Correcting for Bias in Retrospective Data*

Zvi Eckstein

Tel Aviv University† and University of Minnesota‡

Ron Shachar

Tel-Aviv University§ and Duke University¶

Abstract

When panel data is not available, retrospective data is being used in the estimation of dynamic choice models. However, retrospective data is not reliable. Previous studies of voting choices, for example, have shown that respondents misreport their past choices in order to appear more consistent with their current choice. Such retrospective bias leads to inconsistent estimates, especially when there is state dependence in choices. Specifically, observed persistence in retrospective data may be due to (a) true state dependence, (b) unobserved heterogeneity, and (c) retrospective bias in reporting previous choices. Whereas Heckman (1981) deals with (a) and (b), we introduce a method to estimate true state dependence while accounting for both unobserved heterogeneity and retrospective reporting bias. Our method is based on modeling the reporting behavior, and integrating it into the estimation. The identification strategy is based on the correlation between the reported previous choices and current exogenous variables. Using data on Israeli voters, we find that the probability that a respondent whose vote intention in 1991 differed from her past voting choices would lie about her past choices is 0.23.

*Our thanks go to Dmitri Byzalov who was kind enough to share some of his expertise. Our referees and the editor, Herman K. van Dijk, were particularly helpful with their constructive comments, and we very much appreciate their efforts.

†The Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv, Israel 69978. Phone: +972-3-640-9914. E-mail: eckstein@post.tau.ac.il.

‡Department of Economics, University of Minnesota, 1035 Heller Hall, 271 - 19th Avenue South, Minneapolis, MN 55455, USA. Phone: (612) 625-0145. E-mail: eckstein@econ.umn.edu

§Faculty of Management, Tel Aviv University, Tel Aviv, Israel, 69978. Phone: +972-3-640-6311; Fax: +1-240-414-4844; E-mail: rroonn@post.tau.ac.il.

¶Fuqua School of Business, Duke University, 1 Towerview Drive, Durham, NC 27708, USA. Phone: (919) 660-8046. E-mail: shachar@duke.edu

1 Introduction

Voting has been found to be among the most persistent political attitudes and behavior patterns. In most democracies at least two out of any three individuals vote for the same party in sequential elections.[1] Using panel data on the 1972 and 1976 U.S. presidential elections, Shachar (2003) shows that there is state dependence in voting even when one accounts for both observed and *unobserved* heterogeneity. Therefore, when studying voting theories or any other application in which there is state dependence, it is important to control for individuals' previous choices. Panel data sets are the most appropriate source of information about respondents' decisions over time. However, since they are rare and expensive, scholars often use retrospective data sets instead.

The common view is that retrospective data is not reliable since individuals (intentionally or not) distort facts about their past. Moreover, the distortion process cannot be considered as random. Benewick et al. (1964) showed that respondents who were less accurate about their previous voting decision erred in the direction of making the previous vote consistent with the vote they had just cast. Himmelweit et al. (1978) confirmed these results that when respondents make errors they tended in the direction of consistency with the behavior at the time of recall. Another empirical regularity which is consistent with these results is that the percentage of voters who claimed to vote for the same party in sequential elections is ten percent higher in retrospective data than in panel data (see Zuckerman (1990)).

This behavior may lead to spurious state dependence. Thus, observed persistence in retrospective data maybe due to (a) true state dependence, (b) unobserved heterogeneity and (c) retrospective bias in reporting previous votes. While Heckman (1981a) deals with (a) and (b), we suggest a method to deal with (a) and (c), while controlling for (b). In other words, we introduce a method to estimate true state dependence while controlling for both unobserved heterogeneity and retrospective reporting bias. Our method, which is presented in section 3, is based on modelling the reporting behavior. Our assumptions on the reporting behavior are consistent with previous evidence. In particular we assume that individuals' reports about their previous decision may be true or false. When the report is false, individuals are misreporting their past votes to make them appear more consistent with their current vote.

Using data on Israeli voters in 1991, we find that the probability that a respondent whose vote intention in 1991 differed from her past voting choices would lie about her past choices is 0.23. The voting model is also estimated using alternative methods to deal with retrospective bias

[1](a) In Great Britain between 1959 and 1983 two out of three people voted for the same party in sequential elections. In West Germany between 1949 and 1980, three out of four. In Austria between 1962 and 1982, nine out of ten and in the Netherlands between 1946 and 1983, three out of four (Zuckerman (1990)).

(b) Related studies on sources of persistence are Converse (1976), Franklin and Jackson (1983), Franklin (1984), Markus and Converse (1979), various contributions in Niemi and Weisberg (1976, 1984 and 1993) and Page and Jones (1979).

in reporting past choices (which are presented in section 2). A comparison of the results across the suggested method and its alternatives demonstrates that the suggested approach is superior in terms of its fit with the data and the robustness of the estimates.

The problem of retrospective bias is not unique to political data. Many economic studies are based on retrospective data and the issue of reliability has been raised for these applications as well. For example, Morgenstern and Barrett (1974) and Horvath (1982) present evidence on retrospective bias in unemployment reporting. Torelli and Trivellato (1993) focus on a single recall bias phenomenon: the 'heaping effect,' i.e., abnormal concentration of response at certain durations.[2]

2 Model and estimation issues

This section presents a basic dynamic discrete choice model that will serve us in demonstrating the suggested method (in the following section) and its alternatives (in this section). Since the problem of retrospective bias is not unique to voting, the model is general and can describe various types of dynamic decision making.

In each period t, the individual has two choice alternatives – A and B (e.g., political parties). The assumption that the choice set includes only two alternatives is made for two reasons: (1) simplicity, and (2) consistency with the empirical application. However, many choice situations concern a choice between multiple options. As illustrated in subsection 3.1, the suggested approach is not restricted to two alternatives applications.

The decision of individual i at period t is presented by the binary variable $d_{i,t}$, which is equal to 1 if she chooses A and to zero otherwise. Let $x_{i,t}$ be a vector of observable exogenous variables that affect her decision at time t (for example, the difference in her assessment of candidates' competency), and $\varepsilon_{i,t}$ be an unobserved random shock to preferences (for example, personal experiences that affect party preferences). Finally, we assume that previous choices have a genuine behavioral effect on current decisions (i.e., state dependence in choices).

The time interval for the dynamic process is 1 to T_i, where $t = 1$ corresponds to the first period in which a choice is ever made. For now, we assume that the individual is myopic and ignores the future consequences of her current decision. Thus, she chooses the alternative that yields a higher current utility. Under standard assumptions on the utility function and $\varepsilon_{i,t}$, the decision

[2] They present a measurement error model for the heaping process and combine it with a duration model. Here, we consider a problem in which the retrospective bias is a result of individual behavior, not a statistical error. Furthermore, what makes our problem even more interesting is that the retrospective bias of previous choices depends on ex-post assessment of these choices.

rule at t can be written as (See, for example, Heckman [1981a] and Eckstein and Wolpin [1989]):

$$\begin{aligned} d_{i,t} &= 1 \Leftrightarrow \varepsilon_{i,t} \geq \varepsilon_t^*(d_{i,t-1}, ..., d_{i,1}, x_{i,t}; \theta) \text{ for } t > 1 \\ \text{and } d_{i,1} &= 1 \Leftrightarrow \varepsilon_{i,1} \geq \varepsilon_1^*(x_{i,1}; \theta) \end{aligned} \tag{1}$$

where θ is a vector of parameters that fully describe the utility function.

The θ parameter vector can be estimated using a panel data set of the ds and xs of I individuals, each of which makes T_i choices. With i.i.d. $\varepsilon_{i,t}$ the joint likelihood function is given by:

$$L(\theta) = \prod_{i=1}^{I} \prod_{t=1}^{T_i} \left[1 - F_t\left(\varepsilon_t^*(d_{i,t-1}, ..., d_{i,1,}x_{i,t}; \theta)\right)\right]^{d_{i,t}} \cdot \left[F_t\left(\varepsilon_t^*(d_{i,t-1}, ..., d_{i,1,}x_{i,t}; \theta)\right)\right]^{(1-d_{i,t})} \tag{2}$$

where the c.d.f. of $\varepsilon_{i,t}$ is $F_t(\cdot)$. Notice that if $\varepsilon_{i,t}$ has a normal (extreme value) distribution, the likelihood function for each period is the standard probit (logit). Given explicit functional forms for $\varepsilon_t^*(\cdot)$ and $F_t(\cdot)$ the vector of parameters θ can be consistently estimated (assuming that the model is identified).[3]

2.0.1 Unobserved heterogeneity

The likelihood in eq. (2) ignores persistence in the unobserved variables (i.e., unobserved heterogeneity). Such heterogeneity may lead to spurious state dependence, even if we observe true current and past voting. As a solution to this problem, Heckman (1981a) suggested including the unobservable variables in the model and integrating them out from the likelihood function. It is easy to adopt this approach, and we do so in the empirical example. However, in order to focus on the key issue of this study (retrospective bias) and simplify the presentation, we ignore the unobserved heterogeneity in eq. (2).

2.0.2 Initial observations

The likelihood in eq. (2) assumes that the data include the entire choice history for all the respondents. However, frequently, the initial observations of the dynamic process are missing. In other words, the first choice *observed* in the data is not the first choice *made* by the individual. Thus, the initial observations are unobserved. Heckman (1981b) suggested to resolve this issue by integrating out all possible unobserved choice sequences prior to the first observed choice. In the empirical application, we follow his approach and incorporate the distribution of the initial choices

[3] The likelihood in eq. (2) is standard for the limited dependent variable model. Standard regularity conditions on the functions $\varepsilon_t^*(\cdot)$ and $F_t(\cdot)$ should be satisfied.

in the structural estimation. The specifics of our solution are described in subsection (4.3). Once again, to simplify the presentation, we ignore the unobserved choices issue until subsection (4.3).

2.1 Retrospective bias and solutions

So far we have assumed that a panel data exists. However, frequently such data does not exist, and respondents are asked to report their past choices (and exogenous variables) via a survey. Let $d^r_{i,t}$ be the individual's reported choices. If in some cases $d^r_{i,t} \neq d_{i,t}$, there is a retrospective bias. As discussed above, such a bias may result either from the individual's intentional decision and/or from a rationalization process. Using the reported choices in eq. (2) instead of the *unobserved* true choices may generate bias in the estimation of the state dependence. Specifically, if respondents are misreporting their past choices to make them appear more consistent with their current choice, the estimate of the state dependence is inflated.[4]

While others have made progress on a similar problem of vote overreporting – Abramson and Claggett (1992), Abramson and Ostrom (1994), Presser (1990), Presser and Traugott (1992) – we claim that current methods to deal with choice retrospective bias are inaccurate and inefficient. We discuss three practical ways and their disadvantages.

Ignore data on past choices (Model I): The first way is to ignore the data on past ds. This can be done in two ways. First, the reported past choices are not used as explanatory variables, but they are included as endogenous variables. In other words, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{I} \prod_{t=1}^{T_i} \left[1 - F_t\left(\varepsilon^*(x_{i,t};\theta)\right)\right]^{d^r_{i,t}} \cdot \left[F_t\left(\varepsilon^*(x_{i,t};\theta)\right)\right]^{(1-d^r_{i,t})} \tag{3}$$

Second, the reported past choices are completely ignored and the model is estimated only for the current decision (unconditional on past choices). Specifically, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{I} \left[1 - F\left(\varepsilon^*(x_{i,T_i};\theta)\right)\right]^{d_{i,T_i}} \cdot \left[F\left(\varepsilon^*(x_{i,T_i};\theta)\right)\right]^{(1-d_{i,T_i})} \tag{4}$$

In both cases, the likelihood is mis-specified since it ignores the state dependence element.[5] Furthermore, in the first case (eq. 3) the dependent variable is biased. Thus, the estimators of the parameters that associate x with d are inconsistent. Furthermore, if there is state dependence, ignoring past ds leads to inefficient predictions.

Ignore the retrospective bias (Model II): The second way is to ignore the retrospective bias and use the reported values on past decisions as if they are true. Specifically, the likelihood

[4] Note that we assume that there is no retrospective bias in the xs for reasons discussed later.

[5] The likelihood in eq. (4) is consistent with the model only in the case that $T = 1$ for all the respondents (i.e., the actual history of all the respondents include only one choice period).

function is:

$$L(\theta)=\prod_{i=1}^{I}\prod_{t=1}^{T_i}\left[1-F_t\left(\varepsilon_t^*(d_{i,t-1}^r,...,d_{i,1}^r,x_{i,t};\theta)\right)\right]^{d_{i,t}^r}\cdot\left[F_t\left(\varepsilon_t^*(d_{i,t-1}^r,...,d_{i,1}^r,x_{i,t};\theta)\right)\right]^{(1-d_{i,t}^r)} \quad (5)$$

This, obviously, leads to inconsistent estimates. As discussed above (and demonstrated below), respondents tend to misreport their past choices in order to make them appear more consistent with their current choice. In other words, $d_{i,T-1}^r,...,d_{i,1}^r$ depend on $d_{i,T}$. Since $d_{i,T}$ is a function of $\varepsilon_{i,T}$, we get that $E(\varepsilon_{i,T}|d_{i,T-1}^r,...,d_{i,1}^r)\neq E(\varepsilon_{i,T})$. This means that the exogeneity assumption does not hold (even if there is no unobserved heterogeneity), and the estimates are inconsistent.

Integrate the true past choices out (Model III): A third way to deal with the problem of retrospective bias is to assume that past choices are unobserved and to integrate them out of the likelihood function. In this case, the likelihood is:

$$L(\theta)=\prod_{i=1}^{I}\sum_{d_{i,T_i-1}\in\{0,1\}}...\sum_{d_{i,1}\in\{0,1\}}\prod_{t=1}^{T_i}\left[1-F_t\left(\varepsilon_t^*(d_{i,t-1},...,d_{i,1},x_{i,t};\theta)\right)\right]^{d_{i,t}}\cdot\left[F_t\left(\varepsilon_t^*(d_{i,t-1},...,d_{i,1},x_{i,t};\theta)\right)\right]^{(1-d_{i,t})} \quad (6)$$

Unlike the previous two approaches, this solution leads to consistent estimators of the model's parameters. The identification of the state dependence parameters is based on the effect of $x_{i,l}$ on $d_{i,t}$, where $l<t$ (see Chamberlain 1993, and Shachar 1994). Notice that if there is state dependence (i.e., $d_{i,t}$ structurally depends on $d_{i,l}$) then the choice in period t depends indirectly on the variables that determine the decision in period l (i.e., $x_{i,l}$). Therefore, any evidence that the choice probability in the period t (conditional on $x_{i,t}$) is still a function of $x_{i,l}$ indicates state dependence in choice. Thus, this approach leads to consistent estimates.

However, by removing the reported past choices from the data, the researcher ignores information that can assist in improving the efficiency of the estimates. In other words, the estimates are consistent, but not efficient.[6]

3 The suggested method (Model IV)

Since the source of the bias is behavioral, the most appropriate way to treat it is to model it (as we do with other behavioral processes). Thus, we suggest to model the reporting process and integrating it into the likelihood function for estimation. Specifically, the likelihood includes both the reporting behavior and the dynamic choice model. Thus, the reported persistence can be

[6] Furthermore (a) since we focus on state dependence, ignoring this information implies that we have no direct evidence (data) on the key aspect of the study, and (b) the identification of the unobserved heterogeneity in this case is based only on the structure of the model and thus sensitive to functional form decisions.

broken-down into true state dependence and retrospective bias. Furthermore, by integrating out also the unobserved heterogeneity, we can separate its effect as well.

The reporting behavior at time T is modeled as:[7]

$$\left(d^r_{i,T-1}, d^r_{i,T-2}, ..., d^r_{i,1}\right) = I\left(d_{i,T-1}, d_{i,T-2}, ..., d_{i,1}, z_{i,T}, \omega_{i,T}, d_{i,T}; \gamma\right) \tag{7}$$

where $I(\cdot)$ is a vector indicator function, $\omega_{i,T}$ is a vector of unobservable variables at period T, $z_{i,T}$ is a vector of observed variables at period T (excluding $d_{i,T}$), and γ is a vector of parameters that fully describes this process. The vector on the left-hand side of eq. (7) includes the reported past choices, and the elements of the function on the right-hand side are the true choices and additional variables that are involved in the reporting process (such as the current choice).

Since the ω_T are unobserved, $I(\cdot)$ is not a deterministic function for the econometrician. We assume that the decision at the time of the interview is reported without an error, that is $d^r_{i,T}$ $= d_{i,T}$.

Therefore, the likelihood is:

$$\begin{aligned} L(\theta,\gamma) = \prod_{i=1}^{I} \sum_{d_{i,T_i-1}\in\{0,1\}} \cdots \sum_{d_{i,1}\in\{0,1\}} \Pr\left(d^r_{i,T_i-1}, ..., d^r_{i,1} | d_{i,T_i-1}, ..., d_{i,1}, z_{i,T}, d_{i,T_i}; \gamma\right) \\ \cdot \prod_{t=1}^{T_i} \left[1 - F_t\left(\varepsilon^*_t(d_{i,t-1}, ..., d_{i,1}, x_{i,t}; \theta)\right)\right]^{d_{i,t}} \cdot \left[F_t\left(\varepsilon^*_t(d_{i,t-1}, ..., d_{i,1}, x_{i,t}; \theta)\right)\right]^{(1-d_{i,t})} \end{aligned} \tag{8}$$

The first line of (8) is based on the reporting process presented in (7), and the second is based on the choice model (as in eq. 2). If we wish to control for unobserved heterogeneity, we need to include a permanent personal unobserved variable in the decision rule and integrate it out of the joint probability. To close the model for estimation, one needs to parameterize the reporting behavior equation (eq. 7).

3.1 An example of a reporting process

As discussed above, previous studies have suggested that respondents are misreporting their past choices to make them appear more consistent with their current choice. Thus, we consider a simple case of equation (7) where:

$$d^r_{i,T-l} = \left\{ \begin{array}{ccc} d_{i,T-l} & \text{w.p.} & \gamma_l \\ d_{i,T} & \text{w.p.} & 1-\gamma_l \end{array} \right\}, \text{ for } l = 1, ..., T-1. \tag{9}$$

[7] This equation may be given a structural specification, rather than the reduced form specification presented here. As discussed in the conclusions, this is a suggested extension of our approach.

Here γ_l is the probability of telling the truth and $1-\gamma_l$ is the probability of reporting current decision as past behavior, where l is the number of periods before the interview. In other words, a person who switched parties will report her past choice truthfully with probability γ_l.[8]

There are potentially other behavioral models of misreporting that are consistent with eq. (7). Here, we focus on the process in (9), because it clearly exemplifies the most critical threat in the identification of a voting model with state dependence. This reporting process intervenes with the observed dynamics of voting decisions and makes it hard to estimate state dependence consistently. Specifically, we need to determine which part of the correlation between (let's say) d^r_{T-l} and d_T is due to state dependence and which part is due to the reporting process.

Identification: The identification is based on the correlation between d^r_{T-l} and current exogenous variables x_T, conditional on x_{T-l}. Retrospective reporting is the only reason that the reported *previous choice* might be correlated with *current exogenous* variables. In other words, if decisions made in the past (as reported today) depend on the current realizations of the exogenous variables, it implies that there is a retrospective bias in the reports.

The following formal discussion might shed some light on the above intuitive argument. Consider, for simplicity a two periods model. It is easy to show (see Appendix) that

$$\Pr(d^r_1|x_1,x_2)=\gamma\Pr(d_1=d^r_1|x_1)+(1-\gamma)\Pr(d_2=d^r_1|x_1,x_2) \tag{10}$$

The rationale behind this equation is the following. The first element corresponds to the case that the individual report her previous choice without bias. In such a case, her choice (and thus x) in the second period is not relevant. The second element corresponds to the case that the individual report her second choice as if it was her first choice.[9] In such a case, the exogenous variable of the second period, x_2, becomes relevant.

Equation (10) clarifies that the only reason that d^r_1 might depend on x_2 is due to the reporting process. Specifically, when $\gamma=1$ (i.e., there is no retrospective bias) d^r_1 is not a function of x_2. Thus, the correlation between d^r_{T-l} and x_T (conditional on x_{T-l}) identifies γ. Specifically, if this correlation is zero, the estimate of γ is 1, and if the correlation is larger than zero, the estimate of γ is smaller 1.

Furthermore, notice that even if γ is a function of x_T (i.e., x_T is in z_T), the argument above

[8] The process in eq. 9 assumes that reported past choices $d^r_{T-1},d^r_{T-2},...,d^r_1$ are mutually independent (conditional on actual current and past choices). Moreover, given current choice d_T, the reported choice for date $T-l$ is independent of all actual past choices with the exception of actual past choice d_{T-l}. Thus, $\Pr\left(d^r_{T-1},d^r_{T-2},...,d^r_1|d_{T-1},d_{T-2},...,d_1,z_T,d_T\right)=\prod_{l=1}^{T-1}\Pr(d^r_{T-l}|d_T,d_{T-l})$. This assumption is not necessary for identification. It was chosen because (a) it allows us to focus on the most critical threat that retrospective bias introduces and (b) it is consistent with previous findings about the reporting process.

[9] Notice, that this does not necessarily imply that the individual reports her choice incorrectly, since d_2 might be the same as d_1.

still holds. Once again, any evidence that d^r_{T-l} depends on x_T would reflect a retrospective bias.

Monte Carlo experiments: The results of the Monte-Carlo experiments, reported in Table 1, demonstrate that our identification strategy is effective in estimating the model's parameters. Furthermore, while in the model presented above the individual faces two alternatives, in the Monte-Carlo experiments the number of alternatives is three. Thus, the results in Table 1 also show that the suggested approach is not restricted to cases with only two options.

The model used in the experiments has the following properties. There are 500 individuals who make a choice among three alternatives in two periods. The expression $c_{i,t} = j$ is the notation for the event that the choice of individual i at time t is alternative j. The utility is:

$$u_{i,j,t} = \left\{ \begin{array}{cc} \beta_j + \beta x_{i,j,t} + \varepsilon_{i,j,t} & \text{for} \quad t=1 \\ \beta_j + \beta x_{i,j,t} + \varepsilon_{i,j,t} + \delta I\{c_{i,t-1} = j\} & \text{for} \quad t=2 \end{array} \right\} \tag{11}$$

where x is an observed variable, ε is an unobserved variable with a type I extreme value distribution, the βs and δ are parameters, and $I\{\}$ is the indicator function. The individual is maximizing her per-period utility. Thus:

$$c_{i,t} = j \Leftrightarrow u_{i,j,t} = \max(u_{i,1,t}, u_{i,2,t}, u_{i,3,t})$$

The retrospective reporting process is as in eq. (9). Specifically,

$$c^r_{i,1} = \left\{ \begin{array}{ccc} c_{i,1} & \text{with probability} & \gamma \\ c_{i,2} & \text{with probability} & 1-\gamma \end{array} \right\} \tag{12}$$

The model was simulated and estimated 1000 times.

The means of the estimated parameters are very close to their true values. For example, the average of $\widehat{\delta}$ is 0.9943 where the true value is 1, and the average of $\widehat{\gamma}$ is 0.7518 where the true value is 0.75.

Table 1 also reports the results of a model in which respondents misreport their final choice. This could be not uncommon. For example, suppose that a respondent in the past voted for a large mainstream party. At the final election, however, the respondent decided to vote for a more extreme party. The respondent may not want to reveal this to the interviewer and reports votes for the mainstream party at all elections. In this case, the reporting process is:

$$c^r_{i,2} = \left\{ \begin{array}{ccc} c_{i,2} & \text{with probability} & \gamma \\ c_{i,1} & \text{with probability} & 1-\gamma \end{array} \right\} \text{ and } c^r_{i,1} = c_{i,1}$$

The means of the estimated parameters in this case are also very close to their true values. However, the average of $\widehat{\gamma}$ is a bit less precise (0.7593 and standard deviation of 0.083 versus 0.7518 and standard deviation of 0.048 in the first case). The reason for this is probably that in this case the identification is based on the structure of the model (unlike the previous case, in which the identification is based on the effect of x_2 on d_1^r).

It is worth noting that when respondents misreport their past choices, it is necessary to assume that they report their last choice correctly, and, when they misreport their current choice, we need to assume that they report their past choice correctly. In other words, one needs to assume that (at least) one choice is observed without error.

While the approach presented here can handle misreporting of the final choice, and such behavior might not be uncommon, recall that the focus of this study is on misreporting past choices. The empirical example presented in the following section illustrates such a case.

4 An empirical example

This section presents the results from estimating a voting model under the different ways suggested above to handle bias in retrospective data (i.e., Models I-IV). The advantage of using the suggested method is illustrated by a comparison of the empirical results of the four models.

Following Shachar (2003), the exogenous variables, xs, are (a) candidates' competency, (b) policy issues, and (c) voters' demographic characteristics.

4.1 Data

The empirical study is based on a survey which was conducted among 413 respondents in Israel in 1991, a year before the 1992 general elections. The survey was directed by us, and the data was collected by the Dahaf Research Institute. In Israel, like in almost all other countries, there is no panel data on voters. Thus, data available for estimation of a voting model is based on retrospective reports of respondents. The data consist of the *reported* voting behavior in the 1984, and 1988 elections, and respondents' voting intensions in 1991.[10]

There are three different types of respondents: (a) those who were not eligible to vote before 1991 (65 people), (b) those who were eligible to vote in the 1988 elections but not before (55 people), and those who had the right to vote in each election that was covered by the survey (293 respondents). Table 2 presents the voting pattern of these respondents. The decision variable $d_{i,t}$ is equal to 1, if respondent i voted for the Labor party, and to zero, if she voted for the Likud Party.[11] Among the 293 respondents who had the right to vote prior to 1988, 85 percent (248)

[10] Specifically, there were no general elections in 1991. Thus, the respondents were asked: "If the elections had been held today, which party would you vote for?"

[11] The vote variable in a multiparty system such as Israel can be (and has been) defined in different ways, such as (a) all competing parties, (b) only the (two) large parties, (c) the two major political blocs, or (d) several political blocs along a left-right continuum. The definition that we use–only the two large parties–is consistent with the model of section 2.

respondents reported that they voted for the same party in all three periods.

Table 3 reports the sample means and standard deviations of the exogenous variables. We denote the perceived relative competency of the candidates by $w_{i,t}$, and the relative distance between the candidates and the respondent on policy issues by $y_{i,t}$.

The policy distance variable, $y_{i,t}$, measures the policy closeness to the Labor party, where zero is the neutral point. For example, while in the 1984 elections voters were closer (on average) to the Labor ($\frac{1}{I}\sum_{i=1}^{I} y_{i,84} > 0$), in 1991 they were closer (on average) to the Likud ($\frac{1}{I}\sum_{i=1}^{I} y_{i,91} < 0$). On the other hand, there was no significant change in the second exogenous variable, $w_{i,t}$. This variable measures the perceived relative competency of the Labor's candidate on a five-points scale. The insignificant changes in this variable during the studied period might be due to the fact that the candidates of the two parties were the same people on all the three dates.

We, obviously, do not observe $x_{i,t}$ for $t < T$, and thus use the reported values of $x_{i,t}$ in the survey. It is possible that there is a retrospective bias with respect to these variables as well. However, we assume that the reported values are correct for several reasons. First, while there is evidence of a retrospective bias in reported choices, there is no such evidence on bias in the report of past xs. Second, the behavioral explanation of the retrospective bias is more relevant for past choices than for past xs. Specifically, an individual who switched parties in sequential elections may consider her previous decision as a mistake and as a result is more likely to make a false report. The incentive to distort the report on previous xs is significantly less dramatic for various reasons. For example, the policy issues that determined the voting decision in period $T-l$ might not be relevant anymore at period T. Thus, the respondent can comfortably report her views on the policy issues of period $T-l$ without contradicting her current views. Indeed, when constructing the survey we have included, for each election year, questions about specific and unique policy issues that were relevant for that election year, but are not relevant any more. As a result, the respondent can answer these questions freely without being concerned that her current views are not consistent with her previous ones. The data indicate that our effort was successful: as discussed above, the policy distance variable has moved significantly (on average) over time, in our sample.

The four demographic characteristics are: respondent's age–$Age_{i,t}$;[12] a binary variable that is equal to 1 if the individual is from an African or Asian origin, and zero otherwise ($Spharadi_i$); years of schooling ($Education_{i,t}$); and a four-points measure of religious orthodoxy, which is equal to 1 for an orthodox Jew and to 4 for a secular person ($Religious_i$).

[12] The *Age* variable is equal to one for respondents younger than 24 (and older than 18). It is equal to two for those between 25 and 34; three for those between 35 and 44; four for those between 45 and 54; five for those between 55 and 64 and six for those older than 65.

4.2 Functional forms

The model is specified so that the utility function is linear in x and in the "stock" of past voting decisions, and $\varepsilon_{i,t}$ has a standard normal distribution. Since our data includes only three election periods, we assume that the stock variable depends only on the last two lags. Following Shachar (2003), the stock variable for alternative j is $\sum_{l=1}^{2} \delta^l I\{c_{i,t-l} = j\}$, where δ^l is the "carryover" parameter of choices made l periods prior to t. (We discuss the issue initial observations immediately.) Thus, the threshold, $\varepsilon_{i,t}^*$, in the decision rule (eq. 1) is:

$$\varepsilon_{i,t}^* = -\left(\mu_0 + s_{i,t}\mu + \alpha y_{i,t} + \beta w_{i,t} + \sum_{l=1}^{2} \delta^l (2d_{i,t-l} - 1)\right) \tag{13}$$

where $s_{i,t}$ is a 4-dimensional row-vector of demographic variables and μ is a 4-dimensional vector of parameters. Notice that since the threshold, $\varepsilon_{i,t}^*$, is the difference between the deterministic part of the utilities of the two options, the difference in the stock variable is $\sum_{l=1}^{2} \delta^l \left[I\{c_{i,t-l} = j\} - I\{c_{i,t-l} \neq j\}\right]$ which is equal, in our case, to $\sum_{l=1}^{2} \delta^l (2d_{i,t-l} - 1)$.

4.3 Initial observations

Our data includes the entire vote history for all of the respondents who were not entitled to vote prior to 1984. However, the initial choices of the rest of the sample are missing. As discussed earlier, following Heckman (1981b), we incorporate the distribution of the initial choices in the structural estimation (i.e., integrate out all possible unobserved choice sequences prior to 1984.)

We use the age variable to determine the first election of each individual in our data. Consider, for example, an individual whose 1984 election was the second choice period. Her history probability (from the research point of view) is:

$$\sum_{\widetilde{d}_{i,1} \in \{0,1\}} \Pr(d_{i,4}|d_{i,3}, d_{i,2}, x_{i,4}; \theta) \Pr(d_{i,3}|d_{i,2}, \widetilde{d}_{i,1}, x_{i,3}; \theta) \Pr(d_{i,2}|\widetilde{d}_{i,1}, x_{i,2}; \theta) \Pr(\widetilde{d}_{i,1}|\theta)$$

$$\text{where} \quad \Pr(d_{i,t}|\bullet; \theta) = \left[1 - F_t\left(\varepsilon_t^*(\bullet; \theta)\right)\right]^{d_{i,t}} \cdot \left[F_t\left(\varepsilon_t^*(\bullet; \theta)\right)\right]^{(1-d_{i,t})} \tag{14}$$

More generally, for a respondent who voted l times prior to 1984, the number of sequences is 2^l. Notice that although the current choice depends only on the last two lags, we need to account for the entire choice history. For choices made prior to 1984, we do not observe the exogenous variables, such as perceived candidates competency. To capture the variation in these variables, we allow the standard deviation of ε to differ from 1 for these periods.

We revise the likelihood in eq. (8) accordingly.

4.4 Results

Table 4 presents the maximum likelihood estimates of Models I-IV.

Before describing the estimates, we discuss two issues which are general to all dynamic choice models (and not specific for the case of misreporting past choices): unobserved heterogeneity and initial observation.

4.4.1 Unobserved heterogeneity

All models were estimated initially with an unobserved heterogeneity component. The standard deviation of this heterogeneity was minuscule, both behaviorally and statistically (in all cases the t-statistic was smaller than 0.01).

To illustrate this point further, Figure 1 presents the likelihood of Model IV as a function of the standard deviation of the unobserved heterogeneity parameter. The figure is a result of the following exercise: set the standard deviation at a specific value and estimate all the other parameters of the model, then set it at a different value and reestimate, etc. It is evident that the value that leads to the highest likelihood is zero. (It is worth noting that when we set δ at zero, the estimate of the standard deviation is different from zero even at the one percent significance level.) Thus, the estimates reported in Table 4 are based on the assumption that there is no unobserved heterogeneity.

4.4.2 Initial observations

The issue of unobserved initial observations is relevant for all models but the first one. Since in Model I all past choices are ignored, the lack of the initial choices does not pose a problem.

Table 4 presents the estimation results of Models II-IV with and without our solution to the initial observations issue. When the model is estimated without our solution, the stock variable is based only on the observed choices.

Interestingly, accounting for the unobserved choices significantly improves the likelihood of Model IV (which is based on our suggested approach) and hardly changes the likelihood of the alternative approaches. Furthermore, while the state dependence parameter is about the same in both estimations of Model IV (0.622 and 0.641), it is quite unstable for the alternative approaches (0.663 and 0.613 for Model II and 0.642 and 0.747 for Model III).

These results demonstrate the robustness of our model and its estimates, on the one hand, and the instability of the alternative models, on the other. It is encouraging to find that adding just one parameter to the model (via the formulation of the reporting process) improves its stability significantly. Put differently, the comparison suggests that the alternative approaches are either mis-specified or inefficient.

We are now ready to discuss the estimates.

4.4.3 The estimates and their implications

In all cases, the estimates of the parameters of the policy variable, $\widehat{\alpha}$, and the competency variable, $\widehat{\beta}$, are different from zero, even at the one percent significance level.

Model IV: The point estimate of γ is 0.766 and a likelihood ratio test (χ^2 statistic of 11.1681) demonstrates that we can reject the hypothesis that $\gamma = 1$, even at the 1 percent significance level.[13] In other words, the data support the retrospective bias hypothesis. This finding indicates that respondents do not always report their true past voting decisions. The estimate suggests that about 23 percent of the time, voters report their current voting decision as if it was their previous decision. Notice, though, that this does not mean that 23 percent of the reported past choices are incorrect. Since voters tend to choose the same party in sequential elections, the proportion of *incorrect* reports is much smaller. This proportion is equal to $\Pr(d_{i,T} \neq d_{i,T-l})(1 - \gamma_l)$ and $\Pr(d_{i,T} \neq d_{i,T-l})$ is usually quite small. For example, for respondents who were eligible to vote in the 1984 elections, the model predicts that only 3.7 percent of the reported 1984 choices are incorrect.

While the percent of incorrect reported choices might look small, its effect on the model predictions is far from negligible. For example, while only 12.29 percent of the respondents above reported that they have switched parties between 1984 and 1991, the model predicts that the actual percentage is actually 15.74. The difference between the reported and actual proportions of switchers is significant not only because of the political role of these people in determining the elections, but also because the precision of the estimates in dynamic models is relatively high when the proportion of switchers in not negligible.

Therefore, our estimates demonstrate that *when a respondent is faced with the alternative to lie or to tell the truth (i.e., when she switches), she lies with a probability of 23 percent.* Given the relatively small proportion of switchers in our data, this tendency to lie translates to about 3.7 percent of *incorrect* reports.

A comparison of the results of all the estimated models illustrates the disadvantage of Models I-III.

Model I: Two versions of Model I were estimated. In the first, the reported past choices, d^r_{88} and d^r_{84}, are not used as explanatory variables, but they are included as endogenous variables.

[13] The point $\gamma = 1$ is on the boundary of the parameter space. Thus the test statistic for $H_0 : \gamma = 1$ does not have the standard distribution. Instead, it is distributed as a weighted average of χ^2s with zero and one degrees of freedom. Since the weights are based on the null hypothesis, they are 0.5 and 0.5. Thus, the critical value of the χ^2 for the likelihood ratio test at the 1 percent significance level is 5.411894. (see Wolak 1989 for a discussion of testing inequality constraints). The correct distribution for this test was pointed out to us by one of the referees. We appreciate the referee's assistance on this matter.

In other words, the likelihood function is the joint probability of the voting decisions in 1984, 1988 and 1991, unconditional on past voting decisions (eq. 3). In the second, d^r_{88} and d^r_{84} are completely ignored and the model is estimated only for the 1991 decision (unconditional on past voting decisions. i.e., eq. 4).[14]

Model I has several disadvantages. First, it does not account for an important factor in voting decisions – state dependence. Second (and related to the first), it fits the data significantly worse than Model IV, as demonstrated by the likelihood values. Third, its estimates are inconsistent because the omitted explanatory variable (state dependence) is correlated with the included regressors. Specifically, the xs are correlated over time, and thus, for example, x_{91} is correlated with x_{88}. Since d_{88} depends on x_{88}, there is a correlation between the omitted explanatory variable, d_{88}, and the included regressors, x_{91}. Furthermore, since we expect d_{88} and x_{91} to be positively correlated and d_{88} to have a positive effect on d_{91}, we expect that the effect of the xs would be upward biased. Indeed, a comparison between the estimates of models I and IV support this view.

Model II: This approach uses the reported past choices as if there is no retrospective bias and $d^r_{i,t} = d_{i,t}$. In other words, it assumes that $\gamma = 1$.

A comparison between the likelihood values of Models II and IV reveals that the data is supporting the suggested method. The log-likelihood of Model II is -221.91 compared with -216.33 for Model IV. This finding implies that the separation between the state dependence and the retrospective bias ($\widehat{\delta}$ and $\widehat{\gamma}$ in Model IV) improves the fit of the model. Recall that the separation is enabled by the correlation between x_{91} and the reported past choices (d^r_{88} and d^r_{84}). The significant change in the likelihood indicates that the suggested approach is effective in isolating the state dependence effect and, at the same time, identifying the retrospective bias.

Finally, since $\widehat{\gamma}$ is significantly lower than 1 but this model restricts γ to be equal to one, it is mis-specified. This mis-specification is expressed when our solution to the initial observation issue is applied to this model. As discussed above, while Model IV performs well when the solution is applied (i.e., the likelihood improves and the estimates are stable), Model II does not.

Model III: This approach solves the retrospective bias by ignoring the reported past choices and integrating the unobserved true past decisions from the likelihood function. Thus (a) the only observed endogenous variable is the 1991 choice, and (b) the likelihood value should be compared to the second version of Model I.

This approach has two disadvantages in comparison to the suggested method. First, the estimates are not efficient, since this approach does not use all of the available information (i.e., d^r_{88}

[14] The main difference between the estimates of the two versions is that the absolute values of all the significant parameters are higher in the second version. This result is consistent with the assertion that we do not actually observe past decisions. Instead we observe the reported value of past decision which depends indirectly on past exogenous variables. Therefore, the correlation between the reported choices and the xs is expected to be smaller.

and d^r_{84}). The large standard error of $\widehat{\delta}$ (compared to Model IV) and the significant change in this estimate when the solution for the initial observation is applied, may reflect this aspect. Second, it does not provide us with an estimate of the retrospective bias. We might be interested in such an estimate because it represents a behavioral aspect of the data.

5 Forward looking behavior

So far we have assumed, for simplicity, that the individual is myopic.[15] It seems that this assumption is a critical aspect of our identification strategy for the following reason. The identification is based on the correlation between x_T and d^r_{T-l}. However, a forward-looking individual should base her decision on the expected value of future xs. Thus, one might think that a forward-looking behavior leads to a correlation between x_T and d^r_{T-l}, even if there is no retrospective bias, and thus the identification strategy fails. But such a view is not correct, since at time $T-l$ the individual does not know the realization of x_T. Instead, she decides on d_{T-l} based on $E(x_T|I_{T-l})$ where I_{T-l} is her information set at time $T-l$. Thus, one can still identify the model parameters, even when the individual is forward-looking. In such a case, the identification is based on the correlation between $[x_T - E(x_T|I_{T-l})]$ and d^r_{T-l}.

6 Conclusions

In order to estimate a dynamic choice model the researcher needs information on past choices. When panel data are not available, like in the case of voting, scholars use retrospective data. However, previous studies have suggested that respondents mis-report their past choices in order to appear more consistent with their current choice. Such retrospective bias leads to inconsistent estimates, especially when there is state dependence in choices. Here we suggest a method to consistently estimate the model parameters, and at the same time identify the retrospective bias process. The method is based on modeling the reporting process and integrating it into the likelihood function for estimation. This approach leads to consistent and efficient estimates.

Using a particular application (voting in Israel on 1991) we show how the method is used in providing a useful way to interpret individual behavior. We find that Israeli respondents whose vote intention in 1991 differed from their past voting choices report their current choice as if it was their past choice with probability of 0.23.

Our model of the retrospective bias process is quite simple. A more challenging approach is to

[15] This assumption is reasonable with respect to voting behavior, since voters are likely to be unaware of the habit formation (state dependence) process. Shachar (1997) derives the theoretical and empirical implications of such forward-looking behavior.

present a behavioral theory of this process. In other words, instead of considering the retrospective bias process only as an element that distorts the estimates of the choice model, one can focus on this process in an effort to better understand peoples' behavior. Furthermore, a cross-country and cross-time comparison of the percentage of "liars" in the population might provide interesting insights.

7 Appendix

Here we show that:

$$\Pr(d_1^r|x_1,x_2) = \gamma\Pr(d_1 = d_1^r|x_1) + (1-\gamma)\Pr(d_2 = d_1^r|x_1,x_2)$$

Specifically,

$$\begin{aligned}\Pr(d_1^r & |x_1,x_2) = & \Pr(d_1 = d_1^r|x_1)\,[\Pr(d_2 = d_1^r|d_1 = d_1^r,x_2)\,(\gamma + (1-\gamma)) + \Pr(d_2 \neq d_1^r|d_1 = d_1^r,x_2)(\gamma + 0)] \\ & + \Pr(d_1 \neq & d_1^r|x_1)\,[\Pr(d_2 = d_1^r|d_1 \neq d_1^r,x_2)(0 + (1-\gamma)) + \Pr(d_2 \neq d_1^r|d_1 \neq d_1^r,x_2)(0+0)] \end{aligned} \tag{15}$$

where the first line represents the case that the respondent reports her previous choice correctly and the second stands for the case that she mis-reports her previous choice.

When will the respondent report correctly? If the respondent did not switch (i.e., $d_1 = d_2$), she always reports correctly (i.e., with probability $\gamma + (1-\gamma)$) . However, if she switched, she reports correctly, only if she is truthful (i.e., with probability γ).

When will the respondent misreport? Only when she switched ($\Pr(d_1 \neq d_1^r|x_1)\Pr(d_2 = d_1^r|d_1 \neq d_1^r,x_2)$) and she is not truthful (i.e., $(1-\gamma)$).

Rearranging the right hand side of (15), we get:

$$\begin{aligned}\gamma\Pr(d_1 &= d_1^r|x_1)\,[\Pr(d_2 = d_1^r|d_1 = d_1^r,x_2) + \Pr(d_2 \neq d_1^r|d_1 = d_1^r,x_2)] \qquad (16) \\ &+ (1-\gamma)\,[\Pr(d_1 = d_1^r|x_1)\Pr(d_2 = d_1^r|d_1 = d_1^r,x_2) + \Pr(d_1 \neq d_1^r|x_1)\Pr(d_2 = d_1^r|d_1 \neq d_1^r,x_2)]\end{aligned}$$

where the first line represents the cases that the respondent is truthful and the second stands for the cases that she is not. Notice that when the respondent is truthful, it does not matter whether she switched or not:

$$\Pr(d_2 = d_1^r|d_1 = d_1^r,x_2) + \Pr(d_2 \neq d_1^r|d_1 = d_1^r,x_2) = 1 \tag{17}$$

On the other hand, when she is not truthful, it does not matter what was her choice in the previous period:

$$\begin{aligned}\Pr(d_1 &= d_1^r|x_1)\Pr(d_2 = d_1^r|d_1 = d_1^r,x_2) + \Pr(d_1 \neq d_1^r|x_1)\Pr(d_2 = d_1^r|d_1 \neq d_1^r,x_2) \qquad (18) \\ &= \sum_{\widetilde{d}_1 \in \{d_1^r, 1-d_1^r\}} \Pr(d_2 = d_1^r|\widetilde{d}_1,x_2)\Pr(\widetilde{d}_1|x_1) = \Pr(d_2 = d_1^r|x_2,x_1)\end{aligned}$$

Plugging (17) and (18) into (16), we get:

$$\Pr(d_1^r|x_1, x_2) = \gamma \Pr(d_1 = d_1^r|x_1) + (1 - \gamma) \Pr(d_2 = d_1^r|x_1, x_2)$$

References

[1] Abramson, P. and C. Ostrom (1994) "Question wording and partisanship: change and continuity in party loyalties during the 1992 election campaign," *Public Opinion Quarterly*, 58, 21-48.

[2] Abramson, P. and W. Claggett (1992) "The quality of record keeping and racial differences in validated turnout," *The Journal of Politics*, 54, 871-880.

[3] Benewick, R., Birch, A., Blumler, J. and A. Ewbank (1969) "The Floating Voter and the Liberal View of Representation," *Political Studies*, 17, 177-195.

[4] Chamberlain, Gary (1993) "Multivariate Regression Models for Panel Data," in Maddala, G. S., ed. *The econometrics of panel data*. Volume 1. Aldershot, U.K. pages 266-307.

[5] Converse, Philip E. 1976. *The Dynamics of Party Support*. Beverly Hills: Sage.

[6] Eckstein, Z. and K.I. Wolpin (1989) "The specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey", *The Journal of Human Resources*, 24 (4), 562-598.

[7] Franklin, Charles H. (1984) "Issue Preferences, Socialization and the Evolution of Party Identification," *American Journal of Political Science*, 28 (3), 459-478.

[8] Franklin, Charles H. and John E. Jackson (1983) "The Dynamics of Party Identification," *American Political Science Review*, 77(4), 957-973.

[9] Heckman, James (1981a) "Statistical Models for Discrete Panel Data", in Manski, C. and D. McFadden, (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge: The MIT Press.

[10] Heckman, James (1981b) "The Incidental Parameters Problem and the Problem of Initial Condition in Estimating a Discrete Time-Discrete Data Stochastic Process," in Manski, C. and D. McFadden, (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge: The MIT Press.

[11] Himmelweit, H., Biberian, M. and J. Stockdale (1978) "Memory for Past Vote: Implications of a Study of Bias in Recall," *British Journal of Political Science*, 8, 365-375.

[12] Horvath, F. (1982) "Forgotten Unemployment: Recall Bias in Retrospective Data", *Monthly Labor Review*, March, 40-43.

[13] Markus, Gregory B. and Philip E. Converse (1979) "A Dynamic Simultaneous Equation Model of Electoral Choice," *American Political Science Review*, 73 (4), 1055-1070.

[14] Morgenstern, R. and N. Barrett (1974) "The Retrospective Bias in Unemployment Reporting By Sex, Race and Age," *Journal of the American Statistical Association*, 69, 355-57.

[15] Niemi, Richard G. and Herbert F. Weisberg (eds.) (1976) *Controversies in American Voting Behavior*. San Francisco: Freeman & Comp.

[16] Niemi, Richard G. and Herbert F. Weisberg (eds.) (1984) *Controversies in Voting Behavior*. Washington D.C.: Congressional Quarterly Press.

[17] Niemi, Richard G. and Herbert F. Weisberg (eds.) (1993) *Classics in Voting Behavior*. Washington D.C.: Congressional Quarterly Press.

[18] Page, Benjamin I. and Calvin C. Jones (1979) "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote," *American Political Science Review*, 73(2), 1071-1089.

[19] Presser, S. (1990) "Can changes in context reduce vote overreporting in surveys?" *Public Opinion Quarterly*, 54, 586-593.

[20] Presser, S. and M. Traugott (1992) “Little white lies and social science models: correlated response errors in a panel study of voting,” *Public Opinion Quarterly*, 56, 77-86.

[21] Shachar, Ron (2003) “Party Loyalty as Habit Formation,” *Journal of Applied Econometrics*, 18 (3), 251-269.

[22] Shachar Ron (1997) “Party Loyalty as Habit Formation,” Yale School of Management Working Paper Series H, No. 6.

[23] Shachar, Ron (1994) “A Diagnostic Test for the Sources of Persistence in Individuals’ Decisions,” *Economics Letters*, 45, 7-13.

[24] Torelli, N. and U. Trivellato (1993) “Modelling Inaccuracies in Job-Search Duration Data,” *Journal of Econometrics*, 59, 187-211.

[25] Weir, B. (1975) “The Distortion of Voter recall,” *American Journal of Political Science*, 19, 53-62.

[26] Wolak, Frank (1989) “Testing Inequality Constraints in Linear Econometric Models,” *Journal of Econometrics*, 41 (2), 205-235.

[27] Zuckerman, A. (1990) “The Flow of the Vote in Israel: A Reconsideration of Stability and Change”, in Arian, A. and Shamir, M. (eds), *The Elections in Israel - 1988*, (Boulder, Colorado, Westview).

Table 1
Monte Carlo experiments

	True value	Misreporting past choices: $\bar{\theta} = \frac{1}{1000}\sum_{r=1}^{1000}\hat{\theta}$	Misreporting past choices: $\sqrt{\frac{1}{999}\sum_{r=1}^{1000}\left(\hat{\theta}-\bar{\theta}\right)^2}$	Misreporting the final choice: $\bar{\theta} = \frac{1}{1000}\sum_{r=1}^{1000}\hat{\theta}$	Misreporting the final choice: $\sqrt{\frac{1}{999}\sum_{r=1}^{1000}\left(\hat{\theta}-\bar{\theta}\right)^2}$
β_1	0	normalized	---	normalized	---
β_2	0.5	0.5041	0.1044	0.5051	0.1159
β_3	1	1.0048	0.1082	1.0092	0.1108
β	1	1.0046	0.0664	1.0063	0.0674
δ	1	0.9943	0.1524	0.9951	0.2677
γ	0.75	0.7518	0.0477	0.7593	0.0830

Note: The model was simulated and estimated 1000 times, and the starting values for all the parameters were 0.2.

Tables 2
Voting paths

Table 2a Only respondents who could not vote prior to 1991		
d_{91}	Number of respondents	Percentage
1	19	29.2
0	46	70.8

Table 2b Only respondents who could not vote prior to 1988			
d_{88}	d_{91}	Number of respondents	Percentage
1	1	23	41.82
1	0	2	3.64
0	1	2	3.64
0	0	28	50.91

Table 2c Only respondents who had the right to vote prior to 1988				
d_{84}	d_{88}	d_{91}	Number of respondents	Percentage
1	1	1	115	39.25
1	1	0	5	1.71
1	0	1	3	1.02
1	0	0	18	6.14
0	1	1	6	2.05
0	1	0	6	2.05
0	0	1	7	2.39
0	0	0	133	45.39

	Table 3 Exogenous variables					
	could not vote prior to 1991		could not vote prior to 1988		had the right to vote prior to 1988	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
y_{91}	-4.2	7.72	-1.2	7.87	-0.88	9.39
y_{88}			-0.47	8.51	-0.02	8.93
y_{84}					2.95	7.56
w_{91}	2.29	1.2	2.69	1.32	2.87	1.38
w_{88}			2.76	1.14	2.91	1.35
w_{84}					2.9	1.41
Age	1.22	0.67	1.2	0.56	3.62	1.43
Spharadi	0.63	0.49	0.60	0.49	0.53	0.50
Education	12.31	1.73	12.78	1.41	11.85	3.35
Religious	3.22	0.74	3.45	0.66	3.46	0.60

	Table 4 Maximum likelihood estimates							
	Model I		**Model II**		**Model III**		**Model IV**	
	Version 1	**Version 2**	**Without initial obs.**	**With initial obs.**	**Without initial obs.**	**With initial obs.**	**Without initial obs.**	**With initial obs.**
α	0.080 (0.01)	0.141 (0.02)	0.068 (0.01)	0.077 (0.01)	0.131 (0.03)	0.130 (0.03)	0.069 (0.01)	0.077 (0.01)
β	0.917 (0.06)	1.409 (0.18)	0.817 (0.07)	0.916 (0.08)	1.38 (0.19)	1.45 (0.26)	0.825 (0.07)	0.950 (0.09)
δ			0.663 (0.07)	0.613 (0.07)	0.642 (0.13)	0.747 (0.18)	0.622 (0.08)	0.641 (0.07)
μ_0	-3.05 (0.54)	-5.41 (1.34)	-2.39 (0.57)	-2.70 (0.62)	-5.16 (1.28)	-5.39 (1.38)	-2.27 (0.59)	-2.51 (0.66)
μ_{Age}	0.053 (0.04)	0.064 (0.08)	0.018 (0.04)	0.032 (0.05)	0.008 (0.07)	0.046 (0.09)	0.0276 (0.04)	0.025 (0.05)
$\mu_{Spharadi}$	-0.253 (0.14)	-0.093 (0.28)	-0.190 (0.15)	-0.195 (0.16)	0.130 (0.26)	0.181 (0.27)	-0.193 (0.15)	-0.202 (0.17)
$\mu_{Education}$	-0.023 (0.02)	-0.046 (0.05)	-0.042 (0.02)	-0.044 (0.03)	-0.038 (0.05)	-0.029 (0.05)	-0.041 (0.02)	-0.051 (0.03)
$\mu_{\mathrm{Re}\,ligious}$	0.125 (0.11)	0.435 (0.26)	0.112 (0.12)	0.127 (0.13)	0.366 (0.23)	0.334 (0.24)	0.069 (0.12)	0.070 (0.14)
γ							0.812 (0.09)	0.766 (0.08)
$\frac{100}{\sigma_{\varepsilon'}}$				6.166 (8.88)		15.141 (20.47)		0.284 (8.68)
$\ln[L(\theta)]$	-257.07	-56.49	-222.17	-221.91	-48.49	-48.48	-219.39	-216.33

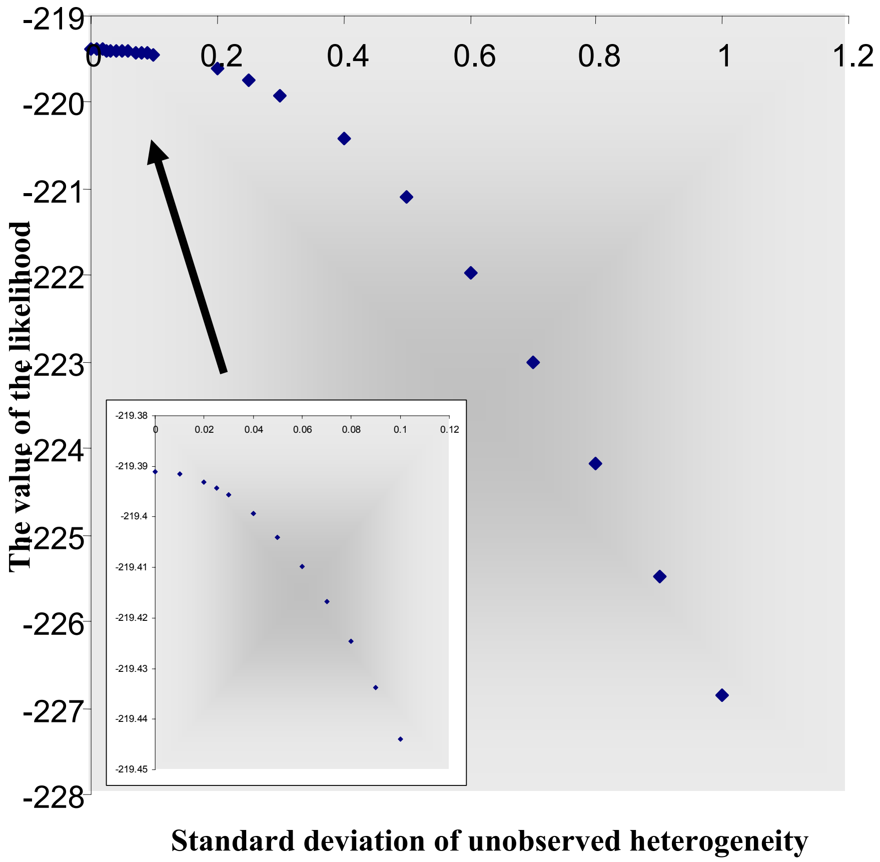
Notes:

(1) the number of observations is 413.

(2) standard errors in parenthesis.

(3) $\sigma_{\varepsilon'}$ is the standard deviation of ε in periods prior to 1984.

Figure 1



Note: The smaller figure enlarge the area for which the standard devision is smaller than 0.1.