their respective point estimates.

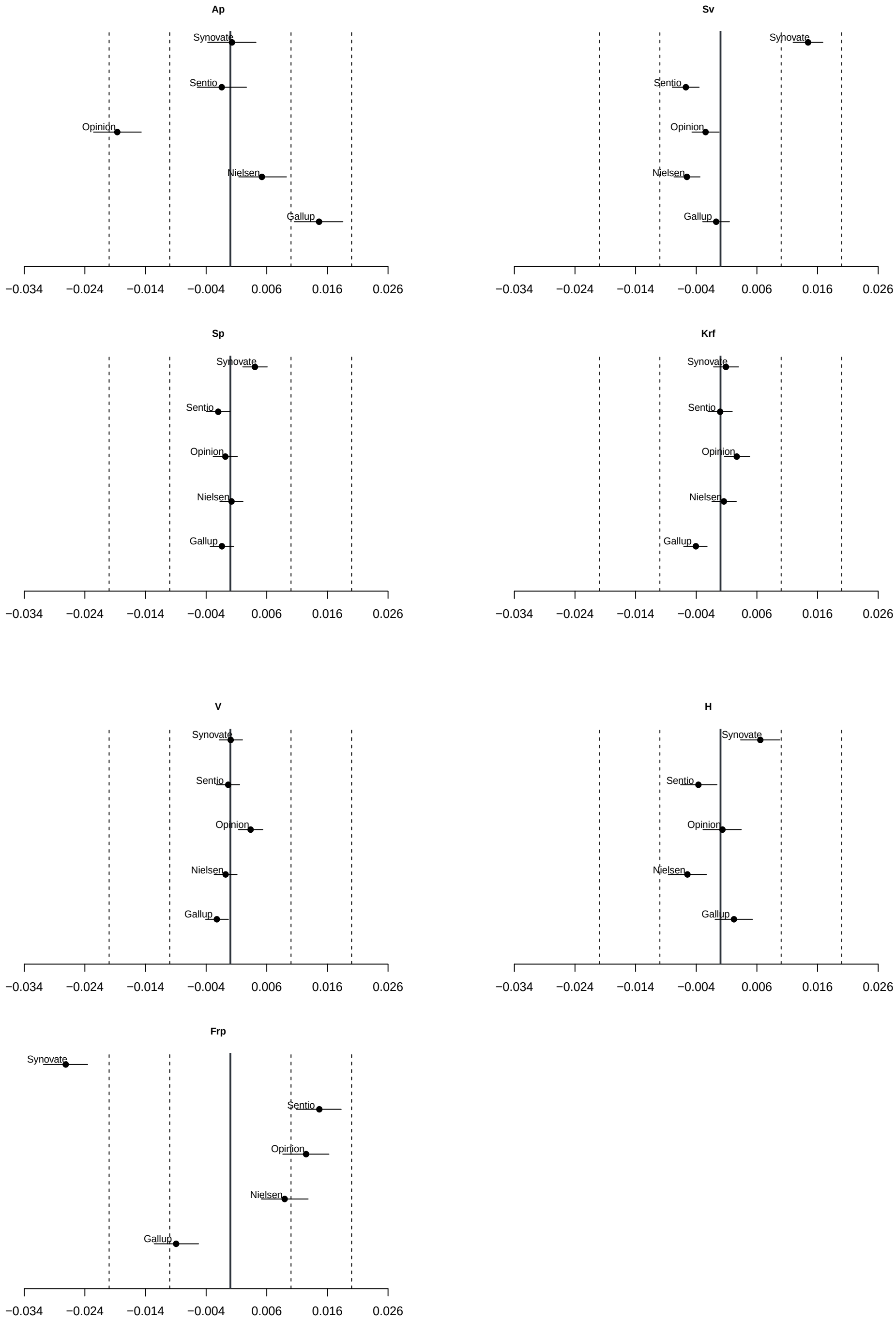


Figure A.4: House effects 2005-2009. Circles are estimated posterior mean, the lines connect the 2.5 and 97.5 percentiles of the posterior distributions.

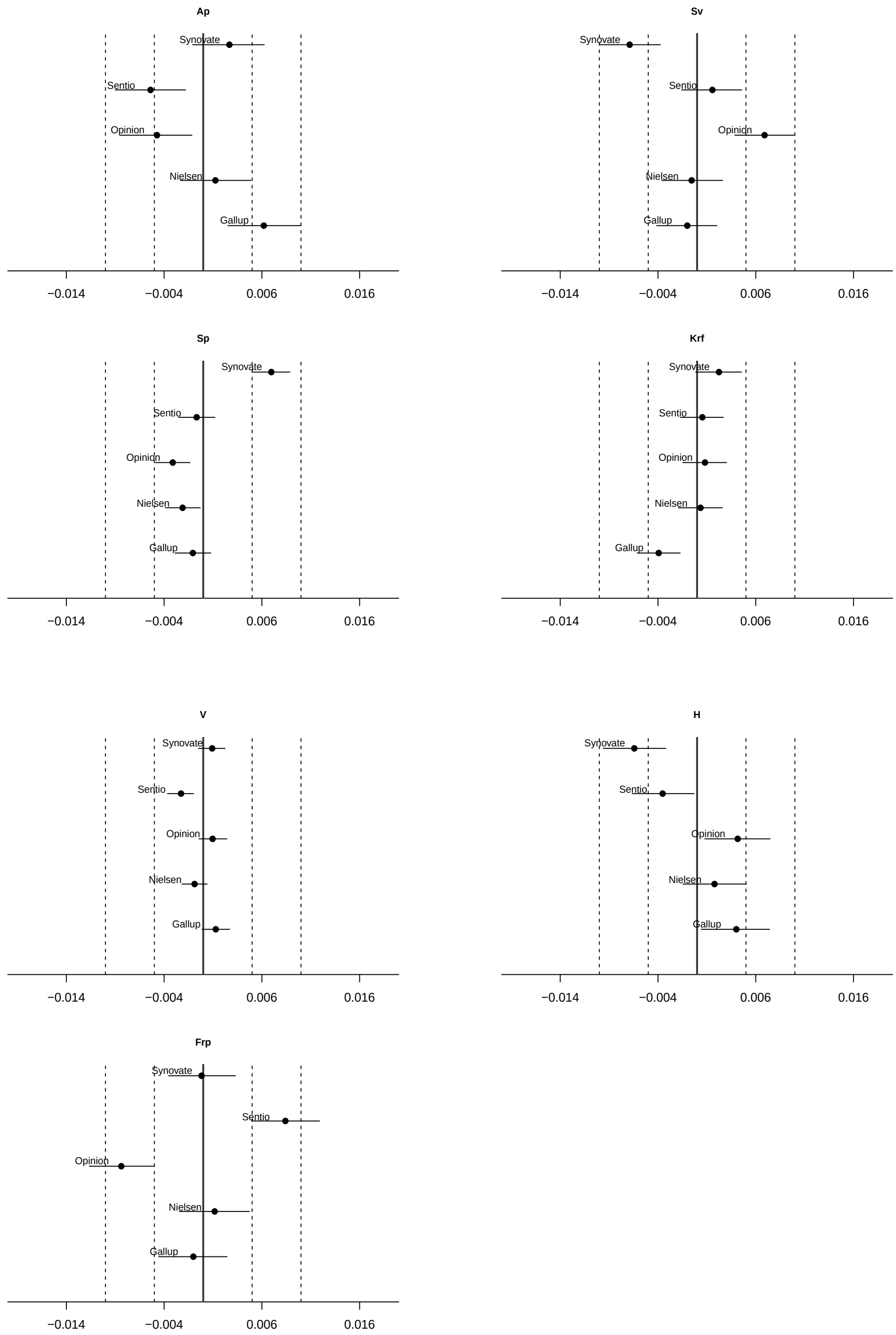


Figure A.5: House effects 2001-2005. Circles are estimated posterior mean, the lines connect the 2.5 and 97.5 percentiles of the posterior distributions.

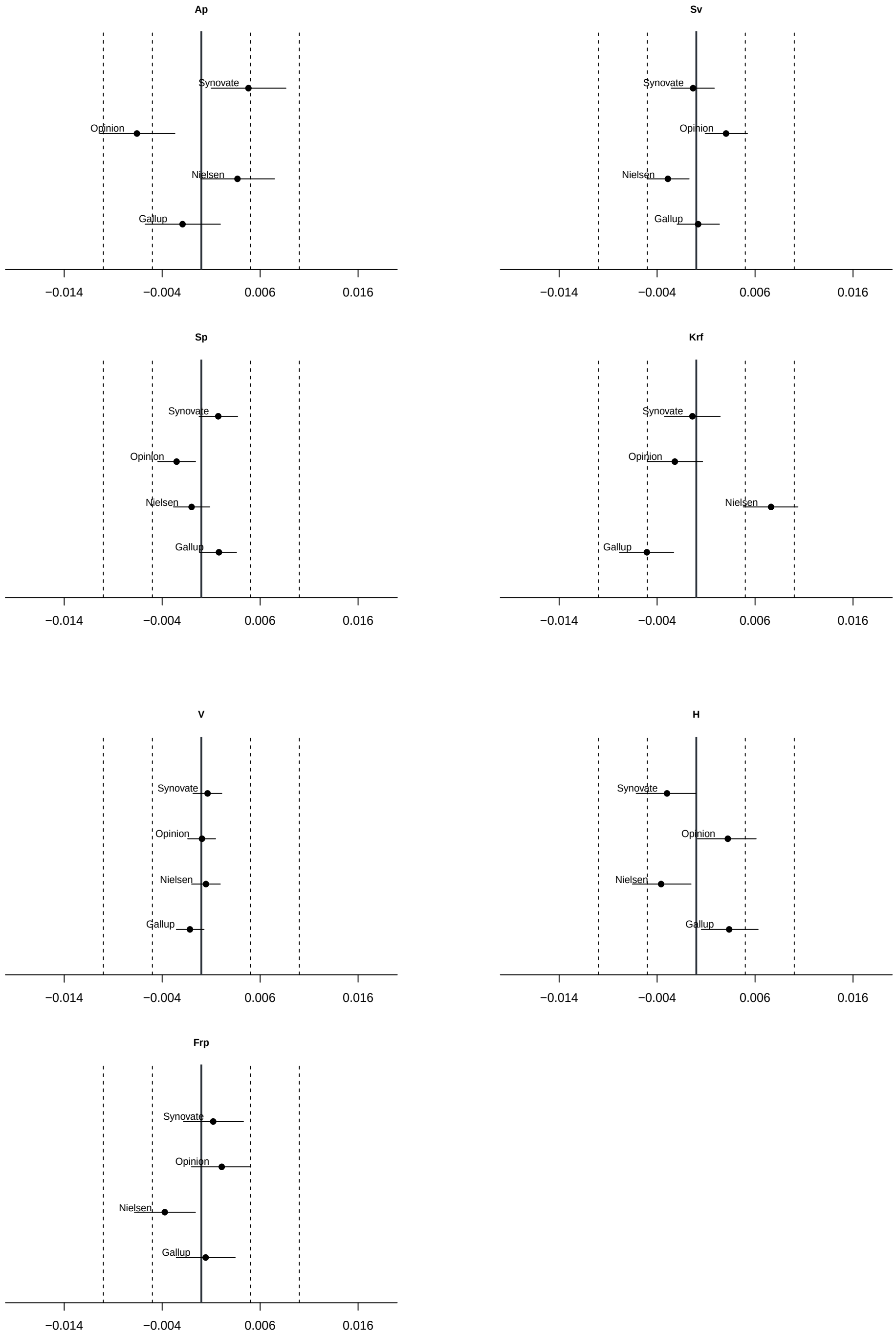


Figure A.6: House effects 1997-2001. Circles are estimated posterior mean, the lines connect the 2.5 and 97.5 percentiles of the posterior distributions.

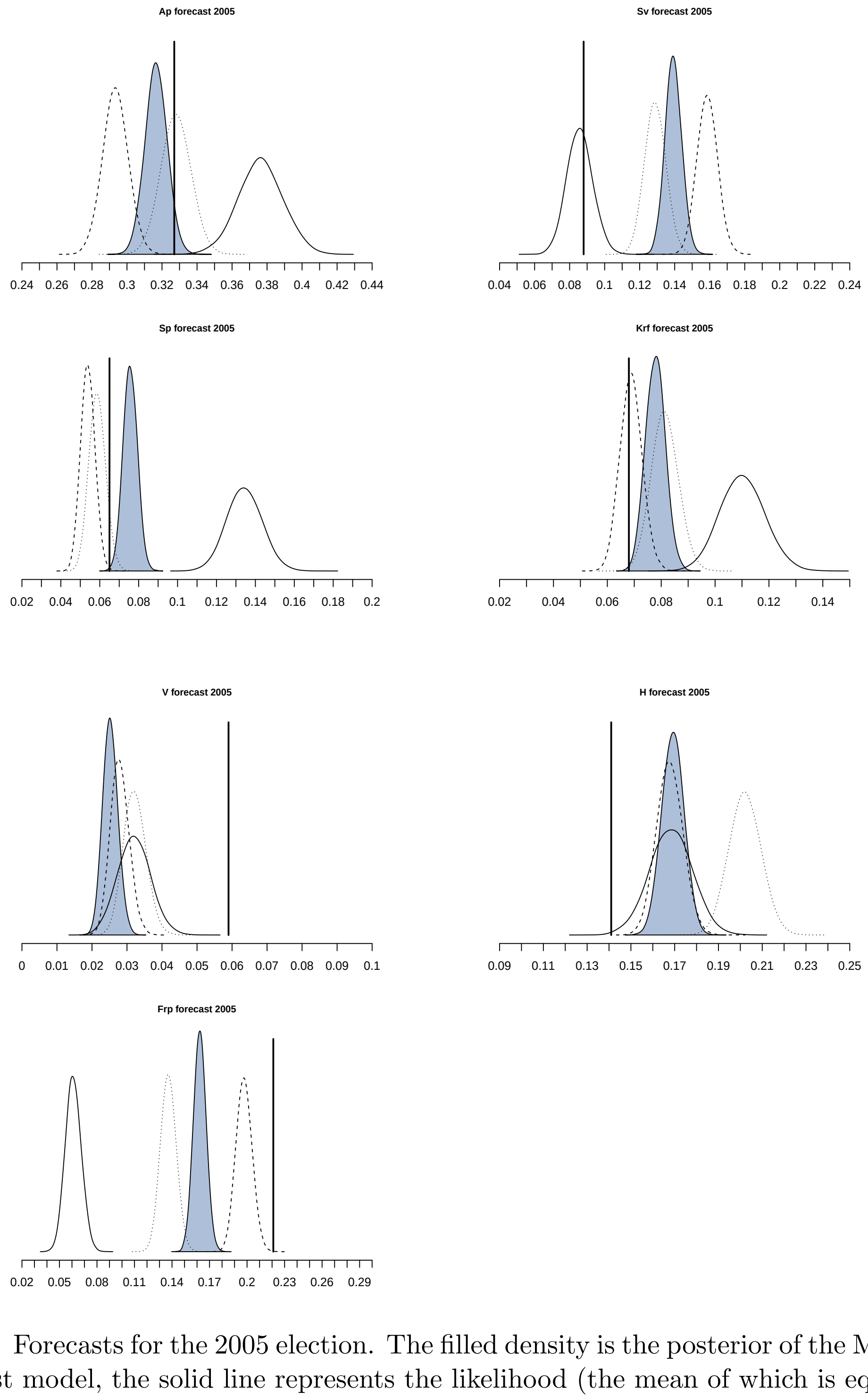


Figure A.7: Forecasts for the 2005 election. The filled density is the posterior of the Multinomial logit forecast model, the solid line represents the likelihood (the mean of which is equivalent to a linear forecast) and the dashed line is the prior. The line consisting of small points is the forecsats distribution produced by the quadratic function.

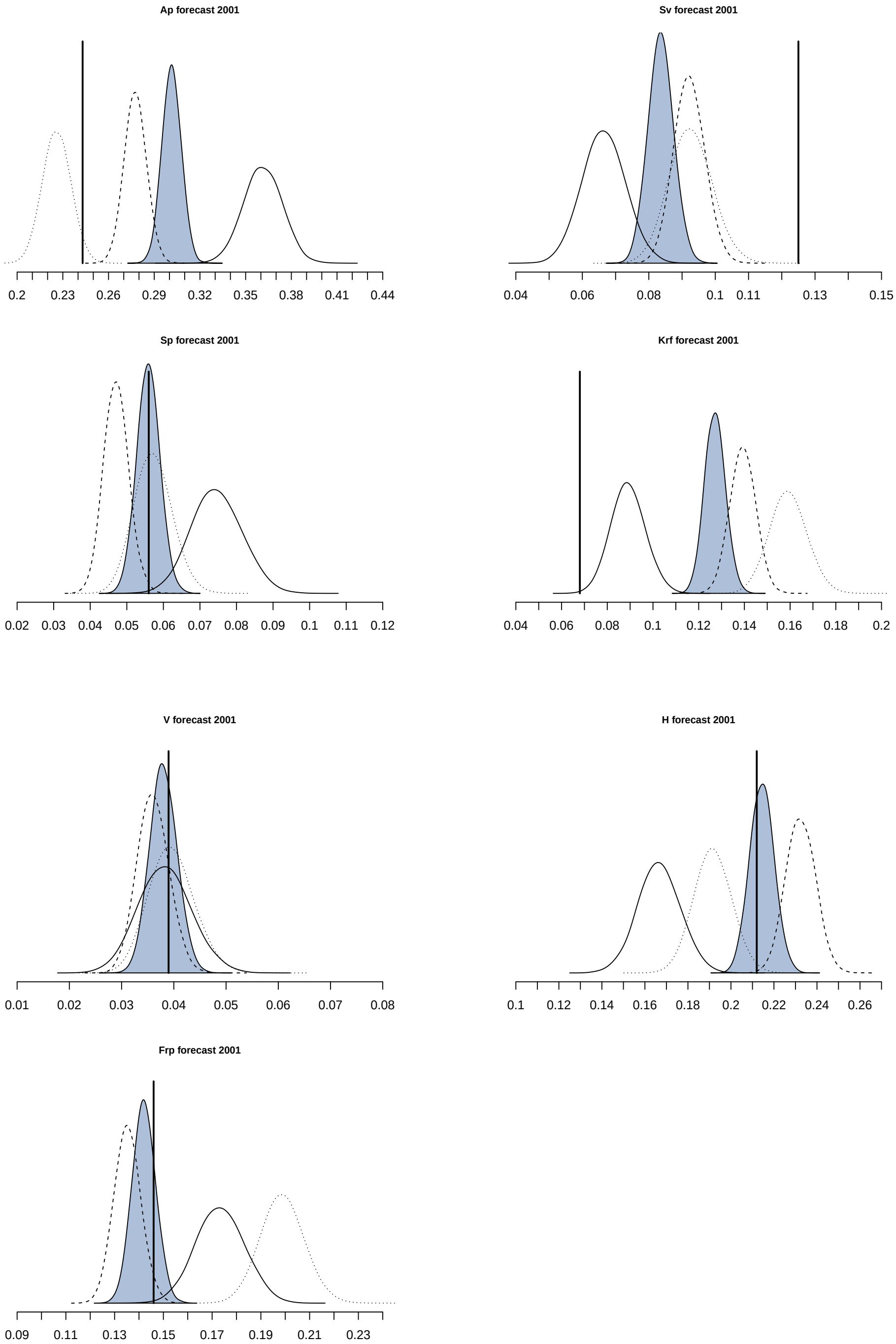


Figure A.8: Forecasts for the 2001 election. The filled density is the posterior of the Multinomial logit forecast model, the solid line represents the likelihood (the mean of which is equivalent to a linear forecast) and the dashed line is the prior. The line consisting of small points is the forecsats distribution produced by the quadratic function.

Appendix B

The GN-model

In the following I give a more formal presentation of the Gamma-Normal model. I substantiate how the priors are set, show how the posterior distributions are dervied and provide an example of how this model is implemented in JAGS (Plummer, 2011).

B.1 The model

As seen in Section 4.1 the time-series' in the Gamma-Normal model were considered independent. Therefore is suffices to consider the time series of one individual party, since they are all derived in the same manner. Let α_t be the intended vote share for one of the parties at time t, and let $i = 1, \ldots, n$ index the polls available for analysis. Each of the polls is assumed generated by

$$y_{ti} = \mu_{ti} + \nu_{ti}, \quad \nu_{ti} \sim N[0, \sigma_{ti}^2] \tag{B.1.1}$$

where y_i is the results of poll i. The expectation μ_{ti} is a linear combination of the latent state α_t and the house specific bias δ_j, thus

$$\mu_{ti} = \alpha_t + \delta_j \tag{B.1.2}$$

To model the change in voting intentions I use a random walk model

$$\alpha_t = \alpha_{t-1} + \epsilon_t \quad \epsilon_t \sim N[0, \omega^2] \tag{B.1.3}$$

with a distribution

$$\alpha_1 \sim \text{Unif}(l, u) \tag{B.1.4}$$

where for H, to pick a party, $l = .10$ and $u = .35$, which brackets the historical range of election results. With this model I assume that today's level of support for H is the same as yesterdays except for random shocks that come from a normal distribution with mean zero and variance σ^2.

B.2 The priors

Viewing the parameters as random and ascribing prior distributions to these is what makes Bayesian statistics Bayesian. Equation B.1.3 and B.1.4 are the priors for the α_t parameters. For the house-effects δ_j I ascribe a normal prior with mean zero

$$\delta_j \sim N[0, d^2] \tag{B.2.1}$$

where d^2 is a large value, which makes the normal distribution flat, so that the data (the likelihood) will dominate the priors. This choice reflects the fact that I posit no prior knowledge about the possible size and direction of the house specific biases. I set $d^2 = 100$ for all the J polling houses, to let the data dominate completely. For the variance parameter ω^2 I consider the precision $\phi = 1/\omega^2$ and assign the precision a gamma distribution,

$$f(\phi) \propto \phi^{a-1} e^{-b\phi} \tag{B.2.2}$$

The values I ascribe the parameters depend on what I view as the probable volatility of party support for each individual party. To take Ap as an example, I set $a = 5000$ and $b = 1$ so that the expected weekly volatility of party support for Ap is at on fifth of a percentage point. This seems little, but since I am looping over the time series several thousand times the total number of paths that $\{\alpha_t\}$ take will provide a good description of the uncertainty associated with the point estimates.

B.3 The posterior distribution

To estimate the model I implement it in a program for Bayesian analysis called JAGS, which is short for Just Another Gibbs Sampler (Plummer, 2011). Gibbs sampling is a Markov Chain Monte Carlo method for obtaining approximately independent samples from a posterior distribution. Gibbs sampling is used for complex multivariate distributions for which the integral is difficult to compute. In short, the Gibbs sampler simulates draws from the (univariate) conditional posterior distributions of the model parameters,

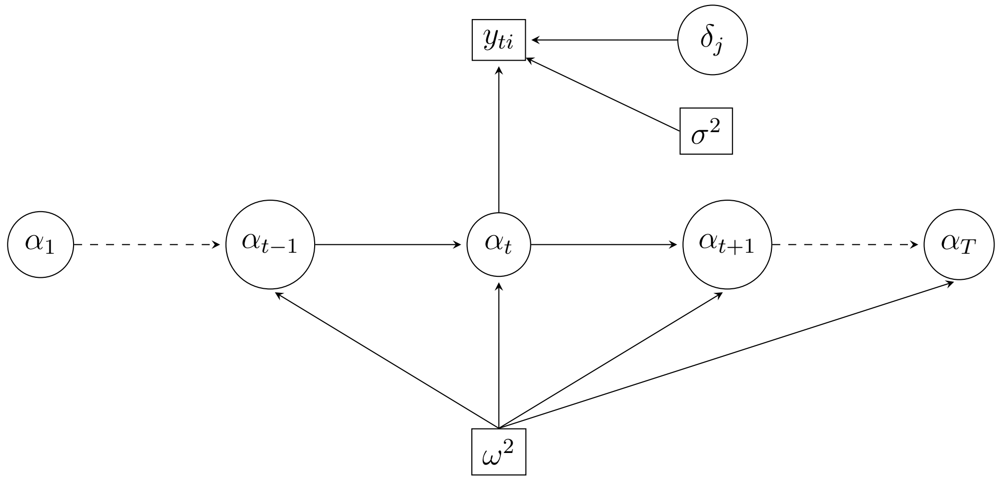


Figure B.1: Equation B.1.1, B.1.2 and B.1.3 given as an Directed Acyclic Graph (DAG). The arrows run from the parent nodes to the child nodes.

holding all the other parmeters fixed at their previously sampled values. After a certain number of iterations the sequence of draws will converge to the target distribution we are trying to simulate (Hoff, 2009, 97). Introductions to Gibbs sampling are found in all books on Bayesian statistics. In order to sample from the conditional distributions of the model parameters a description of these is required. To derive these distributions Propostion 5.2 in Jackman (2009, 227) is helpful. According to this proposition, if a statistical model can be expressed as a directed acyclic graph $\mathcal{G}$ then the conditional density of the node θ_j in the graph is

$$f(\theta_j|\mathcal{G} \setminus \theta_j) \propto f(\theta_j|\text{parents}[\theta_j])\times \prod_{\tau \in \text{children}[\theta_j]} f(\tau|\text{parents}[\tau]) \tag{B.3.1}$$

where $\setminus$ is the set-theoretic difference.[1] In Figure B.1 I show the DAG of the Gamma-Normal model as defined above. When a DAG is used to characterize a Bayesian statisitcal model the circular nodes represents the stochastic quantities, while the square nodes are non-stochastic (Jackman, 2009, 226). The arrows in the DAG run from the parent nodes to the child nodes. As seen in the Figure the observed poll results y_{ti} are the children of the latent states α_t, the house effects δ_j and the variance σ^2_{ti}. This last parameter σ^2_{ti} is in a square box, and is considered non-stochastic. By this I mean that I consider σ^2_{ti} as a given feature of the poll that gives the precision of the measurement, and more importantly, no prior is ascribed to this parameter. With the DAG the conditional distribution of posterior

[1] If A and B are sets then the set theoretic difference is defined by $A \setminus B = \{a|a \in A, a \notin B\}$.

can be derived. Because the model involves extensive use of the Normal distribution it is convenient to derive the conditional distribution $f(x|a)$ when $x \sim N[\mu, \sigma^2]$ and $\mu \sim N[a, b^2]$, before we proceed to the posterior distributions of the GN-model. Since

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

and

$$f(\mu) = \frac{1}{\sqrt{2\pi}b} e^{-\frac{1}{2b^2}(\mu-a)^2}$$

we get via Bayes Theorem that

$$\begin{aligned}
f(\mu|x) \propto f(x)f(\mu) \propto e^{-\frac{1}{2\sigma^2}(x-\mu)^2} e^{-\frac{1}{2b^2}(\mu-a)^2} &= \exp\left\{-\frac{1}{2\sigma^2 b^2}[(x-\mu)^2 b^2 + (\mu-a)^2\sigma^2]\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2 b^2}[x^2b^2 - 2x\mu b^2 + \mu^2 b^2 + \mu^2\sigma^2 - 2\mu a\sigma^2 + a^2\sigma^2]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2 b^2}[\mu^2(b^2+\sigma^2) - 2\mu(xb^2 + a\sigma^2)]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\mu^2\left(\frac{1}{\sigma^2}+\frac{1}{b^2}\right) - 2\mu\left(\frac{x}{\sigma^2}+\frac{a}{b^2}\right)\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{b^2}\right)\left(\mu^2 - 2\mu\frac{\frac{x}{\sigma^2}+\frac{a}{b^2}}{\frac{1}{\sigma^2}+\frac{1}{b^2}}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{b^2}\right)\left(\mu - \frac{\frac{x}{\sigma^2}+\frac{a}{b^2}}{\frac{1}{\sigma^2}+\frac{1}{b^2}}\right)^2\right\} = \exp\left\{-\frac{1}{2\tilde{b}^2}(\mu-\tilde{a})^2\right\}
\end{aligned}$$

Thus, $x|\mu \sim N[\tilde{a}, \tilde{b}^2]$ where the expectation is

$$\tilde{a} = \frac{\frac{x}{\sigma^2}+\frac{a}{b^2}}{\frac{1}{\sigma^2}+\frac{1}{b^2}} \tag{B.3.2}$$

and variance

$$\tilde{b}^2 = \frac{1}{\sigma^2} + \frac{1}{b^2} \tag{B.3.3}$$

With this results at hand it is a matter of insertion to derive the following two conditional posterior distributions. First, the conditional posterior distribution of $f(\alpha_t|\mathcal{G} \setminus \alpha_t)$. From the DAG we see that parents of α_t are α_{t-1} and ω^2, while its children are α_{t+1} and y_i, this gives

$$f(\alpha_t|\mathcal{G} \setminus \alpha_t) \propto f(\alpha_t; \alpha_{t-1}, \omega^2) f(\alpha_{t+1}; \alpha_t, , \omega^2) f(y_i; \mu_{t,i}, \sigma^2) \tag{B.3.4}$$

where

$$\alpha_t \sim N[\alpha_{t-1}, \omega^2] \tag{B.3.5}$$

$$\alpha_{t+1} \sim N[\alpha_t, \omega^2] \tag{B.3.6}$$

so that these can be treated as x and μ, then multiplying these two distributions results in a normal distribution with expectation

$$\tilde{a} = \frac{\frac{\alpha_{t+1}}{\omega^2} + \frac{\alpha_{t-1}}{\omega^2}}{\frac{1}{\omega^2} + \frac{1}{\omega^2}} = \frac{\alpha_{t+1} + \alpha_{t-1}}{2}$$

Multiplying this distribution by

$$y_i - \delta_j \sim N[\alpha_t, \sigma^2]$$

I obtain the conditional posterior distribution of α_t, which is a normal distribution with mean

$$\left(\frac{y_i - \delta_j}{\sigma^2} + \frac{\alpha_{t+1} + \alpha_{t-1}}{2}\right)\left(\sigma^2 + \frac{\omega^2}{2}\right) \tag{B.3.7}$$

and variance

$$\sigma^2 + \frac{\omega^2}{2} \tag{B.3.8}$$

Now, the house effects. The parent distribution of δ_j is its prior given in Equation **??**, and the childre of δ_j is all the poll results published by polling house j. Let P_j be the set of all polls published by polling house j. Given that the conditional posterior distribution for the house effects δ_j is given by

$$f(\delta_j|\mathcal{G} \setminus \delta_j) \propto f(\delta_j; d^2) \prod_{i \in P} f(y_i; \mu_{t,i}, \sigma^2)$$
$$\propto \exp\left\{-\frac{1}{2d^2_{j,i}}\delta^2\right\} \exp\left\{-\frac{1}{2}\sum_{i \in P} \frac{y_i - \alpha_{t,i} - \delta_{j,i}}{\sigma_i^2}\right\}$$

With the same derivations as carried out below this gives a posterior distribution with mean

$$\left(\sum_{i \in P} \frac{y_i - \alpha_{t,i}}{\sigma^2} + \frac{0}{d^2}\right)\left(\sum_{i \in P} \sigma_i^2 + \frac{1}{d^2}\right) = (1 + d^2)\sum_{i \in P}(y_i - \alpha_{t,i}) \tag{B.3.9}$$

and variance

$$\sum_{i \in P} \sigma_i^2 + d^2 \tag{B.3.10}$$

Lastly, the parameter giving the day-to-day variability in the latent state, namely $\phi = 1/\omega^2$. This parameter was ascribed a gamma prior. As seen in Figure B.1 the children of ϕ are all the α_t parameters. The posterior distribution of ϕ is then

$$
\begin{aligned}
f(\phi|\mathcal{G} \setminus \phi) &= f(\phi) \prod_{t=2}^{T} f(\alpha_t|\alpha_{t-1}, \phi) \\
&\propto \phi^{a-1} \exp(-b\phi) \prod_{t=2}^{T} \frac{\phi^{1/2}}{\sqrt{2\pi}} \exp\left[-\frac{\phi}{2}(\alpha_t - \alpha_{t-1})^2\right] \\
&\propto \phi^{a-1+1/2} \exp\left[-\phi(1/2 + b) \sum_{t=2}^{T} (\alpha_t - \alpha_{t-1})^2\right]
\end{aligned}
$$

which is also gamma distributed with scale parameter $a + 1/2$ and rate parameter $(1/2 + b) \sum_{t=2}^{T} (\alpha_t - \alpha_{t-1})^2$.

B.4 Implementation in JAGS

Here is the implementation of the Gamma-Normal model in JAGS.

```
model{
  for(i in 1:NPOLLS){
    A[i] ~ dnorm(alphaA[week[i]]+houseA[org[i]],precA[i])
    Sv[i] ~ dnorm(alphaSv[week[i]]+houseSv[org[i]],precSv[i])
    Sp[i] ~ dnorm(alphaSp[week[i]] + houseSp[org[i]],precSp[i])
    Krf[i] ~ dnorm(alphaKrf[week[i]] + houseKrf[org[i]],precKrf[i])
    V[i] ~ dnorm(alphaV[week[i]] + houseV[org[i]],precV[i])
    H[i] ~ dnorm(alphaH[week[i]] + houseH[org[i]],precH[i])
    Frp[i] ~ dnorm(alphaFrp[week[i]] + houseFrp[org[i]],precFrp[i])
  }
  for(i in 2:NPERIODS){
    alphaA[i] ~ dnorm(alphaA[i-1],phiA)
    alphaSv[i] ~ dnorm(alphaSv[i-1],phiSv)
    alphaSp[i] ~ dnorm(alphaSp[i-1],phiSp)
    alphaKrf[i] ~ dnorm(alphaKrf[i-1],phiKrf)
    alphaV[i] ~ dnorm(alphaV[i-1],phiV)
    alphaH[i] ~ dnorm(alphaH[i-1],phiH)
    alphaFrp[i] ~ dnorm(alphaFrp[i-1],phiFrp)
  }
  ## sum-to-zero constraint on house effects
  houseA[1] <- -sum(houseA[2:NHOUSES]);houseSv[1] <- -sum(houseSv[2:NHOUSES])
  houseSp[1] <- -sum(houseSp[2:NHOUSES]);houseKrf[1] <- -sum(houseKrf[2:NHOUSES])
  houseV[1] <- -sum(houseV[2:NHOUSES]);houseH[1] <- -sum(houseH[2:NHOUSES])
  houseFrp[1] <- -sum(houseFrp[2:NHOUSES])
  ## priors
  phiA ~ dgamma(5000,1)
```

```
  alphaA[1] ~ dunif(.15,.40)
  phiSv ~ dgamma(10000,1)
  alphaSv[1] ~ dunif(.05,.15)
  phiSp ~ dgamma(10000,1)
  alphaSp[1] ~ dunif(.02,.15)
  phiKrf ~ dgamma(10000,1)
  alphaKrf[1] ~ dunif(.02,.15)
  phiV ~ dgamma(10000,1)
  alphaV[1] ~ dunif(.02,.1)
  phiH ~ dgamma(7000,1)
  alphaH[1] ~ dunif(.1,.35)
  phiFrp ~ dgamma(7000,1)
  alphaFrp[1] ~ dunif(.1,.35)
  for(i in 2:NHOUSES){
    houseA[i] ~ dnorm(0,.01);houseSv[i] ~ dnorm(0,.01)
    houseSp[i] ~ dnorm(0,.01);houseKrf[i] ~ dnorm(0,.01)
    houseV[i] ~ dnorm(0,.01);houseH[i] ~ dnorm(0,.01)
    houseFrp[i] ~ dnorm(0,.01)
  }
}
```

Appendix C

The DM-model

The complications that one encounters when working with a Dirichlet-Multinomial model in a DLM setting are different from those of the Gamma-Normal model derived above. While the dynamical part of the Dirichlet-Multinomial model is less intuitive than for the former model, the derivation of the posterior distribution is very straightforward. In the following two sections I present the model and discuss my solution to using this model in a dynamic setting. Then I show the Python-script that implements the DM-model as a DLM.

C.1 The model and the posterior distribution

In Section 4.2 we saw that when using the Dirichlet-Multinomial model the observations are transformed to the actual number of respondents that said they had the intention to vote for party $k = 1, \ldots, 8$. Thus, the time-series is a sequence of count vectors

$$\mathbf{y}_t = (y_{t1}, y_{t2}, y_{t3}, y_{t4}, y_{t5}, y_{t6}, y_{t7}, y_{t8}) \tag{C.1.1}$$

where the elements take on integer values. The elements are the sum of respondents who said they would vote for a given party in a given week/month. For example, if three polls are fielded in the same week, with samples sizes equal to 1000, 1024 and 850, which report 15.6, 22.1 and 18 percent as the respective estimates for Frp, then $y_{t,Frp} = 1000 \times 0.156 + 1024 \times 0.221 + 850 \times 0.18 = 535$, where I have rounded off to the nearest integer. I restate the probability mass function of the Multinomial distribution here

$$p(y_{t1}, \ldots, y_{t8}) = \frac{n_t!}{y_{t8}! \cdots y_{t8}!} \alpha_{t1}^{y_{t1}} \cdots \alpha_{t8}^{y_{t8}} \tag{C.1.2}$$

For a full Bayesian specification of the model a prior must be assigned over the α_{tk}'s. The conjuagate prior distribution is the Dirichlet distribution with probability density function

$$\pi(\alpha_{t1}, \ldots, \alpha_{t8}) = \frac{\Gamma(b_{t1} + \cdots + b_{t8})}{\Gamma(b_{t1}) \cdots \Gamma(b_{t8})} \alpha_{t1}^{b_{t1}-1} \cdots \alpha_{t8}^{b_{t8}-1} \tag{C.1.3}$$

where $\sum_{k=1}^{8} \alpha_{tk} = 1$ for all t. The posterior distribution is

$$f(\alpha_{t1}, \ldots, \alpha_{t1} | \mathbf{y}_t) \propto \alpha_{t1}^{y_{t1}} \cdots \alpha_{t8}^{y_{t8}} \alpha_{t1}^{b_{t1}-1} \cdots \alpha_{t8}^{b_{t8}-1} = \alpha_{t1}^{b_{t1}+y_{t1}-1} \cdots \alpha_{t8}^{b_{t8}+y_{t8}-1} \tag{C.1.4}$$

which is also Dirichlet with parameters $b_{t1} + y_{t1}$. There are two nice features of this approach. First, the manner in which the count vectors are designed gives more weight to poll esitimates that are based on large samples. Second, since the variance of the Dirichlet distribution is decreasing in the size of its parameters, large sample sizes also result in tighter HDRs. In order to specify the Dirichlet-Multinomial model in a DLM setting there are two issues that need to be handled. First, contrary to the GN-model where the posterior estimate of α_t is readily used as prior at $t + 1$, things are not that straightforward with the DM-model. The challenge is simply that the parameters used to specify the Dirichlet prior distribution are the b_{tk}'s which are integers, while the estimates obtained when sampling from the Dirichlet posterior are the $\alpha_{tk} \in [0, 1]$. To have a prior at $t + 1$ the α_{tk}'s must be translated to sensible integer values. A possible solution, that would certainly work if I was not aiming at a moving target, would be to use the $b_{tk} + y_{tk}$'s as the parameters specifying my prior at $t+1$. But, since the targets of interest, the latent states, are evolving through time this solution is only feasible if the value of the $b_{tk} + y_{tk}$'s are modelled as decaying with time. I propose a pragmatic solution: I assume that my confidence in the location of the latent states is never lower than if I had just surveyed thousand respondents. The α_{tk}'s sampled from the posterior at t are therefore multiplied by a thousand to become the parameters of the Dirichlet prior at $t + 1$. The scheme is simply,

$$(b_{t+1,1}, \ldots, b_{t+1,8}) = (\alpha_{t1}, \ldots, \alpha_{t8}) \times 1000$$

which means that for the points in time where I lack observations the parameters of the posterior are equal to the parameters of the prior, save for the random-walk. In Figure A.1, A.2 and A.3 it is easy to spot these points or periods of time where observations are lacking.

C.2 Implementing the DM-model in Python

Here is the Python script I have written to run the Dirichlet-Multinomial model.

```
import random as r
import numpy as np
def multidiri(data,itrs,pseudonobs):
  m = lambda x: np.linspace(0,0,len(data.T)*itrs).\
  reshape(itrs,len(data.T))
  A =m('.');Sv=m('.');Sp =m('.')
  Krf =m('.');V =m('.');H =m('.');Frp=m('.')
  others=m('.')
  i = 0;g = lambda a,b:r.gammavariate(a,b)
  h = lambda l,u:r.randint(l,u);avg=np.zeros(8)
  while i < itrs:
    prior = np.zeros(8)
    prior[0:8] = h(250,350),h(50,100),h(50,100),h(50,100),\
    h(30,62),h(100,160),h(150,250),h(20,140)
    sample = np.zeros(8);s=np.zeros(8)
    for week in range(0,len(data.T)):
      if week == 1:
        post = data.T[week][1:] + prior
      else:
        post = data.T[week][1:] + sample*pseudonobs
      sample[0:8] = g(post[0],1),g(post[1],1),g(post[2],1),\
            g(post[3],1),g(post[4],1),g(post[5],1),\
            g(post[6],1),g(post[7],1)
      sample[0:8] = sample[0:8]/float(sum(sample))
      A[i][week]=sample[0]
      Sv[i][week]=sample[1];Sp[i][week]=sample[2]
      Krf[i][week]=sample[3];V[i][week]=sample[4];H[i][week]=sample[5];
      Frp[i][week]=sample[6];others[i][week]=sample[7];
      # random-walk
      s[0:8] = sample[0:8] + np.array([g(post[0],1),g(post[1],1),g(post[2],1),\
                                g(post[3],1),g(post[4],1),g(post[5],1),\
                                g(post[6],1),g(post[7],1)])
      sample[0:8] = s[0:8]/float(sum(s))
      if week==len(data.T)-1:
        avg += post

    i += 1;print 'iteration %g'%i
  print avg/float(itrs) # print Dirich params
  return {'A':A,'Sv':Sv,'Sp':Sp,'Krf':Krf,'V':V,\
        'H':H,'Frp':Frp,'others':others}
```

Appendix D

Quadratic function

Here is the **R** code for the quadratic forecast function. It takes as arguments the points of the estimated time series and from which point and how far the time series is extrapolated.

```
quadratic <- function(jusquici,ici,elec){
  y <- jusquici
  t <- ici
  X <-cbind(rep(1,t),1:t,(1:t)^2)
  beta <- solve(t(X)%*%X)%*%t(X)%*%y
  t <- elec
  X <-cbind(rep(1,t),1:t,(1:t)^2)
  yhat <- X%*%beta
  return(yhat[elec])
}
```

Appendix E

Multinomial logit regression model

E.1 Model and priors

With

$$\mathbf{y}_t \sim \text{Multi}(n_t, \alpha_{t1}, \alpha_{t2}, \ldots, \alpha_{t8})$$

where α_{ik} represents the probability of the k'th party, and $\sum_{k=1}^{8} \alpha_{tk} = 1$. I follow Gelman et al. (2004) and parametrize the model in terms of the of logarithm of the ratio of the probability of each category relative to a baseline category. The categories are the seven parties (plus others), where I have set Arbeiderpartiet (Ap) as the baseline. I label the baseline as $k = 1$.

$$\ln\left(\frac{\alpha_{tk}}{\alpha_{t1}}\right) = \eta_{tk} = (\mathbf{X}\boldsymbol{\beta}_k)_t$$

where $\boldsymbol{\beta}_j$ is the vector of parameters for the j'th party. The distribution of the data is then

$$p(\mathbf{y}|\boldsymbol{\beta}) \propto \prod_{t=1}^{10} \left(\frac{e^{\eta_{t1}}}{\sum_{k=1}^{8} e^{\eta_{tk}}}\right)^{y_{t1}} \cdots \left(\frac{e^{\eta_{t8}}}{\sum_{k=1}^{8} e^{\eta_{tk}}}\right)^{y_{t8}} \tag{E.1.1}$$

where $\boldsymbol{\beta}_1$ is set to zero. The priors over the $\boldsymbol{\beta}_k$'s are

$$\boldsymbol{\beta}_k \sim N[\mathbf{b}_k, \boldsymbol{\Sigma}_k] \tag{E.1.2}$$

The covariance matrix $\boldsymbol{\Sigma}_k$ is the same for eight categories. It is matrix with 0.01 on the diagonal and zeros off the diagonal. The $\mathbf{b}_k$ vectors are specific for each of the eight categories. My priors for these are

$$\begin{aligned}
\mathbf{b}_{\text{Sv}} &= (0, .2, .1, .1, -.2)\\
\mathbf{b}_{\text{Sp}} &= (0, 0, .04, .04, 0)\\
\mathbf{b}_{\text{Krf}} &= (0, 0, -.2, -.2, .1)\\
\mathbf{b}_{\text{V}} &= (0, -.2, .2, .1, .1)\\
\mathbf{b}_{\text{H}} &= (0, -.4, .4, .4, .1)\\
\mathbf{b}_{\text{Frp}} &= (0, -.4, .4, .4, .1)
\end{aligned}$$

with all zeros in the vector for the category "others". The elements are in the order of the variables in the JAGS-script below. To estimate this model the Gibbs sampler was run for 2 050 000 iterations where the first 50 000 were discarded as burn in. Each 200'th iteration of the remaining two millon were kept for inference. Visual inspection of the trace plot of each of the parameters indicated that the Markov Chain produced by the Gibbs sampler had converged on its stationary distribution. No formal tests were conducted. The coefficients displayed in Table E.1 are the means and the 2.5 and 97.5 percentiles of the sampled values.

E.2 Implementation in JAGS

Here is how to implement the Multinomial logit regression model in JAGS.

```
model{
  for(i in 1:NOBS){
    for(j in 1:8){
      mu[i,j] <-  beta[j,1]
      + beta[j,2]*unempl[i]
      + beta[j,3]*cpigrwth[i]
      + beta[j,4]*gdpgrwth[i]
      + beta[j,5]*left[i] # left in gov.
  emu[i,j] <- exp(mu[i,j])
    }
    for(e in 1:8){
      p[e,i] <- emu[i,e]/sum(emu[i,1:8])
    }
    y[1:8,i] ~ dmulti(p[1:8,i], N[i])
  }
  # priors
  for(e in 1:5){
```

```
    beta[1,e] <- 0 # identifying restriction.
  }
  # separate priors for left and right parties
  for(k in 2:8){
    beta[k,1:5] ~ dmnorm(b0[1:5,k],E0)  # priors are specified in the data
  }
}
```

	Sv	Sp	Krf	V	H	Frp	others
Intercept	-0.89	-2.0378	-1.1619	-2.023	-0.6527	0.642	-1.6673
	[-0.924,-0.8473]	[-2.0759,-1.9996]	[-1.199,-1.1247]	[-2.077,-1.9697]	[-0.679,-0.6259]	[0.6062,0.6784]	[-1.7255,-1.6083]
Unemployment	-0.0612	0.1612	-0.2057	-0.0574	0.0478	-0.251	-0.1925
	[-0.0705,-0.0517]	[0.1526,0.1696]	[-0.2144,-0.1969]	[-0.07,-0.0451]	[0.0415,0.0542]	[-0.2597,-0.2419]	[-0.2069,-0.1783]
CPI growth	-0.0624	0.0078	-0.1097	-0.0878	0.0872	-0.2896	-0.1307
	[-0.0675,-0.0574]	[0.0029,0.0124]	[-0.1142,-0.1053]	[-0.0944,-0.0811]	[0.0839,0.0906]	[-0.2946,-0.2846]	[-0.1386,-0.1231]
GDP growth	-0.0095	0.0102	0.0662	0.0356	-0.0398	0.0015	0.022
	[-0.0127,-0.0062]	[0.0068,0.0136]	[0.0632,0.0692]	[0.0312,0.04]	[-0.0422,-0.0375]	[-0.0011,0.0041]	[0.0173,0.0265]
Left dummy	-0.1013	-0.0149	0.6031	0.1638	-0.1962	0.2118	0.3305
	[-0.1276,-0.0749]	[-0.0415,0.0113]	[0.5802,0.6262]	[0.1276,0.1979]	[-0.215,-0.1787]	[0.1876,0.236]	[0.2926,0.3679]

Table E.1: Multinomial logit regression model with Ap as baseline. Estimated on the last ten elections. Parentheses give the 95% HDRs of the posterior density of each parameter.

Appendix F

Implementing Saint-Laguës method

The Python class that calculates Saint-Laguës method is looped over a list of lists that contains the percentage shares of the seven (plus) one parties. In addition I supply, a data set containing the county wise deviances, data on the turnout in each county for the election of 2009, and the number of mandates from each county for the election of 2013.

F.1 Python script

```
import random as r
import numpy as np
class StLagues:
  def __init__(self,fylkes_forecast,eligible_voters,fylke_name):
    self.fylke_name = fylke_name
    fylkes_forecast = np.zeros(8) + fylkes_forecast
    fylkes_forecast = fylkes_forecast*eligible_voters
    self.fylkes_forecast = fylkes_forecast
  def distribute_them(self,mandates,to1,by1,to2,by2):
    self.to1,self.by1,self.to2,self.by2=to1,by1,to2,by2
    self.mandates = mandates
    fylkes_forecast = self.fylkes_forecast
    divisors = np.linspace(1,to1,by1)
    divisors[0] += .4
    table = []
    for i in range(0,len(divisors)):
      for j in range(0,len(fylkes_forecast)):
        table.append(fylkes_forecast[j]/float(divisors[i]))
      count = 1
    got_them=[]
  def index(table):
    for i,j in enumerate(table):
      if j == max(table):
        got_one = i
```

```
      return got_one
    while count <= self.mandates:
      which_one = index(table)
      got_them.append(which_one)
      table[which_one] = 0
      count += 1
    ap=0;sv=0;sp=0;krf=0;
    v=0;h=0;frp=0;others=0
    for x in got_them:
      if x in np.linspace(0,to2,by2):
        ap += 1
      elif x in np.linspace(0,to2,by2)+1:
        sv += 1
      elif x in np.linspace(0,to2,by2)+2:
        sp += 1
      elif x in np.linspace(0,to2,by2)+3:
        krf += 1
      elif x in np.linspace(0,to2,by2)+4:
        v += 1
      elif x in np.linspace(0,to2,by2)+5:
        h += 1
      elif x in np.linspace(0,to2,by2)+6:
        frp += 1
      else:
        others += 1
    print """ %s
      Ap = %s, Sv = %s, Sp = %s, Krf = %s
      V = %s, H = %s, Frp = %s, others = %s
      """%(self.fylke_name,ap,sv,sp,krf,v,h,frp,others)
    return ap,sv,sp,krf,v,h,frp,others
class deviance_fylke:
  def __init__(self,nationalforecast,fylkes_factors,all19):
    self.all19 = all19
    self.fylkes_factors = fylkes_factors
    self.nationalforecast=np.zeros(8)+nationalforecast
  def get_fylkesvis(self,fylke_name):
    self.fylke_name = fylke_name
    nationalforecast=self.nationalforecast
    def tuple_mult(a,b):
      if len(a) != len(b):
        print 'Incompatible tuples!'
      out = np.zeros(8)
      for i in range(0,len(a)):
        out[i] = a[i]*b[i]
      return out
    for s in range(0,19):
      if self.fylke_name == self.all19[s]:
        fylkes_forecast = tuple_mult(self.fylkes_factors[s]\
                                     ,nationalforecast)
```

```
    return fylkes_forecast

class equalization(StLagues):
  def __init__(self,mandDict,all19,names_in_order,nationalforecast):
    self.mandDict,self.all19=mandDict,all19
    self.names_in_order=names_in_order
    self.nationalforecast=nationalforecast
  def how_many(self,totalmandates):
    self.totalmandates = totalmandates
    mandDict=self.mandDict
    names_in_order=self.names_in_order
    ap=0;sv=0;sp=0;krf=0;v=0;h=0;frp=0;others=0
    for fylke in names_in_order:
      ap+=mandDict[fylke][0];sv+=mandDict[fylke][1]
      sp+=mandDict[fylke][2];krf+=mandDict[fylke][3]
      v+=mandDict[fylke][4];h+=mandDict[fylke][5]
      frp+=mandDict[fylke][6];others+=mandDict[fylke][7]
    CountyMandates=[ap,sv,sp,krf,v,h,frp,others]
    self.countymand=[ap,sv,sp,krf,v,h,frp,others]
    tobedist=self.totalmandates
    for e in range(0,len(self.nationalforecast)):
      if self.nationalforecast[e] < .04:
        tobedist-=CountyMandates[e]
        self.nationalforecast[e]=0
        CountyMandates[e] = 0
    negs = 'negative diffs'
    while negs == 'negative diffs':
      StLagues.__init__(self,self.nationalforecast,2682948,'National')
      StLaguesITER=np.zeros(8)+StLagues.distribute_them(self,\
                        tobedist,201,101,800,101)
      self.diffs = StLaguesITER-CountyMandates;negdiffs=0
    for j in range(0,len(self.diffs)):
      if self.diffs[j]<0:
        tobedist -= CountyMandates[j]
        self.nationalforecast[j]=0
        CountyMandates[j]=0
        negdiffs+=1
      if negdiffs==0:
        negs='No negative diffs'
  def total(self):   # adding them together
    diffs = self.diffs
    county = self.countymand
    return county+diffs
```