No. 835

Handling "Don't Know" Survey Responses: The case of Japanese Voters on Party Support

by

LEE, Sang-Gil and KANAZAWA, Yuichiro

October 1999

Handling “Don’t Know” Survey Responses: The Case of Japanese Voters on Party Support

LEE, Sang-Gil*

Doctoral Program in Policy and Planning Sciences

University of Tsukuba

KANAZAWA, Yuichiro

Institute of Policy and Planning Sciences

University of Tsukuba

October 25, 1999

*Correspondence: 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573 Japan. E-mail address: slee@sk.tsukuba.ac.jp. An earlier version of this paper was presented at the annual conference of Japanese Statistical Society in Tokyo, Japan, July 1998. We wish to thank Ikuo Kabashima and Yoshito Ishio for permission to use the data set.

Abstract

The problem of false negatives, people who really have attitudes but refrain from expressing them, could seriously bias the analysis, but has largely been neglected. Using a survey data including a number of "Don't know" responses, this paper examined whether "Don't know" respondents had underlying attitudes. We treated these nonresponses as nonignorably missing, in the sense that "Don't know" responses are related to the answer of question in some partially unknown way. We proposed a method to estimate parameters in logit model when the covariates are nonignorably missing. The method simultaneously employed two generalized linear models: the proportional odds model for the response variable "Party-Support", and the multinomial logit model for the nonresponse. We found that "Don't know" responses to the Cabinet support question depended on whether the respondents supported the Cabinet, indicating the existence of false negatives. We also found that determining which party to support was based on voters' ideology, city size and stance toward the Cabinet, even with the false negatives.

Keywords: EM algorithm; False Negatives; Missing Mechanism; Multinomial Logit Model; Proportional-odds Model.

1 Introduction

Survey data, collected by a questionnaire or interview, often include "Don't know" (DK) responses to one or several questions. In some surveys, the

DK responses can be considered as valid and should form one answer. One such example would occur in pre-election polls of United States, where DK respondents form undecided voters who might be more easily influenced to one side or the other. Since such voters could affect the election result as a group, their very existence is meaningful. However, in cases where we have no reason to treat DK responses as one independent answer because all the possible answers are prepared or the measurement scale is well established, we can not help but treat them as missing.

The missing responses, however, are subject to interpretation. It may be due to the lack of knowledge or uncertainty about the meaning of the question asked; these unintended DK survey responses are relatively harmless, in the sense that they can be safely excluded from an analysis because we usually have no reason to believe that they do not occur randomly. We should be aware of the existence of ***false negatives***, "people who really have an underlying attitude, nonetheless, do not express their opinion if asked in an interview" (Gilljam and Granberg 1993, pp. 348-49); these intended DK responses are especially to be expected in the questions of embarrassing or private nature, for example, as in alcohol consumption, drug abuse, sexual activities, or income.

There would be two types of false negatives. One is ***ignorable false negatives*** whose missing-data mechanism, the reason whether one's opinion is not expressed, depends only on the fully observed answers to the questions, but not on the unobserved one; the other is ***nonignorable false negatives*** whose nonresponses are allowed to depend on the unobserved one. For instance, we might be able to imagine a case or situation where the voters with the most liberal or conservative ideology tend to answer DK because they do not want to be regarded as being politically biased by interviewers. We hypothesize that the nonignorable false negatives exist among Japanese voters, mostly due to the fifty-four years of almost consecutive conservative party rule in Japanese politics.

A number of studies on Japanese voters' behavior have been published. Watanuki et al. (1986); Miyake (1989, chap. 3) used the ***complete-case analysis***, where the units with at least one DK response were excluded, ignoring the problem associated with missing value. As the excluded units might have an observation on some of the questions, they still carry an information on the questions. Furthermore, if the units are dropped out of the original sample in a systematic manner, the complete-case data would not be a random subsample, and thus the analysis could be highly biased. To overcome these problems, some researches, e.g., Richardson (1988); Miyake (1995, chap. 5); Kabashima (1998, chap. 11), used the ***mean-imputation analysis***, where missing values were replaced with the average of observed values. Even then, estimated variances of the parameter estimates from artificially complete data set are invalid in general (Little and Rubin 1987).

Only Abe et al. (1998) used an ignorable nonresponse model assuming the DK responses are missing at random (see Rubin 1976), and addressed the problem of bias associated with complete-case analysis. The analysis under ignorable assumption, however, could not address the problem of nonignorable false negatives, and could also result in biased estimation (Little and Rubin 1987).

In this article, using a survey data on Japanese voters in the early 1997 including a number of DK responses, we examine whether people who answer DK really have underlying attitudes. We treat these nonresponses as missing because the true intention of voter is unknown but must be in one of the possible answers provided by interviewers, and assume that the occurrences of missing values are nonignorable as defined by Rubin (1976). Note that when there is nonignorable nonresponses, serious biases in parameter estimates may results if one does not model the missing-data mechanism (Rubin 1976; Little and Rubin 1987).

Most models in the literature of this type concern the case where the nonresponses are confined to a single variable (see Baker and Laird 1985; Forster and Smith 1988; Fitzmaurice et al. 1996). For more than two variables with missing values, Ibrahim et al. (1999) proposed a model incorporating the missing-data mechanism. In his approach, the missing-data mechanism is modeled by a sequence of conditional distributions specified by a sequence of logistic regressions, and the indexing parameters for each distribution are assumed to be distinct. We propose instead to represent the distribution for missing-data mechanism as a multinomial and apply the multinomial logit model to the distribution. This renders the additional assumptions on indexing parameters by Ibrahim et al. (1999) to be unnecessary. In this sense this paper employs a model different from that of Ibrahim et al. (1999).

The method used in this paper is principally aimed at the generalized linear model to estimate parameters in logit model, consisting of the proportional odds model for the regression of ordinal response variable "Party-Support" on covariates and the multinomial logit model incorporating the missing-data mechanism on the model. The advantage of this approach is that it makes the missing-data mechanism transparent, so that it enables us to pay careful attention to the problem of false negatives.

We shall apply the EM algorithm by method of weight introduced in Ibrahim (1990) to estimate parameters. The results of estimation and hypothesis testing could be sensitive to the choice of model for the missing-data mechanism. To determine a suitable model, we fit different models and evaluate the fit by model selection criteria.

In section 2, an explanation of the data is given. The method is described in section 3 generally, and in appendix C in greater detail. The model selection and the estimation results are presented in section 4 and 5. Finally some discussion for profiling voters on party support are given in section 6, along with the validity of nonignorable false negatives assumption. The appendices also contain a brief chronology of the realignment of Japanese political parties, and a summary of coding data.

2 Data

The Jiji Monthly (JM) survey, conducted monthly since June of 1960 by the CENTRAL RESEARCH SERVICES, is an ongoing survey designed to examine the contemporary Japanese political attitude and behavior.

We chose the data from the JM survey for three consecutive months from January 1997, because there had been no major changes in Japanese politics and presumably the profiles of Japanese voters on party support would be rather stable during the period (see appendix A).

Using a two-staged cluster sampling method, the survey used in this paper was conducted after the 41st Lower House election of Japan and employed a face-to-face interview of 2,000 eligible voters across the nation. Of the target sample, 69.7% on the average responded to the survey conducted.

We focus attention on the following four key questions concerning the party support.

Party-Support	Which political party do you usually support?
Ideology	What is your political position?
City-Size	How large is your city?
Cabinet-Support	Do you support the Prime Minister Hashimoto's Cabinet?

We shall try to investigate the effects of covariates on a categorical response variable "Party-Support", paying careful attention to the problem of false negatives. As covariates, "Ideology" represents the political position of voters on ten-point ordinal scale, ranging from liberal to conservative; "City-Size" measures the population size of the city where the voters reside in as three-point scale; and "Cabinet-Support" is a binary measurement of the voters' stance toward the Cabinet. Each of variables is re-coded as described in appendix B.

There are a lot of factors that could exert the influence on political party support, for instance, socioeconomic background such as gender, age, income and city size; political attitudes such as political satisfaction and ideology; and so on (see, e.g., Miyake 1998). However, some preliminary analyses indicate that the most parsimonious model is to include these three covariates, and these covariates were found to be statistically significant in explaining the party support as in Inoguchi (1983, chap. 3); Miyake (1995, chap. 5); Abe, et al. (1998); Kabashima and Ishio (1998).

With surveys of this kind, it is inevitable that there are some DK respondents. Table 1 illustrates the problem of DK responses aforementioned.

Table 1: Number of Samples

Month	January	February	March
Sample Size	583	585	521
# of Cases without a DK response	519	514	451
% of Complete Case	89.0	87.9	86.6

The first row is the sample size, the second the number of sample without a single DK response, and the third the percentage of the latter relative to the former. Throughout the three months, only about 87-89% of respondents participating in the survey completed all questions. Thus the complete-case analysis ignoring the remaining 11-13% of the data is very likely to bring nonresponse bias.

In our data, incompleteness is only due to two covariates, "Ideology" and "Cabinet-Support". For the ideology question, we have no reason to treat the DK responses as one answer because the liberal-conservative scale is well established. For the Cabinet support question, although it is conceivable that the voters who lukewarmly support the Cabinet or who do not dislike the Cabinet enough to express their nonsupport probably answer DK, their existence does not warrant a formation of an independent answer DK. Thus, we treat the DK responses to these covariates as missing and broadly assume the missing-data mechanism is nonignorable. These treatments for DK responses will be scrutinized in the following sections.

3 Method

This section presents a regression model for categorical data under the nonignorable nonresponse assumption to missing covariates. We employ simultaneously the proportional odds model for the response variable "Party-Support", and the multinomial logit model for missing-data mechanism. See, e.g., McCullagh and Nelder (1989, chap. 5); Agresti (1990, chap. 9) on these two models.

Let us denote the response variable as Y, the covariates as X, and the missingness indicator representing the missing-data mechanism as R. Our model assumes that the categorical variable, Y or R, has multinomial distribution. For five categories representing an answer to "Party-Support", i.e., 1 (JCP), 2 (SDP), 3 (DP), 4 (NFP), 5 (LDP), we employ the following proportional odds model:

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = b_j - \beta_1 \text{Ideology}_i - \beta_2 \text{City}_i - \beta_3 \text{Cabinet}_i, \qquad j = 1, \ldots, 4 \tag{3.1}$$

where $\gamma_{ij} = \Pr(Y_i \leq j | \text{Ideology}_i, \text{City}_i, \text{Cabinet}_i, \boldsymbol{\beta})$ denotes a cumulative probability up to and including the jth category of "Party-Support" for the ith observation ($i = 1, \ldots, n$), and where $\boldsymbol{\beta} = (b_1, \ldots, b_4, \beta_1, \beta_2, \beta_3)^T$ are the corresponding regression coefficients. We employed the model since many researches argued that Japanese political parties can be ordered with respect to its ideology or political hue. See, e.g., Kabashima (1998, chap. 8).

Next we consider the missing-data mechanism model. Since we have an incomplete data due to two covariates, each observation can be classified into any one of four possible ***patterns of missingness***: 1 ("Ideology" and "Cabinet-Support" are missing), 2 ("Ideology" is observed but "Cabinet-Support" is missing), 3 ("Ideology" is missing but "Cabinet-Support" is observed) and 4 ("Ideology" and "Cabinet-Support" are observed). Then setting the pattern 4 as a baseline, the model is written as

$$\begin{aligned}\log\left(\frac{\psi_{ik}}{\psi_{i4}}\right) &= \alpha_{k0} + \alpha_{k1}\text{Ideology}_i + \alpha_{k2}\text{City}_i + \alpha_{k3}\text{Cabinet}_i \\ &\quad + \alpha_{k4}\text{Party}_i, \qquad k = 1, \ldots, 3, \end{aligned} \tag{3.2}$$

where $\psi_{ik} = \Pr\{R_i = k|\text{Ideology}_i, \text{City}_i, \text{Cabinet}_i, \text{Party}_i, \boldsymbol{\alpha}\}$ is a probability of the kth pattern of missingness, and where $\boldsymbol{\alpha} = (\alpha_{10}, \alpha_{11}, \ldots, \alpha_{34})^T$ are the corresponding regression coefficients. Note that this model consists of three logit equations with separate parameters for each. We employed the model for the patterns of missingness should not be ordered.

In (3.2) we allow the probability ψ_{ik} depends not only on completely observed response variable but on partially missing covariates as well, which implies the missing-data mechanism could be nonignorable. For example, if $\alpha_{k1} = \alpha_{k3} = 0$, then the missingness does not depend on covariates with missing values, implying the missing-data mechanism is ignorable; if $\alpha_{k1} \neq 0$ and/or $\alpha_{k3} \neq 0$ then the mechanism is nonignorable.

Finally, following the formulation of Little and Rubin (1987, chap. 11), the ***complete data likelihood*** for n independent observations is given by

$$f(r, y, x|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho}) = \prod_{i=1}^{n} f(y_i|x_i, \boldsymbol{\beta}) f(r_i|x_i, y_i, \boldsymbol{\alpha}) f(x_i, \boldsymbol{\rho}), \tag{3.3}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ are assumed to be distinct sets of indexing parameters for corresponding distributions. In (3.3), the first two components are specified by (3.1) and (3.2) respectively, and the third component corresponds to the marginal distribution of x_i assumed to follow a multinomial distribution with cell probabilities $\boldsymbol{\rho}$.

Based on the complete data likelihood in (3.3), ML estimates with incomplete data can be obtained via the EM algorithm (Dempster et al. 1977). In particular, we apply the EM by method of weight introduced by Ibrahim (1990), which is equivalent to doing pseudo complete data ML estimation: (1) evaluate weights for "filled-in" $N = \sum_{i=1}^{n} n_i$ observations from parameter estimates, where n_i is the number of possible distinct covariate patterns for the ith observation (E-step), (2) estimate parameters with weighted filled-in data (M-step), (3) iterate E and M-step until convergence. See appendix C for detail.

Our primary interest is in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, with $\boldsymbol{\rho}$ being as nuisance parameter. Note that $\boldsymbol{\alpha}$ is very important since it index the distribution for the missing-data mechanism.

4 Inestimability

The M-step involves three separate maximization for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$. The ML estimates of $\boldsymbol{\rho}$ are easily obtained since a closed form solution can be found. The likelihood equations for $\boldsymbol{\beta}$ and for $\boldsymbol{\alpha}$, however, are nonlinear functions, so the numerical optimization method is required to obtain the ML estimates. In general, the numerical method in logit model will find a solution very readily, but infinite estimators are possible when the model is not appropriate in the sense of providing a good fit to the data, and the data are badly conditioned (see Manski and McFadden 1983).

Therefore, for certain models, it may happen that the parameters are not identifiable. To illustrate the problem, consider the following nonignorable nonresponse model, consisting of the proportional odds model for party support

$$\log(\gamma_{ij}/(1-\gamma_{ij})) = b_j - \beta_1 \text{Ideology}_i - \beta_2 \text{City}_i - \beta_3 \text{Cabinet}_i, \qquad j = 1, 2, 3, 4,$$

and the multinomial logit model for missing-data mechanism

$$\log(\psi_{ik}/\psi_{i4}) = \alpha_{k0} + \alpha_{k1} \text{Ideology}_i + \alpha_{k4} \text{Party}_i, \qquad k = 1, 2, 3.$$

We fitted this model to January data, but some estimates of the multinomial logit model were infinite. In figure 1, we plotted iterative ML estimates of the multinomial logit model up to the 100th EM iteration. We observed that most of parameters converged rapidly, but α_{10} and α_{11} were not bounded and unique.

For the multinomial logit model for missing-data mechanism, our incomplete data can be represented in terms of a $4 \times 5 \times 4$ table, cross-classifying the missingness indicator R for ideology and party support. Table 2 provides the contingency table for the incomplete data of January.

Table 3 provides the contingency tables for pseudo complete data, mentioned in the preceding section, on the 5th, 20th, 50th, and 100th EM iterations. In the table, the partially classified data in table 2 are filled in the completely cross-classified table according to its possible set of realization, along with corresponding weights evaluated in the E-step.

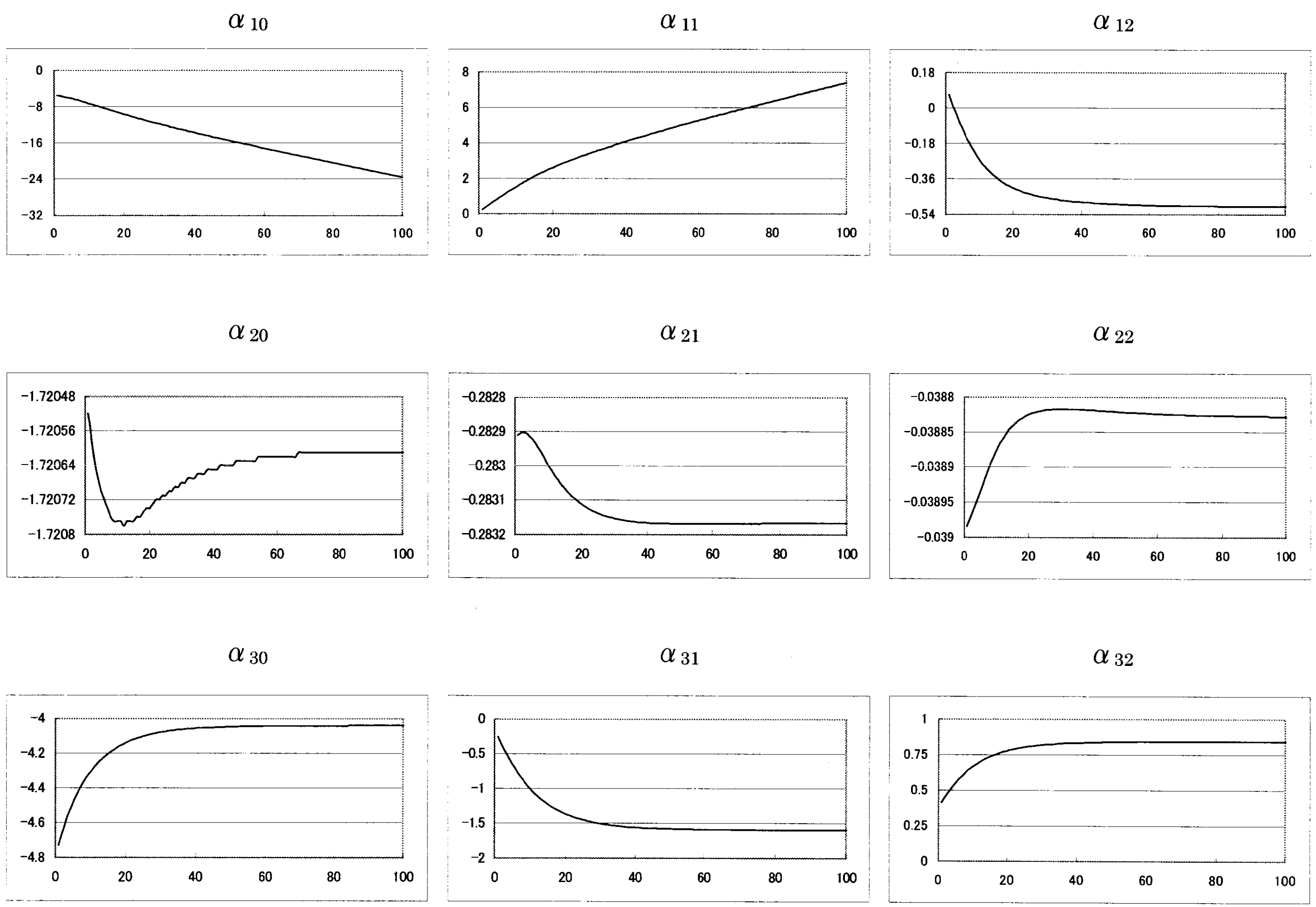

Figure 1: Parameter Estimates up to the 100th EM iteration

Table 2: Contingency Table of Ideology, Party and Missingness Indicator R, and Partially Classified Margin for Missing at Ideology

		R				
Ideology	Party	1	2	3	4	row total
1	1	0	3	0	17	20
	2	0	0	0	22	22
	3	0	1	0	11	12
	4	0	1	0	24	25
	5	0	1	0	15	16
2	1	0	1	0	6	7
	2	0	2	0	21	23
	3	0	2	0	21	23
	4	0	12	0	50	62
	5	0	11	0	120	131
3	1	0	0	0	2	2
	2	0	0	0	3	3
	3	0	0	0	4	4
	4	0	1	0	23	24
	5	0	8	0	180	188
(Missing)	1	0	0	0	0	0
	2	1	0	1	0	2
	3	0	0	0	0	0
	4	0	0	4	0	4
	5	4	0	1	0	15
	column total	5	43	16	519	583

Table 3: Contingency Table of Ideology, Party and Missingness Indicator R for Weighted Filled-in Data

(a) the 5th EM iteration

Ideology	Party	R 1	R 2	R 3	R 4
1	1	0	3	0	17
	2	0.21	0	0.72	22
	3	0	1	0	11
	4	0	1	1.79	24
	5	0.05	1	1.2	15
2	1	0	1	0	6
	2	0.54	2	0.26	21
	3	0	2	0	21
	4	0	12	1.92	50
	5	0.91	11	5.5	120
3	1	0	0	0	2
	2	0.24	0	0.02	3
	3	0	0	0	4
	4	0	1	0.29	23
	5	3.05	8	4.31	180
	total	5	43	16.01	519

(b) the 20th EM iteration

Ideology	Party	R 1	R 2	R 3	R 4
1	1	0	3	0	17
	2	0.02	0	0.85	22
	3	0	1	0	11
	4	0	1	2.59	24
	5	0	1	2.56	15
2	1	0	1	0	6
	2	0.27	2	0.15	21
	3	0	2	0	21
	4	0	12	1.31	50
	5	0.2	11	6.08	120
3	1	0	0	0	2
	2	0.71	0	0.01	3
	3	0	0	0	4
	4	0	1	0.1	23
	5	3.8	8	2.36	180
	total	5	43	16.01	519

(c) the 50th EM iteration

Ideology	Party	R 1	R 2	R 3	R 4
1	1	0	3	0	17
	2	0	0	0.88	22
	3	0	1	0	11
	4	0	1	2.8	24
	5	0	1	3.1	15
2	1	0	1	0	6
	2	0.04	2	0.12	21
	3	0	2	0	21
	4	0	12	1.13	50
	5	0.03	11	6.02	120
3	1	0	0	0	2
	2	0.95	0	0	3
	3	0	0	0	4
	4	0	1	0.07	23
	5	3.97	8	1.88	180
	total	4.99	43	16	519

(d) the 100th EM iteration

Ideology	Party	R 1	R 2	R 3	R 4
1	1	0	3	0	17
	2	0	0	0.88	22
	3	0	1	0	11
	4	0	1	2.81	24
	5	0	1	3.14	15
2	1	0	1	0	6
	2	0	2	0.12	21
	3	0	2	0	21
	4	0	12	1.12	50
	5	0	11	6.01	120
3	1	0	0	0	2
	2	1	0	0	3
	3	0	0	0	4
	4	0	1	0.07	23
	5	4	8	1.86	180
	total	5	43	16.01	519

We see from table 3 that some of the weighted filled-in data collapse to zero as the EM iteration proceeds. The reason for this is, since the weights are probabilities of the possible outcomes of incomplete observations, unusual outcomes tend to receive less weight than typical outcomes. Consequently, the weighted filled-in data are nearly zero when the corresponding cells are unusual in the sense of weight evaluated in each iteration. At this iteration the Hessian is nearly singular, and thus the second order conditions for finding a maximum are not fulfilled. Furthermore, we see that the data are badly conditioned. Compared to the pattern 4 of R, the pattern 1 has extremely small number of data involving many zeros. This leads to the lack of sufficient information on the parameters α_{10} and α_{11}, resulting in infinite estimates.

The number of identifiable models is therefore limited. The models in which the missing-data mechanism is explained by two covariates (Ideology, Cabinet), or (Ideology, City), and by more than three covariates were inestimable in at least one of the three months. The inestimability arises often in the models incorporating missing-data mechanism. See Little and Rubin (1987, p. 239); Baker and Laird (1988) for more details.

5 Regression Results

Table 4 provides the estimation results of January data for three models consisting of missing-data mechanism model with different sets of covariates for $f(r_i|x_i, y_i, \boldsymbol{\alpha})$ but keeping the regression model for $f(y_i|x_i, \boldsymbol{\beta})$ fixed:

A: $\log(\psi_{ik}/\psi_{i4}) = \alpha_{k0} + \alpha_{k2}\text{City}_i + \alpha_{k3}\text{Cabinet}_i,$

B: $\log(\psi_{ik}/\psi_{i4}) \; = \alpha_{k0} + \alpha_{k1}\text{Ideology}_i + \alpha_{k2}\text{City}_i,$

C: $\log(\psi_{ik}/\psi_{i4}) = \alpha_{k0} + \alpha_{k2}\text{City}_i + \alpha_{k4}\text{Party}_i,$

for $i = 1, \ldots, n$; $k = 1, \ldots, 3$. Note that both model A and B assume nonignorable nonresponse mechanism as they allow the probability ψ_{ik} to depend on covariate with missing value, i.e., "Cabinet-Support" and "Ideology" respectively, whereas model C assumes ignorable nonresponse as it depends only on the completely observed "City-Size" and "Party-Support".

Table 4: Estimation Results for Three Models in January

	Y on X†				R on X and Y‡				
Model	Logit	Ideology	City	Cabinet	Logit	Constant	Ideology	City	Cabinet
A:	$\log(\frac{\gamma_j}{1-\gamma_j})$	1.450	0.435	-1.475	$\log(\frac{\pi_1}{\pi_4})$	-1.919		-0.686	-1.151
		(9.929)	(3.062)	(-7.280)		(-0.529)		(-1.020)	(-0.410)
					$\log(\frac{\pi_2}{\pi_4})$	-5.267		0.108	1.709
						(-3.515)		(0.402)	(2.397)
					$\log(\frac{\pi_3}{\pi_4})$	-4.681		0.293	0.431
						(-3.705)		(0.750)	(0.803)
AIC=1615.05, Deviance=1583.05									
B:	$\log(\frac{\gamma_j}{1-\gamma_j})$	1.454	0.433	-1.504	$\log(\frac{\pi_1}{\pi_4})$	-5.979	1.046	-0.639	
		(9.999)	(3.056)	(-7.278)		(-1.269)	(0.608)	(-0.965)	
					$\log(\frac{\pi_2}{\pi_4})$	-1.607	-0.312	-0.103	
						(-2.446)	(-1.431)	(-0.436)	
					$\log(\frac{\pi_3}{\pi_4})$	-4.982	0.456	0.202	
						(-2.561)	(0.607)	(0.533)	
AIC: 1635.04, Deviance: 1603.04									
C:	$\log(\frac{\gamma_j}{1-\gamma_j})$	1.447	0.433	-1.498	$\log(\frac{\pi_1}{\pi_4})$	-4.260		-0.650	
		(9.932)	(3.057)	(-7.250)		(-1.914)		(-0.973)	
					$\log(\frac{\pi_2}{\pi_4})$	-1.835		-0.088	
						(-2.805)		(-0.367)	
					$\log(\frac{\pi_3}{\pi_4})$	-5.171		0.151	
						(-3.500)		(0.399)	
AIC: 1639.48, Deviance: 1607.48									

Asymptotic t-values in parentheses.

†To make the table concise the estimates in intercept term are omitted.

‡In logit, 1: missing at Ideology and Cabinet, 2: missing at Cabinet, 3: missing at Ideology, and 4: All observed

The estimated coefficients, their t-values based on Louis' formula (Louis, 1982), and model selection criteria are given in table 4. We observed that the estimated coefficients and their t-values for the regression of Y on X were slightly different from each other. Based on the AIC and the Deviance, however, we found that the model A provided a better fit to the data than any other model considered. Also, only the model A contained a significant covariate at (two-tailed) 5% level, i.e., the occurrence of the missingness pattern 2 ("Cabinet-Support" is missing) was significantly correlated with the covariate "Cabinet-Support".

Similarly, for February and March, we evaluated the fit by the AIC and the Deviance in table 5 and found the model A fitted the data better.

Table 5: Model Selection in February and March

Month	Model	AIC	Deviance
Feburary	A	1570.10	1538.10
	B	1585.16	1553.16
	C	1601.61	1569.61
March	A	1465.72	1433.72
	B	1470.24	1438.24
	C	1476.18	1444.18

Consequently, among the identifiable models discussed in the previous section, we chose the model A in which the missing-data mechanism related to two covariates "City-Size" and "Cabinet-Support". Monthly results for the model A are given in table 6.

We examine first the results for the regression R on X and Y. In this model, the first three missingness patterns are compared to the completely

Table 6: Monthly Results for Nonignorable Nonresponse Model

Month	Y on X†				R on X and Y‡			
	Logit	Ideology	City	Cabinet	Logit	Constant	City	Cabinet
January:	$\log(\frac{\gamma_j}{1-\gamma_j})$	1.450	0.435	-1.475	$\log(\frac{\pi_1}{\pi_4})$	-1.919	-0.686	-1.151
		(9.929)	(3.062)	(-7.280)		(-0.529)	(-1.020)	(-0.410)
					$\log(\frac{\pi_2}{\pi_4})$	-5.267	0.108	1.709
						(-3.515)	(0.402)	(2.397)
					$\log(\frac{\pi_3}{\pi_4})$	-4.681	0.293	0.431
						(-3.705)	(0.750)	(0.803)
February:	$\log(\frac{\gamma_j}{1-\gamma_j})$	1.435	0.230	-2.249	$\log(\frac{\pi_1}{\pi_4})$	-12.179	1.149	3.177
		(8.951)	(1.529)	(-10.230)		(-1.825)	(1.762)	(0.981)
					$\log(\frac{\pi_2}{\pi_4})$	-4.981	-0.048	1.696
						(-3.526)	(-0.192)	(2.457)
					$\log(\frac{\pi_3}{\pi_4})$	-5.293	0.792	0.200
						(-4.392)	(2.121)	(0.412)
March:	$\log(\frac{\gamma_j}{1-\gamma_j})$	1.300	0.676	-2.426	$\log(\frac{\pi_1}{\pi_4})$	-2.124	0.046	-1.467
		(7.626)	(4.260)	(-10.120)		(-0.836)	(0.096)	(-0.736)
					$\log(\frac{\pi_2}{\pi_4})$	-4.272	0.155	1.067
						(-3.965)	(0.633)	(1.958)
					$\log(\frac{\pi_3}{\pi_4})$	-4.043	0.264	0.194
						(-3.649)	(0.727)	(0.392)

Asymptotic t-values in parentheses.
†To make the table concise the estimates in intercept term are omitted.
‡In logit, 1: missing at Ideology and Cabinet, 2: missing at Cabinet, 3: missing at Ideology, and 4: All observed

observed pattern of missingness.

"City-Size" had positive effect on "Missing at Ideology" versus "All observed", and therefore, compared to "All observed", "Missing at Ideology" was more likely to occur for those who lived in towns or villages.

"Cabinet-Support" exerted a noticeable positive effect on the odds of "Missing at Cabinet" instead of "All observed" for all three months. In particular, for a given "City-Size", the estimated odds that the missingness pattern was 2 (Missing at "Cabinet") instead of 4 (All observed) was about $\exp(1.067) = 2.9$ to $\exp(1.709) = 5.5$ times higher for people who did not support the Cabinet than those who did support. That is, holding other conditions fixed, people who did not support the Cabinet were at least 2.9 times more likely to respond DK.

The data appear to include intended nonresponse justifying the existence of nonignorable false negatives on "Cabinet-Support". Therefore, our finding suggests that the ignorable false negatives assumption made by Abe et al. (1998) may not be sufficient.

Next, we examine the results for the regression Y on X. We found that "Ideology" and "Cabinet-Support" was significant for all three months at (two-tailed) 5% level; "City-Size" was significant in two months, but we chose to leave it in the model.

The signs of estimated regression coefficients indicated that in the early 1997 a Japanese voter whose ideology was more conservative, who supported the Hasimoto's Cabinet, and who lived in smaller city was more likely to support a conservative-leaning party.

Holding other conditions fixed, the odds that a voter with conservative ideology supported a conservative-leaning party rather than a liberal one was about $\exp(2 \times 1.300) = 13.5$ to $\exp(2 \times 1.450) = 18.2$ times as large as a voter with liberal ideology. Therefore, voters were at least 13.5 times likely to support the political party ideologically similar, despite the fact that ideological differences among voters become less pronounced in the study of Japanese voters' behavior. This finding is consistent with the results of Miyake (1989, chap. 3; 1995, chap. 5) and Abe et al. (1998).

"City-Size" was significant except in February, although its magnitude of coefficients were much smaller than those of other two covariates. The odds that a voter who lived in a town or village supported a conservative-leaning party rather than a liberal one was about $\exp(2 \times 0.230) = 1.6$ to $\exp(2 \times 0.676) = 3.9$ times as large as a voter who lived in a big city. It was found in Miyake (1989, chap. 3) that the support for the conservative-leaning LDP was more firm in rural region, while the liberal-leaning parties were more popular in urban area throughout 1980's. Our founding supports this pattern still exist in the early 1997.

Since "Cabinet-Support" was consistently and highly significant, and its coefficient was negative, a voter who affirmatively evaluated the performance of the Cabinet was inclined to support a conservative-leaning party. In particular, the odds that a voter who did support the Cabinet supported the conservative leaning party rather than a liberal one was about $\exp(1.475) = 4.4$ to $\exp(2.426) = 11.3$ times as large as those who did not support the Cabinet. The most conservative party in our research was the LDP which was the senior partner of the coalition government at the time, and the policies of government strongly reflected those of LDP. Consequently, voters who were supporting the Cabinet were likely to support the LDP.

6 Discussion

We found that voters' stance toward the Cabinet had an influence on whether their opinion would be expressed or not to the question of Cabinet support. That is, people who did not support the Cabinet were more likely to stay as nonrespondents than those who did support. Our interpretation is that the critical attitudes or political dissatisfactions toward the Cabinet increased voters' likelihood of becoming false negatives.

The fifty-four years of almost consecutive conservative party rule in Japanese politics could be a factor for the existence of nonignorable false negatives. Since the end of the Second World War, conservative-leaning parties have led Japanese government except for two short periods, 1947-48 and 1993-94.It is thus conceivable that a sizable number of voters have been reluctant to disclose their opposition to the Cabinet, because there has been no realistic expectation that the opposing parties win the election; and they could have felt it unwise to express such a political opinion or have seen no merit in doing so. On the other hand, the spectacular defeat in the Second World War and the resultant American-imposed "Peace Constitution", left a indelible mark of pacifism in the Japanese psyche, in the sense that any actions of Japanese government, if it involves the operation by the Self Defense Forces outside Japan and in Japanese waters, have been seen as a thinly-disguised attempt to restore the Second World War-type militarism. Those who advocate ultra-conservatism have tended to be seen as something of a pariah because they tend to favor strong military build-up. It is therefore probable that a significant number of these people have been hesitant to reveal their ultra conservative ideologies. Whether these interpretations are appropriate should be examined in future research.

Next, we found that a voter whose ideology was conservative, who lived in smaller city and who supported the Cabinet was more likely to support a conservative-leaning party rather a liberal one. Though these factors cannot possibly explain all the major determinants to the political party support, it represents a set of fundamentally important variables that is theoretically sound and statistically significant. For the more systematical approach for profiling voters on the party support, for instance, path analysis could be applied. With such studies one might be able to differentiate direct and/or indirect factors that affect the party support.

Finally, non-partisans, people who answered "Do not support any political party" to the question of party support, was excluded in our research though the rate of non-partisans reached at 50-60% in our data. Since non-partisans cannot be classified according to the liberal-conservative ideology prior to the analysis, we need to use hierarchical response scale, by which the sample is classified first as "partisan" or "non-partisan", and then partisans are classified further according to the party they support. Even in that case, our method can be applied with slight modification.

Appendix A The Realignment of Japanese Political Parties

When we research Japanese party support, we must take into account the realignment of Japanese political parties that have been taking place in 1990's. It began when the Japan New Party was organized in May 1992. In June 1993, the breakup of the Liberal Democratic Party (LDP) launched the Sakigake (Harbinger) and the Shinsei (Renewal). In December 1994, the Democratic Socialist Party, the Clean Government Party, the Japan New Party, the Shinsei Party, and other small parties formed the New Frontier Party (NFP). The Democratic Party of Japan (DP) was established in September 1996 by the members of the Sakigake and the Social Democratic Party of Japan (SDP) to represent city-voters. Eventually, by such political party splits and new party births, there were five major political parties (LDP, NFP, DP, SDP and Japan Communist Party (JCP)), as of January 1997. Political differences and personal animosities led to the breakup of the NFP in December 1997, and paved the way to the creation of six small political parties; Liberal Party, New Party Peace, New Party Fraternity, Voice of the People, Dawn Club and Frontier Net, in January 1998. With the addition of the new and returning members, the LDP's strength in the 500-member Lower House stood at 259 as of January 1998. At the same month, the Good Governance Party, which consisted of members of the Voice of the People, the Taiyo Party and the From Five, arose to effectively counter the powerful LDP.

This realignment is thought to weaken the voters' party loyalty. Kabashima and Yamada (1995) found, in the four panel surveys from July 1993 to February 1995, that voters were changing their parties much like "one might change television channels".

Appendix B Coding of Variables

The answers are coded as follows:

Party-Support 1 (JCP: Japan Communist Party), 2 (SDP: Social Democratic Party of Japan), 3 (DP: Democratic Party of Japan), 4 (NFP: New Frontier Party), 5 (LDP: Liberal Democratic Party).

Ideology 1 (liberal), 2 (moderate), 3 (conservative), -9999 (don't know).

City-Size 1 (thirteen big cities: $\geq 1,000,000$), 2 (other cities: $1,000,000$–$50,000$), 3 (towns or villages: $< 50,000$).

Cabinet-Support 1 (I do support the Hashimoto's Cabinet), 2 (I do not support), -9999 (don't know).

We do not consider non-partisans in our analysis because the behavior of non-partisans is beyond the scope of this article, and also supporters to other small parties because the number of these supporters are very small and our measurement scale for party support is well defined. For "Ideology" we used three-point scale, by combining some of the ten categories used in the survey into the same category, but it is known that the number of categories in case of such an ordinal scale data does not influence the nature of analysis. The categories for other variables are exactly the one used in the survey.

Appendix C Parameter Estimation via EM

Let $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ be a set of n independent observations, where each $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$ is an observation vector of p covariates and y_i is a polytomous response with J categories. We specify the joint distribution of $(\mathbf{x}_i, y_i)$ by a conditional distribution of $y_i|\mathbf{x}_i$ and a marginal distribution of $\mathbf{x}_i$.

Without loss of generality, let us denote $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})^T$ as an indicator vector for the ith response, and suppose $\mathbf{y}_i|\mathbf{x}_i$ has a multinomial distribution with probabilities $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$. If the response categories can be ordered, however, it is more convenient to deal with the cumulative response vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{iJ})^T$ and cumulative probability vector $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{iJ})^T$, instead of $\mathbf{y}_i$ and $\boldsymbol{\pi}_i$ (see McCullagh and Nelder 1989). The probability density function for $\mathbf{z}_i|\mathbf{x}_i$ is given by

$$f(\mathbf{z}_i|\mathbf{x}_i; \boldsymbol{\gamma}_i) = \exp\left\{\sum_{j=1}^{J}(z_{ij} - z_{i,j-1})\log(\gamma_{ij} - \gamma_{i,j-1})\right\}, \qquad \text{(C.1)}$$

where $z_{i0} \equiv 0$ and $\gamma_{i0} \equiv 0$.

For the cumulative probabilities $\gamma_{ij} = \gamma_{ij}(\mathbf{x}_i) = \Pr\{Y_i \le j|\mathbf{x}_i\}$ we consider the following proportional odds model (see McCullagh and Nelder 1989):

$$\log\left(\frac{\gamma_{ij}(\mathbf{x}_i)}{1-\gamma_{ij}(\mathbf{x}_i)}\right) = \sum_{r=1}^{p^*} x^*_{ijr}\beta^*_r, \qquad i = 1, \ldots, n;\ j = 1, \ldots, J-1, \qquad \text{(C.2)}$$

where x^*_{ijr} is the element of an $n(J-1)\times p^*$ matrix $\mathbf{X}^*$, and β^*_r is the element of a vector $\boldsymbol{\beta}^* = (b_1,\ldots,b_{J-1},-\beta_1,\ldots,-\beta_p)^T$ with dimension $p^* = J-1+p$. Note that the negative sign in β's is a convention ensuring that large values of covariates lead to an increase of probability in the higher-number categories, and that b_j must satisfy $b_1 \le b_2 \le \ldots \le b_4$ to ensure that the probabilities are non-negative. The (i,j) row of $\mathbf{X}^*$ has components $(0,\ldots,1,\ldots,0,\mathbf{x}_i)$, with the unit value in position j. Consequently, the ith block of $J-1$ rows is $\mathbf{X}^*_i = [\mathbf{I}_{J-1} : \mathbf{1}_{J-1}\mathbf{x}_i]$, where $\mathbf{I}_{J-1}$ is an identity matrix, and $\mathbf{1}_{J-1}$ a $(J-1)\times 1$ vector with unit values.

Next we consider the distribution of covariates. Since all of the covariates are categorical, there exists $C = \prod_{r=1}^{p} C_r$ different possible combinations of levels of covariates or covariate patterns, where C_r is the number of categories for the rth covariate. We suppose that $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are random sample from the multinomial distribution with cell probabilities $\boldsymbol{\rho} = (\rho_1,\ldots,\rho_C)^T$. Then the probability density function for $\mathbf{x}_i$ is given by

$$f(\mathbf{x}_i;\boldsymbol{\rho}) = \prod_{l=1}^{C} \rho_l^{I_l(\mathbf{x}_i)}, \tag{C.3}$$

where $I_l(\mathbf{x}_i)$ is an indicator function that ith observation $\mathbf{x}_i$ belongs to the lth covariates pattern and ρ_l is the corresponding probability.

In addition to specifying the joint distribution of $(y_i,\mathbf{x}_i)$, we need to specify a distribution of the missing-data mechanism given data when the missing-data mechanism is nonignorable. Without loss of generality, denote $K = 2^{\#mis.\ var.}$ as the number of missingness patterns, where $\#mis.\ var.$ is the number of variables with missing value, and let $\mathbf{R}_i = (R_{i1},\ldots,R_{iK})^T$ as an indicator vector of missingness, where $R_{ik}(k=1,\ldots,K)$ takes value 1 if the ith observation vector $\mathbf{v}_i = (\mathbf{x}_i, y_i)$ belongs to the kth pattern of missingness. Assuming that $\mathbf{r}_i|\mathbf{v}_i$ has a multinomial distribution with probabilities $\boldsymbol{\psi}_i = (\psi_{i1}, \ldots, \psi_{iK})^T$, the probability density function for $\mathbf{r}_i|\mathbf{v}_i$ is given by

$$f(\mathbf{r}_i|\mathbf{v}_i; \boldsymbol{\psi}_i) = \prod_{k=1}^{K} \psi_{ik}{}^{r_{ik}}. \tag{C.4}$$

For $\psi_{ik} = \psi_{ik}(\mathbf{v}_i) = \Pr\{R_i = k|\mathbf{v}_i\}$, we consider the following multinomial logit model or baseline category logit model (see Agresti 1990). The model, when the last category (K) is the baseline, may be written as

$$\log\left(\frac{\psi_{ik}(\mathbf{v}_i)}{\psi_{iK}(\mathbf{v}_i)}\right) = \sum_{r=1}^{p^{**}} v^*_{ikr}\alpha^*_r, \qquad i = 1, \ldots, n;\ k = 1, \ldots, K-1, \tag{C.5}$$

where v^*_{ikr} is the element of a $(K-1)\times p^{**}$ matrix $\mathbf{V}^*_i = [\mathbf{I}_{K-1} : \mathbf{I}_{K-1}\otimes\mathbf{v}_i]$, α^*_r is the element of a vector $\boldsymbol{\alpha}^* = (a_1, \ldots, a_{K-1}, \boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_{K-1}^T)^T$ with dimension $p^{**} = (K-1) + (K-1)(p+1)$, and $\otimes$ denote Kronecker product.

Finally, let $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{*T}, \boldsymbol{\beta}^{*T}, \boldsymbol{\rho}^T)^T$, and suppose $\boldsymbol{\alpha}^*$, $\boldsymbol{\beta}^*$ and $\boldsymbol{\rho}$ are distinct sets of parameters. The complete data loglikelihood of $\boldsymbol{\theta}$ for n independent observations may be written as

$$\begin{aligned}
&\sum_{i=1}^{n} l(\boldsymbol{\theta}|\mathbf{x}_i, \mathbf{y}_i, \mathbf{r}_i) \\
&= \sum_{i=1}^{n}\left\{l_{\mathbf{z}_i|\mathbf{x}_i}(\boldsymbol{\beta}^*) + l_{\mathbf{x}_i}(\boldsymbol{\rho}) + l_{\mathbf{r}_i|\mathbf{v}_i}(\boldsymbol{\alpha}^*)\right\} \\
&= \sum_{i=1}^{n}\left\{\sum_{j=1}^{J}(z_{ij} - z_{i,j-1})\log(\gamma_{ij} - \gamma_{i,j-1}) + \sum_{l=1}^{C} I_l(\mathbf{x}_i)\log\rho_l + \sum_{k=1}^{K} r_{ik}\log\psi_{ik}\right\},
\end{aligned} \tag{C.6}$$

where $l_{\mathbf{z}_i|\mathbf{x}_i}(\boldsymbol{\beta}^*)$, $l_{\mathbf{x}_i}(\boldsymbol{\rho})$ and $l_{\mathbf{r}_i|\mathbf{v}_i}(\boldsymbol{\alpha}^*)$ are the contribution from the corresponding distribution specified by C.2, C.3 and C.5.

Based on the complete data loglikelihood in (C.6), the ML estimation of $\boldsymbol{\theta}$ with incomplete data can be obtained via the EM algorithm. We write $\mathbf{x}_i = (\mathbf{x}_{\mathbf{obs},i}, \mathbf{x}_{\mathbf{mis},i})$, where $\mathbf{x}_{\mathbf{obs},i}$ and $\mathbf{x}_{\mathbf{mis},i}$ denotes the observed and missing components of $\mathbf{x}_i$, respectively. Then, following the formulation of Ibrahim (1990), the E-step of the EM is defined as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[s]}) &\\ = \;& \sum_{i=1}^{n} \mathrm{E}_{\mathbf{X}_{\mathbf{mis},i}} \left[l(\boldsymbol{\theta}|\mathbf{x}_i, \mathbf{y}_i, \mathbf{r}_i) | \mathbf{x}_{\mathbf{obs},i}, \mathbf{y}_i, \mathbf{r}_i, \boldsymbol{\theta}^{[s]} \right] \\ = \;& \sum_{i=1}^{n} \sum_{\mathbf{X}_{\mathbf{mis},i}} w_i^{[s]} \cdot l(\boldsymbol{\theta}|\mathbf{x}_i, \mathbf{y}_i, \mathbf{r}_i) \\ = \;& \sum_{i=1}^{n} \sum_{\mathbf{X}_{\mathbf{mis},i}} w_i^{[s]} \cdot l_{\mathbf{Z}_i|\mathbf{X}_i}(\boldsymbol{\beta}^*) + \sum_{i=1}^{n} \sum_{\mathbf{X}_{\mathbf{mis},i}} w_i^{[s]} \cdot l_{\mathbf{X}_i}(\boldsymbol{\rho}) + \sum_{i=1}^{n} \sum_{\mathbf{X}_{\mathbf{mis},i}} w_i^{[s]} \cdot l_{\mathbf{r}_i|\mathbf{V}_i}(\boldsymbol{\alpha}^*) \end{aligned} \qquad \text{(C.7)}$$

where $w_i^{[s]} = \Pr(\mathbf{x}_{\mathbf{mis},i}|\mathbf{x}_{\mathbf{obs},i}, \mathbf{y}_i, \mathbf{r}_i, \boldsymbol{\theta}^{[s]})$ are the weights corresponding to the incomplete observations, and given by

$$w_i^{[s]} = \frac{\Pr(\mathbf{z}_i|\mathbf{x}_i, \boldsymbol{\beta}^{*[s]})\Pr(\mathbf{x}_i|\boldsymbol{\rho}^{[s]})\Pr(\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\alpha}^{*[s]})}{\sum_{\mathbf{X}_{\mathbf{mis},i}} \Pr(\mathbf{z}_i|\mathbf{x}_i, \boldsymbol{\beta}^{*[s]})\Pr(\mathbf{x}_i|\boldsymbol{\rho}^{[s]})\Pr(\mathbf{r}_i|\mathbf{v}_i, \boldsymbol{\alpha}^{*[s]})}. \qquad \text{(C.8)}$$

The M-step of EM is to estimate $\boldsymbol{\theta}$ which maximizes the loglikelihood in (C.7), which is equivalent to doing complete data maximum likelihood with each incomplete observation replaced by a set of weighted "filled-in" observations. Note that the expression in (C.7) takes the form of a weighted complete data loglikelihood based on $N = \sum_{i=1}^{n} n_i$ observations, where n_i is the number of possible distinct covariate patterns for the ith observation.

Therefore, at the sth iteration, the EM algorithm is carried out by the following steps:

E-step : Calculate the weight $w_i^{[s]}$ in (C.8) from the current estimate $\boldsymbol{\theta}^{[s]} = (\boldsymbol{\alpha}^{*[s]}, \boldsymbol{\beta}^{*[s]}, \boldsymbol{\rho}^{[s]})$.

M-step : Maximize the loglikelihood in (C.7), with respect to $\boldsymbol{\theta}$. Note that this maximization step involves three separate maximization, for $\boldsymbol{\beta}^*$, $\boldsymbol{\alpha}^*$ and $\boldsymbol{\rho}$. These ML estimates are regarded as the new parameter estimates $\boldsymbol{\theta}^{[s+1]}$.

In the M-step, the ML estimate of $\boldsymbol{\rho}$ are easily obtained since a closed form solution for $\boldsymbol{\rho}$ in (C.7) can be found. This ML estimate of $\boldsymbol{\rho}$ is regarded as the estimate $\boldsymbol{\rho}^{[s+1]}$. The likelihood equation for $\boldsymbol{\beta}^*$ and for $\boldsymbol{\alpha}^*$ in (C.7), however, is nonlinear, so the numerical optimization method is required to obtain the ML estimates. We use the Fisher's scoring method.

From the expected information and the score with respect to $\boldsymbol{\beta}^*$, the iterative estimate for $\boldsymbol{\beta}^*$ at the $(t+1)$st scoring iteration is obtained by

$$\boldsymbol{\beta}^{*[t+1]} = \boldsymbol{\beta}^{*[t]} + \left\{ \left(\sum_{i=1}^{N} \tilde{\mathbf{X}}_i^{*T} \mathbf{W}_i^{[s]} \boldsymbol{\Delta}_i \boldsymbol{\Gamma}_i^{-} \boldsymbol{\Delta}_i \tilde{\mathbf{X}}_i^{*} \right)^{-1} \sum_{i=1}^{N} \tilde{\mathbf{X}}_i^{*T} \mathbf{W}_i^{[s]} \boldsymbol{\Delta}_i \boldsymbol{\Gamma}_i^{-} (\mathbf{z}_i - \boldsymbol{\gamma}_i) \right\} \Bigg|_{\boldsymbol{\beta} = \boldsymbol{\beta}^{*[t}} \tag{C.9}$$

where $\tilde{\mathbf{X}}_i^*$ is a $(J-1) \times p^*$ matrix, generated by $\mathbf{X}_i^*$, in which any observation vector of covariates with missing components is replaced by a set of possible distinct covariate patterns; $\mathbf{W}_i^{[s]} = \text{diag}\{w_i^{[s]}\}$ is a $(J-1) \times (J-1)$ diagonal matrix of weight; $\boldsymbol{\Delta}_i = \text{diag}(\gamma_{i1}(1-\gamma_{i1}), \ldots, \gamma_{i(J-1)}(1-\gamma_{i(J-1)}))$ is a $(J-1) \times (J-1)$ diagonal matrix; $\boldsymbol{\Gamma}_i^{-}$ is a generalized inverse matrix of variance-covariance matrix (see McCullagh and Nelder 1989, p. 168). Note that at convergence, we take the convergence value of iterative estimate for $\boldsymbol{\beta}^*$ as $\boldsymbol{\beta}^{*[s+1]}$.

Similarly, we have the iterative estimate for $\boldsymbol{\alpha}^*$ as follows:

$$\boldsymbol{\alpha}^{*[t+1]} = \boldsymbol{\alpha}^{*[t]} + \left.\left\{\left(\sum_{i=1}^{N} \tilde{\mathbf{V}}_i^{*T} \mathbf{W}_i^{[s]} \boldsymbol{\Sigma}_i \tilde{\mathbf{V}}_i^*\right)^{-1} \sum_{i=1}^{N} \tilde{\mathbf{V}}_i^{*T} \mathbf{W}_i^{[s]} (\mathbf{r}_i - \boldsymbol{\psi}_i)\right\}\right|_{\boldsymbol{\alpha}^* = \boldsymbol{\alpha}^{*[t]}}, \tag{C.10}$$

where $\tilde{\mathbf{V}}_i^*$ is a $(K-1) \times p^{**}$ matrix, generated by $\mathbf{V}_i^*$, in which any observation vector of covariates with missing components is replaced by a set of possible distinct covariate patterns; $\mathbf{W}_i^{[s]} = \mathrm{diag}\{w_i^{[s]}\}$ is a $(K-1) \times (K-1)$ diagonal matrix of weight; $\boldsymbol{\Sigma}_i = \{\mathrm{diag}(\boldsymbol{\psi}_i) - \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T\}$ is a variance-covariance matrix for $\mathbf{r}_i$ with rank $(K-1)$. At convergence, we take the convergence value of iterative estimate for $\boldsymbol{\alpha}^*$ as $\boldsymbol{\alpha}^{*[s+1]}$.

For a prescribed convergence criterion $\epsilon > 0$, the EM algorithm terminates if $\left|\boldsymbol{\theta}^{[s+1]} - \boldsymbol{\theta}^{[s]}\right| < \epsilon$, and we take the ML estimate $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{[s]}$. Note that although our main interest is in the estimate of $\boldsymbol{\beta}^*$ and $\boldsymbol{\alpha}^*$, we also estimate $\boldsymbol{\rho}$, which is nuisance parameter, in the M-step to proceed EM algorithm.

The asymptotic variance-covariance matrix of ML estimate is obtained by inverting an observed information matrix based on observed data. We can use Louis' formula (Louis, 1982) to compute the observed information in terms of complete data quantities (see, e.g., Abe et al. 1998).

References

[1] Abe, Y., Kanazawa, Y. and Kabashima, I. (1998). Modeling Party Support of Japanese Voters in the Early 1996. *Behaviormetrika*, **25**, pp. 25-44.

[2] Agresti, A. (1990). *Categorical Data Analysis.* John Wiley & Sons, New York.

[3] Baker, S. G. and Laird, N. M. (1988). Regression Analysis for Categorical Variables with Outcome Subject to Nonignorable Nonresponse. *Journal of the American Statistical Association*, **83**, pp. 62-69.

[4] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society B*, **39**, pp. 1-38.

[5] Fitzmaurice, G. M., Laird, N. M. and Zahner, G. E. P. (1996). Multivariate Logistic Models for Incomplete Binary Responses. *Journal of the American Statistical Association*, **91**, pp. 99-108.

[6] Forster, J. J. and Smith, P. W. (1998). Model-based Inference for Categorical Survey Data subject to Non-ignorable Non-response. *Journal of the Royal Statistical Society B*, **60**, pp. 57-70.

[7] Gilljam, M. and Granberg, D. (1993). Should We Take Don't Know for an Answer?. *Public Opinion Quarterly*, **57**, pp. 348-357.

[8] Ibrahim, J. G. (1990). Incomplete Data in Generalized Linear Models. *Journal of the American Statistical Association*, **85**, pp. 765-769.

[9] Ibrahim, J. G., Lipsitz, S. R. and Chen, M. -H. (1999). Missing Covariates in Generalized Linear Models when the Missing Data Mechanism is Non-ignorable. *Journal of the Royal Statistical Society B*, **61**, pp. 173-190.

[10] Inoguchi, T. (1983). *Gendai Nihon Seiji Keizai no Kouzu: Seifu to Shijo (Contemporary Japanese Political Economy: Government and Market).* Toyo Keizai Shinposha, Tokyo (in Japanese).

[11] Kabashima, I. (1998). *Seiken Koutai to Yukensha no Taido Henyo (Power Changes and Attitude Changes of Eligible Voters).* Bokutakusha, Tokyo (in Japanese).

[12] Kabashima, I. and Ishio, Y. (1998). The Instability of Party Identification among Eligible Japanese Voters. *Party Politics*, **4**, pp. 151-176.

[13] Kabashima, I. and Yamada, M. (1995). Seito Shiji no Henndou to Mutouhasou no Keisei (The Change in Party Support and the Formation of Non-Partisan Voters). *Senkyo*, **48**, pp. 1-17 (in Japanese).

[14] Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* John Wiley & Sons, New York.

[15] Louis, T. A. (1982). Finding the Observed Information When Using the EM Algorithm. *Journal of the Royal Statistical Society B*, **44**, pp. 226-233.

[16] Manski, C. F. and McFadden, D. (1983). *Structural Analysis of Discrete Data with Econometric Applications.* The MIT Press, London.

[17] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London.

[18] Miyake, I. (1989). *Tohyo Kodo (Electoral Behavior).* University of Tokyo Press, Tokyo (in Japanese).

[19] Miyake, I. (1995). *Nihon no Seiji to Senkyo (Japanese Electoral Politics in Disarray).* University of Tokyo Press, Tokyo (in Japanese).

[20] Miyake, I. (1998). *Seitou Shiji no Kouzou (Structure in Political Party Support).* Bokutakusha, Tokyo (in Japanese).

[21] Richardson, B. M. (1988). Constituency Candidates Versus Parties in Japanese Voting Behavior. *American Political Science Review*, **82**, pp. 695-718.

[22] Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, **63**, pp. 581-592.

[23] Watanuki, J., Miyake, I., Inoguchi, T. and Kabashima, I. (1986). *Nihon-jin no Senkyo Kodo (Japanese Electoral Behavior).* University of Tokyo Press, Tokyo (in Japanese).