

# Consideration Sets, Intentions and the Inclusion of “Don’t Know” in a Two-Stage Model for Voter Choice

Richard Paap, Erjen van Nierop, Harald J. van Heerde,  
Michel Wedel, Philip Hans Franses, Karel Jan Alsem<sup>1</sup>

Econometric Institute Report EI2000-33/A

<sup>1</sup>Richard Paap and Erjen van Nierop are, respectively, Postdoc and Ph.D. student, Faculty of Economics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands (email: paap@few.eur.nl and vannierop@few.eur.nl); Harald J. van Heerde is Assistant Professor, Tilburg University, Faculty of Economics, P.O. Box 90153, 5000 LE Tilburg, The Netherlands (email: heerde@kub.nl); Michel Wedel is Professor, University of Groningen, Faculty of Economics, P.O. Box 800, 9700 AV Groningen, The Netherlands (email: m.wedel@eco.rug.nl) and Visiting Professor, University of Michigan, U.S.A. (Wedel@Umich.edu); Philip Hans Franses is Professor, Faculty of Economics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands (email: franses@few.eur.nl); Karel Jan Alsem is Associate Professor, University of Groningen, Faculty of Economics, P.O. Box 800, 9700 AV Groningen, The Netherlands (email: k.j.alsem@eco.rug.nl). The authors thank Neal Beck and Rob Eisinga with their help concerning the political science literature. Additionally, the first author thanks the Netherlands Organization for Scientific Research (N.W.O.) for its financial support.

## **Abstract**

We present a statistical model for voter choice that incorporates a consideration set stage and final vote intention stage. The first stage involves a multivariate probit model for the vector of probabilities that a candidate or a party gets considered. The second stage of the model is a multinomial probit model for the actual choice. In both stages we use as explanatory variables data on voter choice at the previous election, as well as socio-demographic respondent characteristics. Importantly, our model explicitly accounts for the three types of “missing data” encountered in polling. First, we include a no-vote option in the final vote intention stage. Second, the “do not know” response is assumed to arise from too little difference in the utility between the two most preferred options in the consideration set. Third, the “do not want to say” response is modelled as a missing observation on the most preferred alternative in the consideration set. Thus, we consider the missing data generating mechanism to be non-ignorable and build a model based on utility maximization to describe the voting intentions of these respondents. We illustrate the merits of the model as we have information on a sample of about 5000 individuals from the Netherlands for who we know how they voted last time (if at all), which parties they would consider for the upcoming election, and what their voting intention is. A unique feature of the data set is that information is available on actual individual voting behavior, measured at the day of election. We find that the inclusion of the consideration set stage in the model enables the user to make more precise inferences on the competitive structure in the political domain and to get better out-of-sample forecasts.

Key words: Choice model, Probit model, Election data, Polling, Bayesian method

# 1 INTRODUCTION

Modeling and forecasting voter choice is a key topic in the political science literature. The models may lead to an understanding of the determinants of voting behavior, while the forecasts based on these models can for example be used to modify campaign strategies. Given the use of the models for such an important topic, it should not come as a surprise that there is abundant literature on the design and implementation of models for voter choice. For example, a recent special issue of the *International Journal of Forecasting* (1999, volume 15, issue 2) contained a variety of articles on models that can be useful to predict the outcomes of elections. Typically, the models take stated intentions, previous choice and voter-specific characteristics as explanatory variables, where the data can be available at the individual voter level or at a certain level of aggregation. Techniques that are commonly used range from the familiar multinomial logit and probit models for individual data to ecological inference methods for grouped data.

The overall conclusion from the literature is that there is no single method that is to be preferred as it outperforms all other methods. There may be various reasons why predictions turn out to be wrong or inaccurate. One reason may be that the forecasting methodology itself is inadequate. Another reason may be that in between elections, and perhaps also in the period in between stated intentions and final elections, economic or other conditions change, see Fair (1996). Note that this possibility touches upon the very nature of forecasting, which is that it is usually assumed that the past can somehow be extrapolated into the future. A third possible reason for inaccurate predictions, which has received quite some attention recently, concerns the individuals who state not to know who to vote for or who intend not to vote at all, as discussed in particular by Lynn and Jowell (1996). In fact, Sanders (1998) convincingly shows that any forecasting model that does not take those individuals into account is misspecified.

There have been various attempts in the political science literature to incorporate the decision to vote in multinomial choice models, see Dubin and Rivers (1990) and also Sanders (1998), among others. These models usually contain a model component that can handle only two or three voting options, preceded by a component that takes care of the sample selection issue, that is, the notion that, simply, no vote is observed for those individuals who do not vote. The currently available models suffer from various drawbacks. The first is that these models are difficult to extend to the case of (many) more than, say, three voting options, where such a large amount of options is quite common for several European countries, for example. Additionally, the models are difficult to extend if one would want to allow for other types of missing data, which may be generated by individuals who state that they do not know which party or candidate they would favor, or those individuals who do not want to tell about their voting intentions.

Thirdly, and perhaps most importantly, the typical view on non-voters and the like is to treat these as missing cases, for which one should correct in a second stage of modeling. In other words, the non-voters cause some form of selection bias for the model of interest. However, as it is sometimes suggested, see Gilljam and Granberg (1993) and Lynn and Jowell (1996) among others, it seems worthwhile to examine the behavior of non-voters and try to include the relevant observations into a forecasting model. Indeed, it may be that non-voters are informative in the context of a forecasting model, that is, they can have predictive value for the final outcome.

The key problem for the inclusion of non-voters, that is, those who do not know and those who do not want to say into a forecasting model is of course the fact that there is no answer to the question concerning the voting intention for these individuals. This may become even more problematic if the same individual indeed did not vote at the previous election, and perhaps indeed does not vote at the current election. One may resort to various imputation techniques, see for example Gelman, King, and Liu (1998) and Gelman and King (1995) for recent approaches to item nonresponse based on complete and available cases and multiple imputation. However, another viable strategy may be to ask for additional information in the relevant questionnaire. For example, one may ask individuals to indicate which parties or candidates they might consider voting for, and next, one can ask which of these (if at all) would be preferred (see, for example, Yu and Lam 1997). It is known from the marketing and econometrics literature, see Siddarth, Bucklin and Morisson (1995) and Chiang, Chib and Narasimhan (1999), that the introduction of an additional layer in a choice model can lead to better predictions of the actual choice. Consideration set formation is supported by consumer behavior theory (Bettman 1979), that indicates that consumers are likely to make decisions from larger sets of alternatives in more than one stage. Taking this notion to voting behavior, one may also view the voting process as a multi-stage process, where the final choice is made based on a set of alternatives which is presumably smaller than the set containing all choice options. Hence, it is of interest to include questions on consideration sets in the questionnaire, also since these may receive less “do not know” or “do not want to say” responses. If so, then the information in the consideration set can have a predictive value for the final choice.

In this paper we present a statistical model for voter choice that incorporates a consideration set stage and final vote intention stage. The first stage involves a multivariate probit [MVP] model for the vector of probabilities that a candidate or a party gets considered (Chib and Greenberg 1998, van Nierop et al. 2000). The second stage of the model is a multinomial probit [MNP] model for the actual choice, see Alvarez and Nagler (1998a, 1998b), Lacy and Burden (1999), Quinn, Martin, and Whitford (1999) and Schofield, Martin, Quinn and Whitford (1998) for

other applications of this MNP model in political science. In both stages we use as explanatory variables data on voter choice at the previous election, as well as socio-demographic respondent characteristics. Importantly, our model explicitly accounts for the three types of “missing data” encountered in polling. First, we include a no-vote option in the final vote intention stage. Second, the “do not know” response is assumed to arise from too little difference in the utility between the two most preferred options in the consideration set. Third, the “do not want to say” response is modelled as a missing observation on the most preferred alternative in the consideration set. Thus, we consider the missing data generating mechanism to be non-ignorable and build a model based on utility maximization to describe the voting intentions of these respondents.

In this paper we illustrate the merits of the model as we have information on a sample of about 5000 individuals from the Netherlands for who we know how they voted last time (if at all), which parties they would consider for the upcoming election, and what is their voting intention. A unique feature of the data set is that information is available on actual individual voting behavior, measured at the day of election. This enables us to validate our approach at the individual level and to reveal the gain in predictive accuracy relative to a few other individual level forecasting approaches.

The outline of our paper is as follows. In Section 2 we put forward our two-stage model, which explicitly incorporates the possibility that individuals “do not vote”, “do not know” or “do not want to say”. In Section 3 we give a sketch of a Bayesian method for parameter estimation, where we relegate more technical details to the Appendix. In Section 4 we provide a description of the data and of the political context in the Netherlands. In Section 5 we report on the empirical results and we provide a discussion of the main conclusions that can be drawn from these results. Finally, in Section 6, we conclude our paper with a few remarks.

## 2 MODEL

### 2.1 Preliminaries

In this section we propose a two-stage choice model to describe the vote behavior of an individual  $i$  ( $i = 1, \dots, I$ ), who faces a choice between  $J$  political parties, where we make this behavior dependent on intentions and considerations. We denote the stated vote intention prior to the election by the variable  $d_i$ , which can take  $J + 3$  different values. If  $d_i = j$  with  $j \leq J$  the individual has a single most preferred party. If  $d_i = J + 1$  the individual does not intend to vote, while  $d_i = J + 2$  indicates that the individual does not know yet and  $d_i = J + 3$  indicates that the

individual does not want to say his or her vote intention. Individuals are assumed not to consider all parties in their decision process, but to choose one party from a particular subset of the parties at a particular point in time. This consideration or choice set may contain any combination of available parties. Hence, for each individual there are  $Q = 2^J$  potential consideration sets. Note that an individual may have an empty consideration set, which implies that he or she does not consider any party at all. In that case the outcome of the voting intention  $d_i$  may be “do no vote” ( $J + 1$ ) or “do not know yet” ( $J + 2$ ) or even “do not want to say” ( $J + 3$ ).

## 2.2 Two-stage Model

The goal of this paper is to construct a choice model for the vote intention of the individuals stated prior to the elections and to use this model to predict the actual vote of the individuals at the election. The model we propose consists of two components. In the first stage we describe the consideration set of an individual. In the second stage we model the actual choice among the parties in the consideration set of this individual.

**Consideration set formation** The stated consideration set of an individual  $i$ ,  $c_i$ , can be represented by a  $J$ -dimensional vector containing binary variables

$$c_i = \begin{pmatrix} c_{i1} \\ \vdots \\ c_{iJ} \end{pmatrix}, \quad (1)$$

where  $c_{ij} = 1$  if party  $j$  is in the consideration set of individual  $i$ , and 0 otherwise, see van Nierop et al. (2000) for a similar approach. To describe the inclusion of a party in the consideration set of an individual  $i$ , we use a multivariate probit model [MVP]

$$C_{ij}^* = x_i' \alpha_j + I[d_i^{-1} = j] \beta + \varepsilon_{ij}, \quad j = 1, \dots, J, \quad (2)$$

where  $C_{ij}^*$  is the latent variable modeling the inclusion,  $x_i$  is a  $k$ -dimensional vector of explanatory variables including for example age and gender,  $\alpha_j$  a  $k$ -dimensional parameter vector and  $\varepsilon_{ij}$  is a disturbance term. The term  $I[d_i^{-1} = j]$  is a dummy variable that equals 1 if the actual vote of the individual in the previous election is party  $j$  and zero elsewhere. The parameter  $\beta$  therefore models the effect of state-dependence.

Party  $j$  enters the consideration set of individual  $i$ , that is  $c_{ij} = 1$ , if  $C_{ij}^* > 0$ . The probability that this party is included depends on the distribution of  $\varepsilon_{ij}$ . We assume that the vector of disturbances  $\varepsilon_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,J})'$  is normally distributed, that is,

$$\varepsilon_i \sim N(0, \Sigma). \quad (3)$$

The parameters in the model are theoretically identified if we restrict one of the diagonal elements of  $\Sigma$  to be 1. In our empirical illustration below, it turns out that the cross restriction of the state-dependence parameter  $\beta$  across the equations is not strong enough for empirical identification and hence we impose the restriction that the diagonal elements of  $\Sigma$  are all equal to 1.

**Party choice** Given the consideration set of individual  $i$ , we describe his or her party choice by a multinomial probit model. An individual chooses among the parties in his or her consideration set. Additionally, we assume that an individual may always decide not to vote at all and hence we impose that  $c_{i,J+1} = 1$  for all  $i$ . Individual  $i$  perceives utilities  $D_{ij}^*$ , for  $j = 1, \dots, J+1$ , with  $c_{ij} = 1$ , from voting for party  $j$  or for not voting ( $J+1$ ), that is,

$$D_{ij}^* = x_i' \gamma_j + I[d_i^{-1} = j] \delta + \eta_{ij}, \quad \text{for } j = 1, \dots, J+1 \text{ with } c_{ij} = 1, \quad (4)$$

where  $x_i$  is again a  $k$ -dimensional vector of explanatory variables,  $\gamma_i$  is a  $k$ -dimensional parameter vector and  $\eta_{ij}$  is a disturbance term. The  $\delta$  parameter describes again state-dependence. For identification purposes we impose that  $\gamma_{J+1}$  equals zero. Note that the explanatory variables may affect the consideration set membership and the choice in a different way.

We assume that individual  $i$  states to vote on party  $j$  ( $d_i = j$ ) if the perceived utility of voting on party  $j$  is more than  $\theta$  larger than the perceived utilities for voting for the other parties in the consideration set or for not voting. Hence, we assume that

$$D_{ij}^* > D_{il}^* + \theta \text{ for } l = 1, \dots, j-1, j+1, J+1 | c_{il} = 1. \quad (5)$$

An individual  $i$  does not vote ( $d_i = J+1$ ) if the perceived utility of not voting is more than  $\theta$  higher than the perceived utilities of the parties in the consideration set, that is,

$$D_{i,J+1}^* > D_{il}^* + \theta \text{ for } l = 1, \dots, J | c_{il} = 1. \quad (6)$$

To describe the utility structure of individuals who do not know yet ( $d_i = J+2$ ), we consider the maximum of the perceived utilities in the consideration set and the “no vote” option

$$D_{i,\max}^* = \max(D_{ij}^*, j = 1, \dots, J+1 | c_{ij} = 1) \quad (7)$$

and the value of the second largest perceived utility

$$D_{i,\max-1}^* = \max(D_{ij}^*, j = 1, \dots, J+1 | c_{ij} = 1 \wedge j \neq \max). \quad (8)$$

We assume that individual  $i$  states not to know which party to vote ( $j = J+2$ ) if the difference between the two largest perceived utilities is smaller than or equal to  $\theta$ , that is,

$$D_{i,\max}^* \leq D_{i,\max-1}^* + \theta. \quad (9)$$

Note that, as such, an individual may also exercise doubt between voting and non-voting. Finally, an individual may also not want to say the voting intention ( $d_i = J + 3$ ). In this case we do not impose any restrictions on the perceived utilities.

The probabilities of the outcome of the voting intention depend on the distributional assumptions on  $\eta_{ij}$ . We assume that the vector of the disturbances  $\eta_i = (\eta_{i1}, \dots, \eta_{iJ})'$  is normally distributed, that is,

$$\eta_i \sim N(0, \Omega). \quad (10)$$

Although Keane (1992) shows that the off-diagonal elements of the covariance matrix in a multinomial probit [MNP] model are empirically identified if we have a parameter that is the same across the  $J + 1$  equations, it turns out to be that the cross restriction on the state-dependence parameter  $\delta$  parameter is again not strong enough for empirical identification. Hence, we assume that  $\Omega$  is a diagonal matrix and impose that  $\Omega_{J+1, J+1} = 1$  for theoretical identification.

### 3 ESTIMATION AND FORECASTING

#### 3.1 Estimation

To estimate the model parameters in our two-stage model, we consider the likelihood function for the stated considerations and party choices of the individuals,  $c = \{c_i, i = 1, \dots, I\}$ ,  $d = \{d_i, i = 1, \dots, I\}$ , that is,

$$\mathcal{L}(c, d | \Phi) = \prod_{i=1}^I \Pr[d_i | c_i, \gamma, \delta, \theta, \Omega] \Pr[c_i | \alpha, \beta, \Sigma], \quad (11)$$

where  $\Phi = (\alpha, \beta, \Sigma, \gamma, \delta, \theta, \Omega)$  with  $\alpha = (\alpha_1 \dots \alpha_J)$  and  $\gamma = (\gamma_1 \dots \gamma_J)$ . The likelihood function involves the product of the probability that the consideration set of individual  $i$  is  $c_i$  and the probability that the party choice is  $d_i$  given  $c_i$  for all individuals, and thus for a given individual the product of a multivariate and a multinomial probit probability.

To estimate the model parameters  $\Phi$  we opt for a Bayesian approach. We assume a flat prior distribution for the model parameters  $\Phi$  and use the Gibbs sampling approach of Geman and Geman (1984) to obtain posterior results. In each step of the Gibbs sampler, we draw the model parameters  $\alpha, \beta, \Omega, \gamma, \delta, \theta$  and  $\Sigma$  from their full conditional posterior distribution. The unobserved utilities  $D_{ij}^*$  and  $C_i^*$  are sampled alongside the model parameters (data augmentation), see Tanner and Wong (1987). For details on a Gibbs sampling approach in multivariate and multinomial probit models, we refer to Albert and Chib (1993), McCulloch and Rossi (1994), Chib and Greenberg (1998), and McCulloch, Polson and Rossi (2000). In the Appendix we



provide a short outline of the derivations of the full posterior distributions of the model parameters and unobserved utilities.

### 3.2 Forecasting

The goal of this paper is to model the vote intention of individuals prior to the election and to use the estimated model to predict the actual voting behavior of the individuals in the upcoming election. As we are dealing with choice models, we obtain the predictive probabilities for all possible realizations of  $d_i$  given  $x_i$  and the actual vote in the previous election. These predictive probabilities are given by

$$\Pr[d_i] = \int_{\Phi} \Pr[d_i|c_i, \gamma, \delta, \theta, \Omega] \Pr[c_i|\alpha, \beta, \Sigma] p(\Phi|\text{data}) d\Phi, \quad (12)$$

where we integrate with respect to the posterior distribution of  $\Phi$ , denoted by  $p(\Phi|\text{data})$ , to deal with parameter uncertainty.

These predictive probabilities can easily be computed using the Gibbs output. For each draw of  $\Phi$  we generate  $C_{ij}^*$  according to (2) for  $j = 1, \dots, J$ . The latent  $C_{ij}^*$  provides us a consideration set for each individual and given these consideration sets we generate utilities  $D_{ij}^*$  according to (4). The party choices for each draw are determined according to the rules given above. The relative frequencies of the choices provide the predictive probabilities, where we use the stated intentions for those subjects for which those are available. If the vote intention is unknown, we use our model and parameter estimates to predict the vote. In forecasting, we do not reweight the sample. We are not interested in forecasts of the outcomes “does not want to say” and “does not know”, but we are in forecasting the option not to vote. In that case the choice simply corresponds to the largest utility.

## 4 BACKGROUND AND DATA DESCRIPTION

In this section we first discuss the political context of the Netherlands and next we provide some details on the data.

### 4.1 Political Context

The political context in this paper concerns the Dutch 1994 parliament elections. Every four years, the electorate in The Netherlands chooses a new parliament, consisting of 150 members. The number of representatives of each of the political parties in the parliament is simply determined by the share of votes for a certain party. Shifts in the preference distribution can have

substantial political consequences. To form the government a coalition of two or more parties has to represent 50% or more of the voters. Currently the four important political parties are CDA (Christian Democrats), D'66 (Democrats), PvdA (Social Democrats), and VVD (Liberals). The government of The Netherlands has formed a coalition between two or three of these four parties for the past several decades. At the 1989 elections, close to 90% of the votes were for these four parties. Pre-election polls, conducted by several market research agencies in the Netherlands, are used by the parties (amongst others) to investigate with which parties they can and would like to form a coalition and serve as the basis for modifying campaign strategies.

## 4.2 Data

Our data were collected by the market research agency Inter/View in April and June 1994 (elections were held in June 1994). We are interested in forecasting the election outcomes for the four major parties, CDA ( $j = 1$ ), PvdA ( $j = 2$ ), VVD ( $j = 3$ ) and D'66 ( $j = 4$ ). We combine the other parties into one "Other party" ( $j = 5$ ) category. In addition, we include the "Do not know" ( $j = 7$ ), category and the "No vote" ( $j = 6$ ) category. Finally, there is a small fraction of the respondents who do not want to say for which party they intend to vote. We add a separate category for this group. The data concern four (sets of) variables. We have the stated consideration set of individual  $i$  denoted by  $c_i$ , with  $c_i$  a  $J$ -dimensional 0/1 vector. Next, we have the stated party choice  $d_i$ , with  $d_i = j$  for  $j \leq J$  indicating a single most preferred party,  $j = J + 1$  indicating that the individual does not want to vote,  $j = J + 2$  indicating that the individual does not yet know, and  $j = J + 3$  indicating that the individual does not want to say.

As explanatory variables, individual-specific socio-economic and demographic variables are contained in the  $k$ -dimensional vector  $x_i$ . Based on previous literature (Schofield et al. 1998, Swank and Eisinga 1999) and preliminary analyses we selected the following seven explanatory variables, that is, Looking for a job (1=yes, 0=no), Home ownership (1=own house, 0=rental house), Gender (1=female, 0=male), Dwelling (1=countryside, 2=urbanized countryside, 3=small town, 4=city), Income (1=below modal, 2=modal, 3=above modal), Age (1=age  $\leq 34$ , 2= $34 < \text{age} \leq 54$ , 3=age  $> 55$ ), and Labor union membership (1=FNV, 2=CNV (Christian union), 3=Other, 4=none).

The actual voting behavior in 1989 is used as a predictor variable and enables us to assess state-dependence. Finally, we have the actual behavior in 1994, that is, the party that has been voted for at the election day in June 1994, assessed by a telephone interview among the sample on the day after the election. This variable is used for identifying the predictive validity of our approach.

We believe this stated voting behavior to be the best yardstick available for validation in the absence of the option of retrieving the real votes. The bias relative to actual vote behavior is expected to be small given its measurement right after the elections.

### 4.3 Descriptive Statistics of the Data

Table 1 summarizes the actual votes in 1989, the intentions in 1994, the consideration set 1994, the actual votes in 1994, as well as a cross-tabulation of the vote intention variable with the explanatory variables.

[Insert Table 1 about here]

A noticeable figure is that about 10% of the 4620 individuals does not vote in 1994. The profile of the political parties is as expected CDA (Christian democrats) seems to attract the home-owning, countryside-situated, modal-income earning, older respondents. D'66 (democrats) seems to represent home owners living in urban areas with above-modal wages. PvdA-voters (socialists) rent houses, have below-modal incomes, and are highly represented in the FNV (a socialist union). The VVD (liberals) appear to be home owners with high incomes.

It is of some interest to consider the profiles of the no-vote category, the do not know category, and the do not want to say category. The no-voters seem to belong to the young, renting, modal income category. This category could consider politics to be something quite distinct from their own interests. The “do not know” category is large and also shows a mixed profile. It is, however, noteworthy that in particular females have doubts for which party to vote. Finally, the “do not want to say” category does not have a particularly clear profile.

[Insert Table 2 about here]

Table 2 displays a cross-tabulation of the intended and actual vote. Only for 2289 (49.5%) of the individuals, the vote intention is the same as their actual vote, which may be considered as a rather low number. Additionally, 81% of the individuals who state that they will vote for PVDA will actually vote for PVDA. For CDA and VVD, this percentage is about 76%, while for D'66 it is only 61%. For the individuals who do not know what to vote (1289), PVDA (290 votes) and CDA (253 votes) are the most popular parties, while the other parties are roughly equally popular.

## 5 RESULTS AND COMPARISONS

Next to the two-stage model presented above, we estimate three simplified versions of it. The first, which we label model A, is a multinomial probit model for the stated vote intention. In this

model we use no consideration set information and omit all individuals  $i$  with  $d_i > 5$ . Note that this model does not predict the “no vote” option. In model B we include the options “no vote”, “do not know” and “do not want to say”. Hence in this model we have 6 choice categories (5 parties plus no vote), and the two remaining options are modelled as described in Section 2.2. This model describes and predicts the option “no vote” explicitly. The third model, model C, is again an extended version of the previous model. We include state-dependence with the dummy of party choice in 1989. As we do not have an actual vote for every individual in 1989, we include three extra explanatory dummy variables which take on the value 1 if the individual answered “do not know” (do not recall), “no vote”, or “do not want to say”, respectively. The fourth model, labeled model D, is our full-fledged two-stage model, in which we use the consideration sets knowledge and allow for all options, including state-dependence. In this case we use 2 methods for prediction, both of which use stated party intentions if available. The difference is in the use of the consideration set knowledge for respondents with  $j > 5$ . With the first method (D1) we predict the consideration sets for these respondents. Alternatively, in the second method (D2), we use stated consideration sets, unless the consideration set is unknown for a certain individual, in which case we predict it.

[Insert Table 3 about here]

[Insert Table 4 about here]

[Insert Table 5 about here]

The posterior means and posterior standard deviations of the model parameter of the two-stage model are given in Tables 3–5. Table 3 shows that the posterior means of the parameters capturing the effects of the explanatory variables on the consideration sets are much farther away from zero than twice their posterior standard deviations. Importantly, the state-dependence parameter  $\beta$  is positive and highly significant. Thus, the probability of including a party in the consideration set is positively affected by the fact a respondent voted this party in previous (1989) elections.

Table 4 shows that the explanatory variables are also significant predictors for vote intention. Again, the state-dependence parameter  $\delta$  is positive and significant. One of key parameters in our models is  $\theta$ . Its posterior mean indicates that if the absolute difference in the utilities of the two most preferred options (including either two parties or a party and “no-vote”) is smaller than 0.39, the respondent indicates not to know which party to vote for. Although not shown, we note that the posterior standard deviations decrease in the order of model D, C, B and A, where those of the D-models are substantially smaller than those of the others.

Table 5 shows the posterior mean and standard deviation of the covariance matrix  $\Sigma$  of the multivariate probit model for the consideration set formation. Notably, the 95% confidence interval for none of the off-diagonal elements covers the zero value. A positive (negative) off-diagonal element indicates (non-) substitutability of political parties *beyond that what the covariates account for*. Substitutes are, therefore, CDA and VVD (correlation = 0.20) and PvdA and D'66 (correlation = 0.12). Non-substitutes are, for example, PvdA and VVD (correlation = -0.15). The face validity of these findings is high given the Dutch political situation.

The four models and prediction methods are compared on their election forecasts. These are calculated by averaging the posterior predictive probabilities over all individuals. Table 6 displays for each of the models A through D the predicted aggregate election results (party shares). For comparison, we give the true election results (derived from the actual choices in 1994). In the final column we display mean squared forecast errors.

[Insert Table 6 about here]

Table 6 shows that models D1 and D2 have the lowest squared errors across the five alternatives. Second best is C, next B, and worst is A. Hence the most sophisticated models perform best, and each of their components contribute to better predictions. Model B improves on model A as it explicitly accounts for including individuals whose response is no-vote, do not know, or do not want to say. On top of that, the use of state-dependence variables in model C decreases the forecast error. And, if we finally add the consideration set stage as in model D, the prediction becomes even better. It seems that the use of predicted consideration sets as in model D1 is slightly better than the use of stated consideration sets (D2). Apparently, it is important to allow for choices outside the stated consideration set, which is accommodated for in the forecasts of the latter model.

As an aside, we note that all models underpredict PvdA's share and all models overpredict the share of D'66. Apparently, none of the models captures the switch from D'66-intention to PvdA-vote, which is also visible in Table 2: 97 respondents make this switch. Factors external to the models have caused this. Nevertheless, models D1 and D2 suffer the least. For these two models we computed the posterior mean forecast hit-rates: the percentage of time the model accurately predicts the actual vote across individuals in the sample. These are respectively 62.6% and 61.7%, showing that the two models forecast the votes at the individual level quite well.

We depict in Figure 1 the marginal predictive posterior densities for each of the six parties based on model D1. The posterior mass is quite concentrated around the mean.

[Insert Figure 1 about here]

Table 7 yields insight in the bivariate predictive distribution of the percentages for all six options. In this table we present the variance-covariance matrix of models C and D1. It is interesting to see how the exclusion of this stage affects the (co-)variances.

[Insert Table 7 about here]

Table 7 shows that the variances of the predictions decrease for model D1 relative to model C for all options except for the no-vote option. In addition, the covariance structure has changed in a manner consistent with the pattern observed in Table 5. That is, the model with the consideration set stage, model D1, has more “profiled” covariances, in that the covariances between party shares converge to zero for non-substitutable parties, and become more negative for substitutable parties. A strong example is CDA: in D1 relative to C, the covariance for CDA with non-substitutes such as D’66 and PvdA becomes less negative, whereas the covariance with a substitute such as VVD is more so. Similarly, VVD appears to be competing less with PvdA and D’66 in model D1 relative to C. We also note that PvdA and D’66 become less substitutable in D1. Apparently, these two parties tend to be included jointly in the consideration set (positive covariance in Table 5), but this affects the intention in the opposite direction. Thus, the inclusion of the consideration set stage in the model enables the user to make more precise inferences on the competitive structure in the political domain and to get better out-of-sample forecasts.

## 6 CONCLUSIONS AND LIMITATIONS

The prediction of outcomes of elections for political parties and candidates receives much interest in democracies across the world, an interest that is reflected intensive media coverage. Forecasts based on pre-election polls in addition have important political consequences, since they enable parties and candidates to assess their relative strengths and weaknesses as input to campaign strategy and coalition formation. In this paper we proposed a model for pre-election polling data to predict election outcomes. The model is based on a utility framework and explicitly accommodates consideration and choice stages of voting behavior. We find that the addition of the consideration set stage improves predictive accuracy and strongly decreases the posterior variance of the predictions. For the design of pre-election polls these results imply that it is important to include a question on the party voted at the previous elections as well as questions on the set of parties considered.

We find that the predictive capacity of the model is further improved by the inclusion of a state-dependence variable accounting for previous voting behavior as an explanatory variable. A final conclusion from our study is that it is important to account explicitly for “missing data”, that

is, for the responses “do not vote”, “do not know”, and “do not want to say”. Rather than imputing those missing responses based on available procedures, we have postulated a plausible behavioral mechanism for them that enables forecasts of votes of subjects in each of those categories. This significantly adds to the predictive validity of our model.

Our study also has several caveats. One limitation is that the “other party” option includes very heterogeneous parties, whereas our model assumes that the explanatory variables affect the choice for all parties within this set in the same way. However, we cannot split this set into all the parties it consists of, because the response parameters would be unidentified due to too few observations. Further, we choose to concentrate on within-sample forecasts, but forecasts of all procedures could also be projected to the population using weighting procedures. Further, we use stated votes, assessed through telephone interviews after the election, as a forecasting benchmark. Although we do believe that the bias in these stated votes is small, having the actual votes would be preferable. In spite of the fact that our model attained a high hit-rate of votes at the individual level, it could not predict a swing in choice between two parties from intentions to actual votes. Such a swing is caused by other factors than the demographic predictors used in our model, for example the media coverage of the pre-election poll results itself. The development of models to accommodate such effects is an important topic for future research. Finally, our model does not account for heterogeneity in the response parameters across individuals. Unfortunately, this is not possible as we have a vote intention at only one point in time. If we would have had access to intention data at multiple points in time we could have studied the effect of heterogeneity on the predictive performance and we could have modeled the trends in the vote intentions, which could also lead to better prediction of swings in the vote share. We leave these issues for future research.

# APPENDIX:

## FULL CONDITIONAL POSTERIOR DISTRIBUTIONS

In this appendix we provide a sketchy derivation of the full posterior distributions of the model parameters and unobserved utilities, which are needed in the Gibbs sampler.

**Sampling of  $C_i^*$**  Denote the vector of latent variables (utilities) for the consideration set by  $C_i^* = (C_{i1}^*, \dots, C_{iJ}^*)'$ . The model for these latent variables can be written in matrix form as

$$C_i^* = x_i' \alpha + I_i^{-1}[D_i] \beta + \varepsilon_i \quad (\text{A.1})$$

where  $I_i^{-1}[D_i]$  is a  $J$ -dimensional vector containing 0/1 dummy variables describing the state-dependence. The latent variables  $C_i^*$  are truncated multivariate normally distributed. The individual elements  $C_{is}^*$  can be sampled from their full conditional distribution on truncated regions:

$$C_{is}^* | C_{i,-s}^*, \alpha, \beta, \Sigma, \text{data} \sim \begin{cases} \text{normal on } (0, \infty) & \text{if } c_{is} = 1 \\ \text{normal on } (-\infty, 0] & \text{if } c_{is} = 0, \end{cases} \quad (\text{A.2})$$

for  $s = 1, \dots, J$  and where  $C_{i,-s}^* = (C_{i1}^*, \dots, C_{i,s-1}^*, C_{i,s+1}^*, \dots, C_{iJ}^*)'$ , see also Geweke (1991). For individuals who “do not want to say” which party/parties they consider, the latent consideration set utilities can be sampled from their unrestricted distributions according to (A.1).

**Sampling of  $\alpha$  and  $\beta$**  Model (A.1) can be seen as a multivariate regression model with regression parameters  $\alpha$  and  $\beta$ . The full conditional posterior of  $\alpha$  and  $\beta$  are therefore (matrix) normal.

**Sampling of  $\Sigma$**  To sample  $\Sigma$  we note that

$$\begin{aligned} p(\Sigma | C_i^*, \alpha, \beta, \text{data}) &\propto \pi(\Sigma | C_i^*, \alpha, \beta, \text{data}) \\ &= \exp\left(-\frac{1}{2} \sum_{i=1}^I (C_i^* - x_i' \alpha - I_i^{-1}[D_i] \beta)' (C_i^* - x_i' \alpha - I_i^{-1}[D_i] \beta)\right). \end{aligned} \quad (\text{A.3})$$

As  $\Sigma$  is not an unrestricted covariance matrix (the diagonal elements are 1), the full conditional distribution is not inverted Wishart. In fact the full conditional posterior distribution of  $\Sigma$  is not standard. To sample  $\Sigma$  we use a sampler based on Basag and Green (1993). Loosely speaking, this sampler switches the two steps in the Metropolis-Hasting sampler of Metropolis et al. (1953). We first draw  $u$  from a uniform distribution on the interval  $[0, 1]$ . In the second step we keep sampling candidate draws of the elements of  $\Sigma$  from a uniform distribution on the interval  $[-1, 1]$  until  $\pi(\Sigma^{\text{new}})/\pi(\Sigma^{\text{old}}) > u$ . The advantage of the latter approach is that it always results in a new drawing of  $\Sigma$ , which is not the case for the Metropolis-Hasting sampler.



**Sampling of  $D_i^*$**  Denote the vector of unobserved utilities by  $D_i^* = (D_{i1}^*, \dots, D_{i,J+1}^*)'$  which is generated by

$$D_i^* = x_i'(\gamma \ 0) + I_i^{-1}[D_i]\delta + \eta_i, \quad (\text{A.4})$$

where  $I_i^{-1}[D_i]$  is a  $J + 1$ -dimensional vector containing 0/1 dummy variables describing the state-dependence. Hence the full conditional posterior distribution of the vector of unobserved utilities  $D_i^*$  is truncated normal. To simplify notation we define the largest and second largest utility in the consideration set as

$$\begin{aligned} D_{i,\max}^* &= \max[D_{ij}^*, j = 1, \dots, J + 1 | c_{ij} = 1] \\ D_{i,\max-1}^* &= \max[D_{ij}^*, j = 1, \dots, J + 1 \wedge j \neq \max | c_{ij} = 1]. \end{aligned} \quad (\text{A.5})$$

Additionally, if want to sample the utility  $D_{is}$  we need the largest and second largest utility in the consideration set apart from utility  $D_{is}$

$$\begin{aligned} D_{i,-s}^* &= \{D_{ij}^*, j = 1, \dots, s - 1, s + 1, \dots, J | c_{ij} = 1\} \\ D_{i,\max-s}^* &= \max[D_{ij}^*, j = 1, \dots, J + 1 \wedge j \neq s | c_{ij} = 1] \\ D_{i,\max-s-1}^* &= \max[D_{ij}^*, j = 1, \dots, J + 1 \wedge j \neq \max-s \wedge j \neq s | c_{ij} = 1]. \end{aligned} \quad (\text{A.6})$$

To derive the full conditional posterior distributions, we consider three cases. The first case concerns the utilities for  $d_i = j$  observations for  $j \leq J + 1$ . We have to sample  $D_i^*$  from a truncated multivariate distribution such that  $D_{id_i}^* > D_{is}^*$  for all  $s$  with  $c_{is} = 1$  and  $D_{id_i}^* - D_{i,\max-1}^* > \theta$ . This can be done by sampling the separate elements of  $D_i^*$  from their full conditional posterior distribution

$$D_{is}^* | D_{i,-s}^*, \gamma, \delta, \Omega, \theta, \text{data} \sim \begin{cases} \text{normal on } (-\infty, D_{id_i}^* - \theta) & \text{if } s \neq d_i \\ \text{normal on } (D_{i,\max-s}^* + \theta, \infty) & \text{if } s = d_i. \end{cases} \quad (\text{A.7})$$

The second case concerns the utilities corresponding to the  $d_i = J + 2$  observations. In that case we have to sample  $D_i^*$  from a truncated multivariate distribution such that  $D_{i,\max}^* - D_{i,\max-1}^* \leq \theta$ . This can again be done by sampling the elements of  $D_i^*$  from their full conditional posterior distribution, that is,

$$D_{is}^* | D_{i,-s}^* \sim \begin{cases} \text{normal on } (-\infty, D_{i,\max-s}^* + \theta] & \text{if } D_{i,\max-s}^* - D_{i,\max-s-1}^* \leq \theta \\ \text{normal on } [D_{i,\max-s}^* - \theta, D_{i,\max-s}^* + \theta] & \text{if } D_{i,\max-s}^* - D_{i,\max-s-1}^* > \theta \end{cases} \quad (\text{A.8})$$

The final case concerns the utilities corresponding to the  $d_i = J + 3$  observations. In that case we can sample the latent utilities for the parties in the consideration set according to (A.4).

**Sampling of  $\gamma$  and  $\delta$**  Model (A.4) can be seen as a multivariate regression model with regression parameters  $\gamma$  and  $\delta$ . The full conditional posterior of  $\gamma$  and  $\delta$  are therefore normal. Note that one only has to consider the equations for which  $c_{ij} = 1$ .

**Sampling of  $\Omega$**  To sample the diagonal elements of the covariance matrix  $\Omega$  denoted by  $\Omega_{jj}$  for  $j = 1, \dots, J$  we use

$$p(\Omega_{jj}|\gamma, \delta, \text{data}) \propto \frac{1}{\Omega_{jj}\nu_j} \exp\left(-\frac{1}{2\Omega_{jj}^2} \sum_{i=1}^I I[c_{ij} = 1](D_{ij}^* - x_i'\gamma_j - I[d_i^{-1} = j]\delta)^2\right) \quad (\text{A.9})$$

and hence that

$$\frac{\sum_{i=1}^I I[c_{ij} = 1](D_{ij}^* - x_i'\gamma_j - I[d_i^{-1} = j]\delta)^2}{\Omega_{jj}^2} \sim \chi^2(\nu_j) \quad (\text{A.10})$$

with  $\nu_j = \sum_{i=1}^I I[c_{ij} = 1]$  for  $j = 1, \dots, J$ .

**Sampling of  $\theta$**  To sample  $\theta$  we consider the maximum of the differences between the two largest utilities of the “do not know” voters

$$lb = \max[D_{i,\max}^* - D_{i,\max-1}, i = 1, \dots, I | d_i = J + 2] \quad (\text{A.11})$$

and the minimum of the difference between the two largest utilities of the voters and no voters

$$ub = \min[D_{i,\max}^* - D_{i,\max-1}, i = 1, \dots, I | d_i \leq J + 1]. \quad (\text{A.12})$$

It is easy to see that the likelihood as a function of  $\theta$  is constant in the region between the two values. Therefore, we can sample

$$\theta \sim \text{Unif}(lb, ub). \quad (\text{A.13})$$

# REFERENCES

- Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, 88, 669-679.
- Alvarez, R. M., and Nagler, J. (1998a), "When Politics and Model Collide: Estimating Models of Multiparty Elections," *American Journal of Political Science*, 42, 55-96.
- Alvarez, R. M., and Nagler, J. (1998b), "Economics, Entitlements, and Social Issues: Voter Choice in the 1996 Presidential Election," *American Journal of Political Science*, 42, 1349-1363.
- Basag J., and Green P. J. (1993), "Spatial Statistics and Bayesian Computation", *Journal of the Royal Statistical Society B*, 55, 25-37.
- Bettman, J. R. (1979), *An Information Processing Theory of Consumer Behavior*, Reading, MA: Addison-Wesley.
- Chiang, J., Chib, S., and Narasimhan, C. (1999), "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity," *Journal of Econometrics*, 9, 223-248.
- Chib, S., and Greenberg, E. (1998), "Analysis of Multivariate Probit Models," *Biometrika*, 85, 347-361.
- Dubin, J. A., and Rivers, D. (1990), "Selection Bias in Linear Regression, Logit and Probit Models," *Sociological Methods and Research*, 18, 360-390.
- Fair, R. C. (1996), "Econometrics and Presidential Elections," *Journal of Economic Perspectives*, 10, 3, 89-102.
- Geman, S., and Geman, D. (1984), "Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Gelman, A., and King, G. (1995), "Why Are American Election Campaign Polls So Variable, When Votes Are So Predictable?," *British Journal of Political Science*, 23, 409-451.
- Gelman, A., King, G., and Liu, C. (1998), "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys," *Journal of the American Statistical Association*, 93, 846-857.
- Geweke, J. F. (1991), "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints", in *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, 571-578, Alexandria, VA: American Statistical Association.
- Gilljam, M., and Granberg, D. (1993), "Should We Take Don't Know For an Answer?," *Public Opinion Quarterly*, 57, 348-357.
- Keane, M. P. (1992), "A Note on Identification in the Multinomial Probit Model", *Journal of*

- Business & Economic Statistics*, 10, 193–200.
- Lacy, D., and Burden, B. (1999), “The Vote-Stealing and Turnout Effects of Ross Perot in the 1992 U.S. Presidential Election,” *American Journal of Political Science*, 43, 233-255.
- Lynn, P., and Jowell, R. (1996), “How Might Opinion Polls Be Improved?: The Case for Probability Sampling,” *Journal of the Royal Statistical Society*, Ser. A, 159, 21-28.
- McCulloch, R. E., and Rossi, P. (1994), “An Exact Likelihood Analysis of the Multinomial Probit Model,” *Journal of Econometrics*, 64, 207-240.
- McCulloch, R. E., Polson, N. G., and Rossi, P. E. (2000), “A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters”, *Journal of Econometrics*, 99, 173-193
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), “Equations of State Calculations by Fast Computing Machines”, *Journal of Chemical Physics*, 21, 1087-1092.
- Quinn, K. M., Martin, A. D., and Whitford, A. B. (1999), “Voter Choice in a Multi-Party Democracy: A Test of Competing Theories and Models,” *American Journal of Political Science*, 43, 1231-1247.
- Sanders, M. S. (1998), “Unified Models of Turnout and Vote Choice for Two-Candidate and Three-Candidate Elections,” in *Political Analysis*, Volume 7, ed. W. R. Mebane, Ann Arbor, MI: The University of Michigan Press.
- Schofield, N., Martin, A. D., Quinn K. M., and Whitford, A. B. (1998), “Multiparty Electoral Competition in the Netherlands and Germany: A Model Based on Multinomial Probit,” *Public Choice*, 97, 257-293.
- Siddarth, S., Bucklin, R. E., and Morisson, D. G. (1995), “Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data,” *Journal of Marketing Research*, 33, 255-266.
- Swank, O. H., and Eisinga, R. (1999), “Economic Outcomes and Voting Behaviour in a Multi-Party System: An Application to the Netherlands,” *Public Choice*, 101, 195-213.
- Tanner, M. and Wong, W. (1987), “The Calculation of Posterior Distributions by Data Augmentation”, *Journal of the American Statistical Association*, 82, 528–550.
- Yu, P. L. H., and Lam, K. (1997), “How To Predict Election Winners From a Poll,” *Journal of Applied Statistics*, 24, 11-23.
- Van Nierop, E., Paap, R., Bronnenberg, B., Franses, P. H., and Wedel, M. (2000), “Modeling Unobserved Consideration Sets for Household Panel Data”, *ERIM Report Series Research in Management*, ERS-2000-42-MKT, Erasmus University Rotterdam.

Table 1: Descriptive statistics

		CDA	D'66	PvdA	VVD	Other party	No vote	Do not know	Do not want to say	Total
		$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 7$	$j = 6$	$j = 8$	
Actual vote 1989		1093	535	1037	520	311	813	238*	73	4620
Vote intention 1994		662	572	663	566	499	300	1289	69	4620
In consideration set 1994		1651	2123	1702	1458	1652	-	-	-	-
Actual vote 1994		893	710	1113	774	663	467	0	0	4620
Vote intention 1994 crossed with explanatory variables										
Looking for job	Yes	7	23	29	15	29	11	69	1	184
	No	655	549	634	551	470	289	1220	68	4436
Home ownership	Own house	406	331	290	390	244	120	676	38	2495
	Rental house	256	241	373	176	255	180	613	31	2125
Gender	Female	362	280	332	259	245	147	784	38	2447
	Male	300	292	331	307	254	153	505	31	2173
Dwelling	Countryside	105	63	73	63	41	30	144	5	524
	Urb. countryside	310	192	232	225	165	103	497	25	1749
	Small town	154	174	209	148	158	93	369	16	1321
	Large town	93	143	149	130	135	74	279	23	1026
Income	Below modal	152	95	199	65	140	95	326	19	1091
	Modal	272	199	270	177	199	144	589	31	1881
	Above modal	238	278	194	324	160	61	374	19	1648
Age	age $\leq 34$	173	210	166	151	179	118	459	16	1472
	34 < age $\leq 54$	209	284	261	253	198	89	497	24	1815
	age > 55	280	78	236	162	122	93	333	29	1333
Union membership	FNV	59	137	202	46	93	44	222	9	812
	CNV	66	29	16	24	19	7	74	1	236
	Other	56	49	60	53	39	17	100	6	380
	None	481	357	385	443	348	232	893	53	3192

\*238 means "do not recall".

Table 2: Cross-tabulation of intended and actual vote

		Actual vote June 1994						
		CDA	D'66	PvdA	VVD	Other Party	No Vote	Total
Vote intention	CDA	499	45	32	34	28	24	662
April 1994	D'66	31	349	97	41	36	18	572
	PvdA	13	28	538	9	27	48	663
	VVD	42	30	13	443	19	19	566
	Other	23	43	65	18	315	35	499
	No vote	21	19	54	25	36	145	300
	Do not know	253	193	290	192	191	170	1289
	Do not want to say	11	3	24	12	11	8	69
Total		893	710	1113	774	663	467	4620

Table 3: Posterior results MVP-part of the 2-stage model

posterior mean $\alpha$					
	CDA	D'66	PVDA	VVD	other
intercept	-0.8016	-0.4950	-1.2280	-0.4969	-1.0090
job searcher	-0.0675	0.1621	0.1136	-0.0822	0.1279
home owner	0.1904	-0.0256	-0.0039	0.2110	-0.0350
female	0.0566	-0.0803	0.0155	-0.0288	-0.0804
urbanisation	-0.1001	0.0790	0.0391	-0.0244	0.0855
below modal	0.0476	-0.2463	0.2371	-0.4211	0.2724
modal	0.1413	-0.2118	0.0637	-0.1177	0.1283
young	0.1950	0.3810	0.2388	0.1328	0.3710
middle aged	-0.0019	0.3786	0.2978	0.0163	0.2737
FNV member	-0.4073	0.2140	0.2348	-0.3823	0.2255
CNV member	0.1513	-0.0432	-0.0176	-0.2238	0.2458
other union member	0.1025	0.1365	0.0838	-0.1952	0.0891
no vote (1989)	0.3692	-0.1372	0.4480	0.0797	0.0452
do not know (1989)	0.6086	-0.0362	0.4971	0.0818	0.2571
do not want to say (1989)	0.5401	0.1172	0.6780	0.3449	0.1905
posterior standard deviation $\alpha$					
intercept	0.0098	0.0079	0.0087	0.0091	0.0082
job searcher	0.0125	0.0107	0.0114	0.0129	0.0102
home owner	0.0024	0.0021	0.0022	0.0023	0.0020
female	0.0020	0.0016	0.0019	0.0018	0.0017
urbanisation	0.0006	0.0005	0.0005	0.0005	0.0005
below modal	0.0043	0.0036	0.0041	0.0042	0.0038
modal	0.0027	0.0024	0.0026	0.0027	0.0024
young	0.0035	0.0029	0.0034	0.0035	0.0031
middle aged	0.0033	0.0026	0.0031	0.0031	0.0029
FNV member	0.0040	0.0029	0.0033	0.0034	0.0028
CNV member	0.0103	0.0082	0.0102	0.0093	0.0087
other union member	0.0061	0.0059	0.0060	0.0060	0.0054
no vote (1989)	0.0035	0.0033	0.0035	0.0034	0.0031
do not know (1989)	0.0090	0.0086	0.0089	0.0097	0.0086
do not want to say (1989)	0.0492	0.0477	0.0432	0.0512	0.0467
posterior mean $\beta$			1.6980		
posterior standard deviation $\beta$			0.0009		

Table 4: Posterior results MNP-part of the 2-stage model

posterior mean $\gamma$					
	CDA	D'66	PVDA	VVD	other
intercept	1.2230	1.1630	1.2050	1.0170	1.1520
job searcher	-0.4882	-0.2095	0.1897	0.0886	0.1515
home owner	-0.0710	-0.0541	-0.0737	0.1019	0.0201
female	-0.1518	-0.1730	-0.1703	-0.2645	-0.0591
urbanisation	-0.1508	-0.1007	-0.0162	0.0505	0.0219
below modal	-0.1636	-0.2827	-0.3505	-0.3840	-0.1994
modal	-0.2521	-0.2660	-0.3496	-0.2946	-0.2785
young	0.0962	0.4956	-0.1364	0.1854	0.0219
middle aged	0.0191	0.2268	-0.2294	0.3358	-0.0674
FNV member	-0.0372	0.0684	0.0594	-0.1129	-0.0217
CNV member	0.1834	0.3524	0.0961	-0.2046	-0.4478
other union member	0.0572	-0.0185	0.2953	0.2758	0.0506
no vote (1989)	0.7953	0.5509	0.9859	0.8285	0.6974
do not know (1989)	-0.3833	-0.0476	-0.1202	-0.9143	-0.2858
do not want to say (1989)	0.1418	-1.4820	-0.4770	-0.8322	-1.0250
posterior standard deviation $\gamma$					
intercept	0.0573	0.0570	0.0595	0.0656	0.0651
job searcher	0.1158	0.0627	0.0601	0.0897	0.0610
home owner	0.0132	0.0108	0.0122	0.0154	0.0121
female	0.0113	0.0105	0.0116	0.0137	0.0117
urbanisation	0.0034	0.0031	0.0032	0.0037	0.0034
below modal	0.0254	0.0214	0.0240	0.0326	0.0243
modal	0.0149	0.0135	0.0178	0.0165	0.0176
young	0.0193	0.0228	0.0219	0.0220	0.0229
middle aged	0.0180	0.0204	0.0191	0.0223	0.0205
FNV member	0.0272	0.0166	0.0166	0.0289	0.0185
CNV member	0.0453	0.0583	0.0951	0.0762	0.0559
other union member	0.0493	0.0397	0.0455	0.0521	0.0518
no vote (1989)	0.0304	0.0213	0.0253	0.0253	0.0206
do not know (1989)	0.0477	0.0392	0.0472	0.0732	0.0413
do not want to say (1989)	0.3593	0.9197	0.2786	0.4626	0.3143
posterior mean $\delta$		1.4680			
posterior standard deviation $\delta$		0.0020			
posterior mean $\theta$		0.388			
posterior standard deviation $\theta$		0.0001352			

Note: For empirical identification we restrict  $\Omega$  to be a  $J - 1$ -dimensional identity matrix.

Table 5: Posterior results for  $\Sigma$  in MVP-part of the 2-stage model

posterior mean $\Sigma$						
	CDA	D'66	PvdA	VVD	other	
CDA	1.0000	-0.0170	0.0899	0.1985	-0.1726	
D'66	-0.0170	1.0000	0.1182	0.0505	0.0764	
PvdA	0.0899	0.1182	1.0000	-0.1519	0.0237	
VVD	0.1985	0.0505	-0.1519	1.0000	-0.2087	
other	-0.1726	0.0764	0.0237	-0.2087	1.0000	
posterior standard deviation $\Sigma$						
	CDA	D'66	PvdA	VVD	other	
CDA	0.0000	0.0009	0.0009	0.0009	0.0008	
D'66	0.0009	0.0000	0.0007	0.0008	0.0006	
PvdA	0.0009	0.0007	0.0000	0.0008	0.0007	
VVD	0.0009	0.0008	0.0008	0.0000	0.0007	
other	0.0008	0.0006	0.0007	0.0007	0.0000	

Table 6: Prediction results of total election for the four models and fit statistics

Model	CDA	D'66	PvdA	VVD	other	no vote	sq. error
A	22.37	19.21	22.70	18.47	17.24	0	139.44
B	21.41	18.54	21.61	17.98	16.60	3.86	66.21
C	20.99	17.94	21.45	17.16	16.27	6.20	35.52
D1	20.67	17.98	21.21	16.44	15.47	8.24	21.75
D2	20.58	17.89	20.28	16.26	15.10	9.89	23.25
True value	19.33	15.37	24.09	16.75	14.35	10.11	

Table 7: Covariance matrix of the predictions

Model C						
	CDA	D'66	PvdA	VVD	other	no vote
CDA	1.3510	-0.2797	-0.2002	-0.2096	-0.3246	-0.3369
D'66	-0.2797	1.1570	-0.2756	-0.1993	-0.1738	-0.2287
PvdA	-0.2002	-0.2756	1.4390	-0.2414	-0.3118	-0.4101
VVD	-0.2096	-0.1993	-0.2414	1.2560	-0.2817	-0.3234
other	-0.3246	-0.1738	-0.3118	-0.2817	1.4370	-0.3451
no vote	-0.3369	-0.2287	-0.4101	-0.3234	-0.3451	1.6440

  

Model D1						
	CDA	D'66	PvdA	VVD	other	no vote
CDA	1.1070	-0.2180	-0.0933	-0.2252	-0.1529	-0.4178
D'66	-0.2180	1.0870	-0.2405	-0.1319	-0.1901	-0.3070
PvdA	-0.0933	-0.2405	1.2180	-0.2153	-0.2158	-0.4530
VVD	-0.2252	-0.1319	-0.2153	0.9437	-0.1453	-0.2259
other	-0.1529	-0.1901	-0.2158	-0.1453	1.1160	-0.4114
no vote	-0.4178	-0.3070	-0.4530	-0.2259	-0.4114	1.8150

Note: The entries in this table should be multiplied with e-005.

Figure 1: Marginal predictive posterior densities of total election results

