第一年期中報告

報告人：劉正山

102.5.21

本研究計畫為三年期的應用/技術型計畫，目標在於結合理論及調查的實務，探求隱藏黨派傾向選民之真實意向，增加學術界對於隱藏黨派傾向選民之真實意向的洞察能力。我國有相當高比例的選民會在調查中隱藏其真實的政黨傾向。這樣的選民不見得只是項目無反應（item-nonresponse）而已，他們往往宣稱自己是中間或中立選民。從理論上來看，學者有足夠的理由相信，多數選民有其真實的政黨傾向，只是現有的測量方式不易洞察。也因為這個長久以來理論與實務的落差，造成目前要瞭解政黨支持分佈，得依靠有實際開票結果作對照的選舉調查；非選舉期間所進行的調查之預測力十分有限。本年度的主題文獻仍在整理中，但資料收集的工作已經開始。本年度除了文獻之外，亦啟動了第二年的部份工作，先精進研究方法，將 multiple imputation 應用在今年度收集的電話調查資料，來增強對中間選民的政黨傾向的預測能力。

以下為後學依第一年研究所收集到的文獻及電話訪問資料所完成的國際學術會議論文。此成果已於 5 月 19 日發表於美國民意調查研究學會（AAPOR）在美國波士頓舉辦的年會。以下為會議論文全文，此文將依評論人建議修正並潤飾後投稿國際期刊。

Adjust Survey Response Distributions Using Multiple Imputation: A Simulation with External Validation

This paper is to be presented in

<u>AAPOR 2013 Annual Conference, Boston, USA, April 16-19, 2013</u>

This is an early draft; please inform the authors for citation.

Frank C.S. Liu

Associate Professor

Institute of Political Science

National Sun Yat-Sen University, Taiwan

E-mail: csliu@mail.nsysu.edu.tw

Yu-Sung Su

Associate Professor,

Department of Political Science

Tsinghua University, China

E-mail: yusung@gmail.com

Abstract

One commonly acknowledged challenges in polls or surveys is item non-response, i.e., a significant proportion of respondents conceal their preferences about particular questions. This paper presents how multiple imputation (MI) techniques are applied to the reconstruction of vote choice distribution in telephone survey samples. Given previous studies about using this method in adjusting vote share information drawn from pre-election survey/poll data, this paper gives more attention to external validity of this method. Using a telephone survey dataset collected in Taiwan early 2013, the authors take two steps to study the utilities of this method. First, they randomly take out a proportion of values in a variable with few or no missing values to examine the assumption of missingness patterns. Then the follow up survey responses are compared against the "guesses" generated by MI. This paper reports and concludes the utility of applying MI to point-estimation adjustment.

Keywords: multiple imputation, item-non-response, missing values, external validation

Scholars have become aware today that it is not proper to make electoral forecasts based on information simply drawn from raw survey and poll data. When it comes to controverisal or sensitive survey questions, such as partisanship and stance on moral issues, respondents who hesitate to disclose their preferences and attitudes usually create a lost of data so called item-non-response problem. Calculating proportions based on raw data and omitting the non-response data result in biased proportion of interested variables (Bernhagen & Marsh, 2007). This paper is an attempt to apply multiple imputation (MI) to improve electoral forecast. The author echoes the findings of other similar studies and argues that MI is a cost-efficient and methodological sound approach for better use of raw survey and poll data (e.g.,Barzi, 2004).[1]

Multiple Imputation for Electoral Studies with Missing Values

Multiple imputation (MI) refers to a technique by which researchers replace missing or deficient values with a number of alternative values representing a distribution of possibilities (Paul, Mason, McCaffrey, & Fox, 2008; Rubin, 2004).[2] MI has been widely adopted for regression analysis. Researchers draw auxiliary variables, those related to a target variable of interest, from theories and the literature, and then use MI algorithms or software to generate “guessed” values for each missing value based on the distributions of selected auxiliary variables. This procedure will create a number of supplemental data sets in which all missing values are filled. To obtain unbiased and robust regression coefficients, the researcher first runs models using every data set generated

1 For a summary of other approaches to deal with this item-non-response problem, see Liu (2010) and Florez-Lopez (2010). Allison (2001) holds the conventional view that when it comes to linear regression, list-wise deletion is the least problematic and safer method to deal with missing data. As this paper focuses on advancing the accuracy of the proportion of a dependent variable, such as voter turnout, vote choice, etc., this paper is not devoted into the debate about the choice of approach.

2 MI is a method commonly used to deal with missing data problem, including item-nonresponse (nonresponse to some, but not all, survey questions) and unit-nonresponse (nonresponse to all survey questions). A common and still useful alternative is list-wise deletion of observations due to both item-nonresponse and unit-nonresponse in the regression analysis. However, because a significant number of observations are excluded from analysis, this method may yield biased parameter estimates. While the default procedure of most statistical packages excludes the observations with missing values, list-wise deletion has been identified as a problem for most electoral studies (Gelman, King, & Liu, 1998). This concern regarding biased estimates can be minimized if the loss of cases due to missing data is less than about 5%, and if pretest variables can reasonably be included in the models as covariates (Graham, 2009).

from the above procedure. This is followed by averaging the coefficients and standard errors across the models. (King, Honaker, Joseph, & Scheve, 2001; Snijders & Bosker, 2011; Stuart, Azur, Frangakis, & Leaf, 2009).[3]

Electoral scholars have started paying attention to this MI approach and applying it to electoral forecast. Bernaards et al. (2003) compares descriptive statistics of data drawn from the MI procedure to determine if two or more methods generate similar results. Bernhagen and Marsh (2007) adopt this approach by treating non-voters and non-party identifiers as missing and recreate "hypothetical (100% turnout)" votes for individual elections and for individual parties. Although their works uses the conventional way of MI to study relationship between explanatory variables and a chosen response variable, these works imply that using MI to study dependent variable is one way to go. That is, scholars can pay attention more to advancing the accuracy of descriptive statistics of dependent variable than explaining the variance of dependent variable. Although this novel focus is absent from Rubin (1987), there is little methodological reason to object this scholarly action to impute the dependent variable. In effect, studying vote choices of non-voters have been proposed at the time when MI was introduced to the discipline (King et al., 2001; Snijders & Bosker, 2011). As G. David Garson confirms this perspective in his course website, "for purposes of univariate analysis (e.g., understanding the frequency distribution of how subjects respond to an opinion item) imputation can reduce bias and often is used for this purpose if data are missing at random."[4]

3 While some scholars may think this technique is unrealistic, or have concerns about "making up" data, we need to acknowledge that "complete-case analyses require [even] stronger assumptions than does imputation" (Stuart, Azur, Frangakis, & Leaf, 2009, p. 1134).

4 See, http://faculty.chass.ncsu.edu/garson/PA765/missing.htm

Software selection for the MI analysis

Two major algorithms are commonly used in the existing MI software packages. One is joint MI and the other is conditional MI. R as free (as freedom) software for statistical computation and conduct MI inclucing tools for both joint MI and conditional MI.[5] The packages conducting joint MI is Amelia (II)[6]; for coniditional MI one can choose mi and MICE[7].

Joint MI takes much less time than conditional MI to complete calculation. As King, Honaker, Joseph, and Scheve (2001) argue, EM is a faster and less complex alternative to imputation posterior (IP). Concerned about that EM algorithm ignores estimation of uncertainty, they propose EMis (EM with importance re-sampling) to solve the uncertainty problem in EM. This implies that Amelia will be more time efficient than tools based on chain equations like MICE and mi when handling computation. To them, if using IP the algorithms will frequently draw an estimated mean and variance from the disputed data sets created from entire multivariate models of observed data posterior. In order to obtain an exact result as expected, it may take a substantial amount of time drawing infinitely before convergence occurs.

Conditional MI weights more on assumptions held in algorithms than on calcuation speed. Joint MI assumes that the data follow a multivariate normal distribution. The joint MI uses variable transformation before the imputation to make the data distribute normally and then uses

---

5 Besides R, commercial packages such as SAS, SPSS and STATA also supports the procedure of MI. For example, an illustration of using STATA for MI can be found: http://www.stata.com/meeting/spain12/abstracts/materials/Escobar_Jaime.pdf

6 *Amelia II* is a cross-operation system package designed to process EMis (Expectation Maximization with importance re-sampling), one of the suggested algorithms using Markov chain Monte Carlo (MCMC) methods to calculate imputed values (Honaker, King, & Blackwell, 2011). Expectation Maximization (EM) is a method of joint MI and it is also called "joined modeling" (JM), "specifying a multivariate distribution for the missing data, and drawing imputation from their conditional distributions by MCMC techniques" (Buuren & Groothuis-Oudshoorn, 2011, pp. 1–2). *Amelia II* is characterized by its speed using EMis and a handy graphical user interface (GUI), allowing the researcher to manage types of variables by simply designates them as nominal or ordinal variables. After variables are specified, it automatically transforms nominal variables into dummy variables and regards them as categorical variables during the imputation process. For researchers following the three-step procedure of conducting MI, Zelig, another package compatible with R, is suggested for the combination stage (Imai, King, & Lau, 2004). Since hypothesis testing is not the goal of the present study, the analysis below will concentrate on using Amelia II for the first two stages of MI.

7 MICE is a package similar to mi in terms of the adoption of chain equations and proceeding one-variable-at-a-time. What distinguishes mi from MICE is that mi adds algorithms to do with semi-continuous data (such as age of survey respondents) and adds Bayesian models to assist researchers to construct more stable estimates of imputation models (Su, Gelman, Hill, & Yajima, 2011).

transformation after the imputation to recover the original format of the data. That joint distribution is normal seems a naive assumption, because the data might contain binary, ordinal, (unordered) categorical and other special types of variables. In effect, all of which are not of normal distribution. A joint distribution may hardly be a multivariate normal distribution. As Kropko, Goodrich, Gelman and Hill (2013) show, joint MI performs less accurately when a dataset contains many non-normal variables. If this is the case, they propose to use conditional MI.

Conditional MI relaxes the assumption of multivariate normality of the data. The R package mi takes advantage of existing regression models to handle various kinds of variables types: using a logistical regression model to predict a binary outcome, an ordered logit regression model to predict an ordinal outcome, and a multinomial logit regression model to predict an unordered categorical outcome (Su, Gelman, Hill, & Yajima, 2011).

He and Raghunathan (2009) conduct a series of experiment and compare the performance of MI using sequential regression or chain equations. They find that all methods using chain equations perform well for estimating the marginal mean and proportion, as well as regression coefficients even when the error distribution is non-normal. However, they warn that the limit of this method is that the MI results can be very biased when error distributions are strongly heavy tailed, i.e., when data include extreme values. Therefore, we think it proper to use mi as a tool to avoid extreme or impossible values. Therefore, mi that relaxes this joint normal assumption looks more reasonable for such case.

Data

The dataset used for this project was collected from January 23 to February 4, 2013 by a telephone survey center of a research university in Taiwan. The population was eligible voters above 20. Sampling was based on the telephone book published by Chung-Hua Telecom in 2010. The computer assisted telephone interview (CATI) system removes last two digits of all telephone numbers and replaces a full set of 100 double-digit figures from 00 to 99. Specific numbers were then randomly selected from the database by computers. The survey has 1,078 completed interviews. The response rate is 21.56% following American Association of Public Opinion Research (AAPOR) formula 3. Based on population information of 2012, raking weights were applied to the sample and it was ensured that the distributions of sample age, gender, and education level do not substantially differ from the population.

The target variable is political camp to which the respondent belongs, i.e., pro-KMT or pro-DPP. As Table 1 shows, the missing rate of this variable is 61%; 658 out of 1,078 respondents conceal their partisanship in the telephone survey. Partisanship, measured by the question "Which political party do you support for?" has been a "sensitive" question in Taiwan. It is common in surveys that one-third (in face-to-face surveys) to half (in telephone survey) of the samples refuse to tell interviewers their party orientation. Therefore, 61% is even higher than the common cases, indicating a phenomenon of this survey that a increase number of citizens conceal their partisanship and a fact that political party identification has been the most sensitive question in telephone surveys in Taiwan.

The auxiliary variables chosen for MI are listed in Table 1. These variables are chosen based on the empirically supported evidence that a Taiwanese national identification is strongly related to their partisanship .

[Table 1 is about here]

Research Design

The dataset is characterized by a battery of categorical variables. This requires us to adopt conditional MI algorithm to check its missingness patterns and to conduct MI. The targeted variable camp is one that has the highest level of missingness in the survey (See Table 1): 61.1% of respondents fail to provide a valid answer.

Multiple imputation is not a magic algorithm to recover the missing values. There are several assumptions that needed to comply to ensure the quality of the imputation. Many practitioners use imputation software to impute the missing data without checking the validity of the assumptions. Firstly, the data should be at least missing at random (MAR). MAR is a commonly held assumption about the missing patterns of the original data, meaning that the chosen missingness indicators are independent of the unobserved data. In other words, conditional on the other observed variables, the missingness mechanism does not depend on the unobserved data. As Snijders and Bosker (2011) suggests, it is recommended "to collect auxiliary data that are predictive of missingness indicators and of the values of unobserved data points. Including such auxiliary data can push the design in the direction of MAR" (p.150). The other two assumptions about the pattern of missingness are missingness completely at random (MCAR) and missingness not at random (MNAR). MCAR means that the missingness indicators are independent of the complete data; MNAR is a situation where missingness is not at random and will always depend on untestable assumptions. MNAR and will be more complicated and leave open more questions than the MAR case.

Often the cases, researchers ignore the missing data and use complete case analysis assuming that their data is MCAR. Clearly, this is a rather strong assumption. MAR is somewhat weak assumption about the missingness mechanism. However, for those who use implement joint MI or conditional MI commit equal crime if they do not check whether or not their data is at MAR.

Secondly, the conditional model should be appropriate specified. Since the conditionality is a major component of the imputation procedure, a conditional model that is inappropriate might lessen the accuracy of the prediction; hence the imputation. Fortunately, this assumption is less of a problem if our conditional models contains many variables (the ignobility can be reached) and are valid (each conditional family contains the true probability distribution) (Liu, Gelman, Hill, & Su, 2012). In the present study we have selected 18 variables from the dataset. Since most of them are categorical, we believe that a linear combination of these variables as predictors with appropriate regression models should ensure that the second assumption is complied.

Henceforth, we are going to evaluate our imputation result by only checking the first assumption: the missingness pattern of the data. Choosing a baseline for comparison is challenging here because there is no truth. The truth can only be done by a follow up survey to ask those who failed to provide a valid answer in the first round.

[Figure 1 is about here]

We choose the imputed datasets as the baseline for comparison. Then we create three copies and randomly remove values from the completely imputed data according to the original missing rate of the data. As illustrated by Figure 1, we remove 61% of the values of camp variable, 2.7% of the values of age variable and so on so forth (see Table 1 for the missing rate of each variable). To create MAR, we model the missingness of each variable conditional on a linear combination of the rest of variables with logistic regressions. Then we use the predictive missingness to create missing values on the three imputed data. We impute these MCAR and MAR datasets using conditional MI again. After that, we will compare the results of these imputation with the original imputation that imputes the original data.

Given the check of the missing pattern of the dataset, we then will conduct external validation of MI by the following steps: compare respondents' answers with MI guesses and figure out how well the MI prediction works; next, understand why prediction performs not so well, if this is the case.

First, given the data set as detailed above, we impute "camp," denoted by 1 (the Blue camp) and 2 (the Green camp), by using the selected auxiliary variables that are highly correlated with camp. At this stage we will have a probability value for each respondent, indicating their propensity to be 1 or 2. The imputed probability score that is close to 1 indicates supporting for the green camp; one that is close to 0 indicates supporting for the green camp.

The second stage is to contrast respondents' answers against our guesses. Limited by time and funding, we found it is difficult to reach all of the 1,078 respondents and check their answers to the question that they rejected to answer in the first time. In fact, we could only contact those who allowed us to re-contact (only 484 out of 1,078 agreed). Therefore, we chose an alternative strategy: reaching out for those who did not give answers to the political party question, the 61% of the sample, and took a risk of contacting some of them without their consent. We called out for the 658 respondents from April 13 to April 15, 2013 and, not to our surprise, 143 completed this round of survey. Given the second data set, we contrast our guesses to their answers to the forced choice question (support for the Blue or the pan green camp). The questionnaire for the second round of telephone interviews is shorter than the first one, composed of only few questions, including demographics, two questions of checking the consistency of their answers—whether ever going to mainland China in the past two years and the frequency of watching political news, and the political camp questions that they avoided to answer in the first run of telephone survey.

The third stage is checking out how well our method is to predict the preferences of the group of respondents whose answers are least likely to predict correctly. As we focus on those whose answers are predicted incorrectly. We then make the third round of calls between April 20 to May 6, 2013, to those who still accepted our request for recontact (45 out of 143). We then select 5 respondents from this pool of 45 people for further face-to-face interviews.

Results and Findings

Checking the Missingness Pattern of the Dataset

Checking the missingness pattern of the dataset is an important but commonly skipped step for MI researchers. We first check whether or not our data is MCAR or at least MAR.

<u>The Result of MCAR</u>

[Figure 2 is about here]

Figure 2 displays the missingness pattern plot of original data and 3 copies of simulated MCAR mechanism on the imputed data. It utilizes a hierarchical cluster analysis algorithm (Murtagh, 1985) to group the data according to the missingness mechanism. Visually speaking, the missingness pattern plots of simulated datasets look dissimilar with that of the original data. Compared to the original data, the first and second simulated MCAR data have more missing on the right end of the data; and the third simulated MCAR data have some missing on the center of the data.

[Table 2 is about here]

Next, we proceed to formally compare these three dataset with the previous imputed one by imputing those three dataset using conditional MI. If they are similar in some basic statistics, we can still assume that the data are MCAR. Table 2 shows the summary statistics of camp variable under different scenarios. The summary of the original data is also presented here as a reference. Again, we don't know the true response of the missing data, so the summary of the original data serves only as a reference, not a truth.

Overall, Table 2 shows that the imputed MCAR data predict more people belong to the blue camp than the imputed original data does. The difference between the imputed MCAR data and the original data is even larger. At this step, we have some clues that perhaps our data is not MCAR.

[Figure 3 is about here]

Figure 3 further plots the bivariate tabulation of the camp variable against the other variables of one MCAR imputation. The purple line represents the "true" value of imputed original data.

The three red lines represents three difference MCAR imputations results. The dark green lines represents the result of the completed case of the original data.[8] The blue lines represents the available case of the original data.[9] The size of the circle indicates the number of cases of the plotted variables. Henceforth, the blue circles should always be larger than the green ones.

Overall, the good news is that all the purple lines lies in the middle of the three red lines, meanings that all MCAR imputation are reasonable computed: after pooling, the red lines are close to the purple lines: he imputed original data, which we set to be the baseline here. However, there are some ups and downs of the blue and dark green lines of the middle categories of v6, v17, v18, v22, v27, v28 and v29. This indicates that after imputation, the imputed values are quite dissimilar to the original data in the completed and available cases. However, the small circles in these categories indicate that the number of observation is small here. Therefore it ought to bear great uncertainty. However, this is still a warning sign to our imputation model.

In short, the MCAR imputations result in somewhat dissimilar camp variable to the original imputed ones. Based on Figure 3, it seems that the differences may exist in those middle categories. Those respondents are those who have relative neutral answers to several survey questions.

<u>The Result of MAR</u>

Next, we relax the assumption of MCAR by checking whether or not the data is MAR.

[Figure 4 is about here]

Figure 4 demonstrates the missingess pattern of the original data and the three copies of simulated MAR data. These three copies of data are still dissimilar to the original one in that

8 The completed case here means that there is no missing in any of the variable in the dataset.
9 The available case here means that there is no missing of the two plotted variables in the dataset.

patterns of missing on the right and left ends are not quite similar to the original one. However, the middle range of the data has no missing except for the camp variable. This does look similar to the original data. At least, the MAR data look more similar to the original one than the MCAR data do.

[Table 3 is about here]

Table 3 shows the summary statistics of the camp variable under different scenarios. Unlike the ones in the MCAR setting, the imputation of MAR data predicts more green than the imputation of original data does. Still, there are more people that belong to the blue camp than those that belong to the green camp. The result is somewhat close to what the original data shows. However, our major comparison is between the imputation of the MAR data and the imputation of the original data. Henceforth, it seems that our imputation of the MAR data does not perform well enough.

[Figure 5 is about here]

Figure 5 shows detailed bivariate tabulation of the camp variable against others. Similar to what we observed in Figure 3, the purple lines lies in the middle of three red lines in most panels. Our imputation of MAR does relative reasonable job in terms of not deviating away from the baseline imputation. Moreover, the phenomenon of the ups and downs of the blue and dark green lines in the middle categories of other variables is still significant. This is also due to the small size of observation in these categories.

Summary of Assumption Check

Our checking of missingness pattern of the data yield tentative results. To some extent, the imputation based on the MCAR or MAR data do perform reasonably. The imputed estimates are close to the ones of the imputation of the original data. However, the overall statistics of the camp variable do not look good. In the MCAR setting, we overestimate the number of blue camp people. In the MAR setting, we underestimate the number of blue camp people. Figure 3 and Figure 5 give us reasons why there is such a deviation. In some middle categories of the covariates, the imputation perform not so well because the number of observation is small in these categories. Henceforth, the deviation in these categories might affect the average predictions. In the next section, we are going to demonstrate the follow up responses of those people who are missing at the camp variable.

External Validation

The first wave of telephone survey provides raw data of the target variable camp and auxiliary variables. In a follow up telephone interview, we forced the 658 respondents who did not answer their camp preferences to select between the two political camps by providing only two options. While the majority of respondents still refused to answer this question, 143 respondents did give answers, including 74 choosing the Blue camp and 69 choosing the Green camp.

We first set 0.5 as the naive cutting point where imputed probabilities lower 0.5 set to 1 (Blue camp) and those higher than 0.5 set to 2 (Green camp). We find that 100 out of 143 respondents whose political camp choice were correctly predicted (the success rate **70%**). As listed in Appendix, for the 43 respondents that we predict incorrectly 19 have values falling between .45 and .55. We invited and interviewed 5 of them to explore the causes of this ambivalence.[10]

[Table 4 is about here]

10 It is very difficult to contact this list of defensive respondents. We tried called out for 20 of them and felt fortunate to have 5 chose to show up.

Three patterns can be drawn from the talks of the five “mysterious” respondents whose MI values are close to 0.5 but gave political camp choice inconsistent with our prediction. Note that we did not talk about our MI predictions until the end of the interviews. The profiles of the five respondents are summarized in Table 4.

The first pattern is that the five respondents are very aware of politics and do not avoid discussing politics with us. They can talk about politics for more than 30 minutes in our personal interview and some can even criticize our question wording (for being too narrow in the definition about Chinese, for example). We observed that they are active in absorbing political information from TV, newspaper, and online news sources. Therefore they aware of controversial issues and hav influenced by impression about political issues obtained from the news media.

The second pattern is that they do not want to claim a supporter for a political party without criticism. They chose an opposite political camp (contrasted to our prediction) in the second telephone interview primarily because they felt that it was a forced answer and therefore felt forced to respond with a short-term evaluation about politics. Because in April 2013 Taiwanese people are concerned about the reform on a number of domestic policies, such as nuclear power plans, retirement plans, health care reforms, etc., therefore, we found that the four potential KMT supporters (ID 905, 206, 384, and 286) chose DPP because of their concerns about KMT's leadership in domestic politics and policy reform.

Feeling cross-pressured is a third reason for hiding partisan orientation. Respondents ID 140, 384 and ID 286 are representative of classic cases of being living in heterogeneous political communication networks. They can choose either political camp but what they chose in the forced-to-answer question is based on their short-term evaluation of politics. We found that pro-KMP camp supporters seem to be influenced by evaluation about policy and concurrent political issues, while pro-DPP camp supporters are affected by nationalism.

Conclusion and Discussion

One commonly acknowledged challenges in polls or surveys is item non-response, i.e., a significant proportion of respondents conceal their preferences about particular questions. In this paper we take two steps to study the external validity of applying the multiple imputation method to the study of "independent" voters who conceal their answers about their partisanship. The overall prediction using selected auxiliary variables perform well. From those who gave their answer in the re-contact telephone interview we find that MI scores reflect their partisan orientation, including the level of their ambivalence. In our follow-up face-to-face interviews of 5 respondents, we found that the inconsistency of their answers can be explained.

Our study shows that MI has a great potential to solve item non-response problems in telephone surveys. Even that the missing rate in the target question "political camp" (pro-KMT or pro-DPP) is over 60%, MI helps the reconstruction of the distribution and we suspect that the probability of using MI to make correct guesses about respondents' partisan orientation can be higher than 70%. The findings drawn from only five individuals may not represent the whole body of respondents who are unwilling to be bothered with political questions. But we find that information revealed from those individuals are helpful for us to ensure that the scores of MI procedures were not misleading at all; instead, we see that the scores are quite consistent and representative of their status of being caught in the middle.

The panel-like re-contact shows us that so-called independent voters in Taiwan are likely to be those partisans that have difficulty to make a out-of-head, quick choice in telephone surveys. They will fail to give an answer correspond to their overall evaluation about the parties but give a quick answer that reflects their short-term evaluation of politicians, policy issues, or just emotion. Independent voters who really are ambivalent about making a choice about political camp are not apolitical or indifferent of politics. From the in-depth interview we found that they can be partisans hiding behind the excuses of "choosing candidates, not parties."

We hence have a good reason to continue to worry about using telephone surveys in probing citizens' partisan orientation and/or other controversial issues. Respondents, particularly ambivalent ones, are likely to dodge such questions or give an answer that do not correspond to or consistent with their belief systems. We would suggest that survey institutes avoid asking for quick answers but (1) encourage respondents to give up answering those they could not answer and (2) encourage them to answer sincerely other auxiliary questions that seem less sensitive to them. Using MI researchers will be able to use such sincerely answers to reconstruct the distribution of the target variable of high missing values.

We do not know how successful it is that the MI method we used for this study correctly predict 70% to the missing values. Although we suspect that the figure could in effect higher (as we use naive cutting point of 0.5 instead of a more sophisticated one such as avoid counting those whose score falling between 0.4 and 0.6), we see room to do better in future studies.

First, we need to start thinking collecting more questions and dimensions as auxiliary variables. In the present study more than 25 questions asked were found having statistically significant relationship with choice of political camp. We chose 18 of them including demographics. These questions are mostly related to the concept of state identification and national identification, the two dimensions that have been found empirically related to one's party identification in Taiwan. We suggest that future studies continue to explore other dimensions and concepts and test how other measurements and questions contribute to the success rate of prediction.

Second, we did not (and was not able to) interview all of the 1,078 respondents and make it a panel. This prohibits us from confirming the exact external validation of our guess. In future studies using a panel that is composed of volunteer respondents will give a more solid ground for testing external validity of this method.

Third, we found in in-depth interviews that those respondents of 0.5 (or close to 0.5) are likely to be influenced by their feelings and emotion about controversial issues at the time of survey. Therefore, we suggest researchers consider add some questions of evaluation of policy and performance as auxiliary variables.

Fourth, using questions that force respondent to choose one answer may not the best strategy for digging out “true” answers from defensive respondents. Alternative methods or more skillful question wording are needed. Before such method is found to replace questions of direct asking for party or camp choice, we propose in this paper that MI can be seen as the best indirect approach.

Table 1. Summary of Variables

Variables	Question Wording	Distribution	Missing (%)
camp (v33)	This is a binomial variable. This variable is derived from the re-coding of the following question: Do you support for any political party? 1 = KMT; 2 = DPP; 3 = New Party; 4 = PFP; 5 = TSU; 6 = TIP; 7 = pro-KMT; 8 = pro-DPP; 10 = green parties; **NA = other parties, don't know, forget, or refuse to answer.**	1 = Pro-KMT (Blue) camp: 215 2 = Pro-DPP (Green) camp: 204	**658 (61.04)**
v6	Do you agree that we can influence the government with our votes?	Strongly disagree: 135; Disagree: 254; Neutral: 19; Agree: 354; Strongly agree: 297	19 (1.76)
v7	Do you agree that we have little influence on what the government plans to do?	Strongly disagree: 107; Disagree: 276; Neutral: 27; Agree: 282; Strongly agree: 361	25 (2.32)
v9	Do you agree that we should use Taiwanese as the major language in Taiwan?	Strongly disagree: 201; Disagree: 417; Neutral: 89; Agree: 186; Strongly agree: 163	22 (2.60)
v10	Do you agree that Taiwanese children perform better than those in mainland China?	Strongly disagree: 185; Disagree: 478; Neutral: 58; Agree: 164; Strongly agree: 118	75 (6.96)
v17	Do you agree that those identifying with Taiwan can be titled Taiwanese?	Strongly disagree: 138; Disagree: 332; Neutral: 17; Agree: 287; Strongly agree: 270	34 (3.15)
v18	Do you agree that Chinese from mainland China have more money than sense?	Strongly disagree: 117; Disagree: 323; Neutral: 36; Agree: 312; Strongly agree: 230	60 (5.57)

v20	Do you agree that those people should not be called Taiwanese if they don't know Matsu (name of a sea goddess widely worshipped on the SE China coast and in SE Asia)?	Strongly disagree: 322; Disagree: 530; Neutral: 14; Agree: 119; Strongly agree: 61	32 (2.97)
v22	Do you agree that our government should give more restrict policy on mainland China tourists?	Strongly disagree: 169; Disagree: 358; Neutral: 19; Agree: 246; Strongly agree: 258	28 (3.90)
v27	Some call them selves Taiwanese, some Chinese, and some Both, what about you?	Taiwanese: 562; Chinese: 46; Both: 440	30 (2.78)
v28	Do you agree that “Taiwan” is the formal name of our country?	Strongly disagree: 117; Disagree: 195; Neutral: 20; Agree: 300 Strongly agree: 404	42 (3.90)
v29	Do you agree that Taiwanese people and those in mainland China belongs to a same nation?	Strongly disagree: 82; Disagree: 144; Neutral: 15; Agree: 491; Strongly agree: 322	24 (2.23)
v32	Do you agree that we should seek for unification with mainland China if it becomes a democracy?	Strongly yes: 270; Yes: 229; No: 326; Strongly no: 147	106 (9.83)
v37	Do you agree that the two sides of Taiwan Straight will eventually becomes one country?	Strongly disagree: 378; Disagree: 283; Neutral: 22; Agree: 238; Strongly agree: 70	94 (8.72)
v38	Do you ever go to mainland China in the past two years?	1 = yes; 2 = no.	4 (0.37)
age	Which year were you born? (re-coded to real age)	a continuous variable. Mean=46.5 years; SD=14.1 years	29 (2.69)
edu	What is your education level?	1 = Junior high school and below: 147; 2 = High school and vocational school: 334; 3 = College: 491; 4 = Graduate, plus: 99	7 (0.65)
sex	(coded by interviewer)	1 = male; 2 = female	0 (0.00)

Source: this study; N = 1,078

Note:

1. “missing” includes “refuse to answer,” “don’t know,” and “skip”
2. all of the chosen auxiliary variables are correlated with the target variable “camp” at the 0.001 significance level.

Figure 1: The Flowcharts of Checking the Missing Pattern of the Dataset

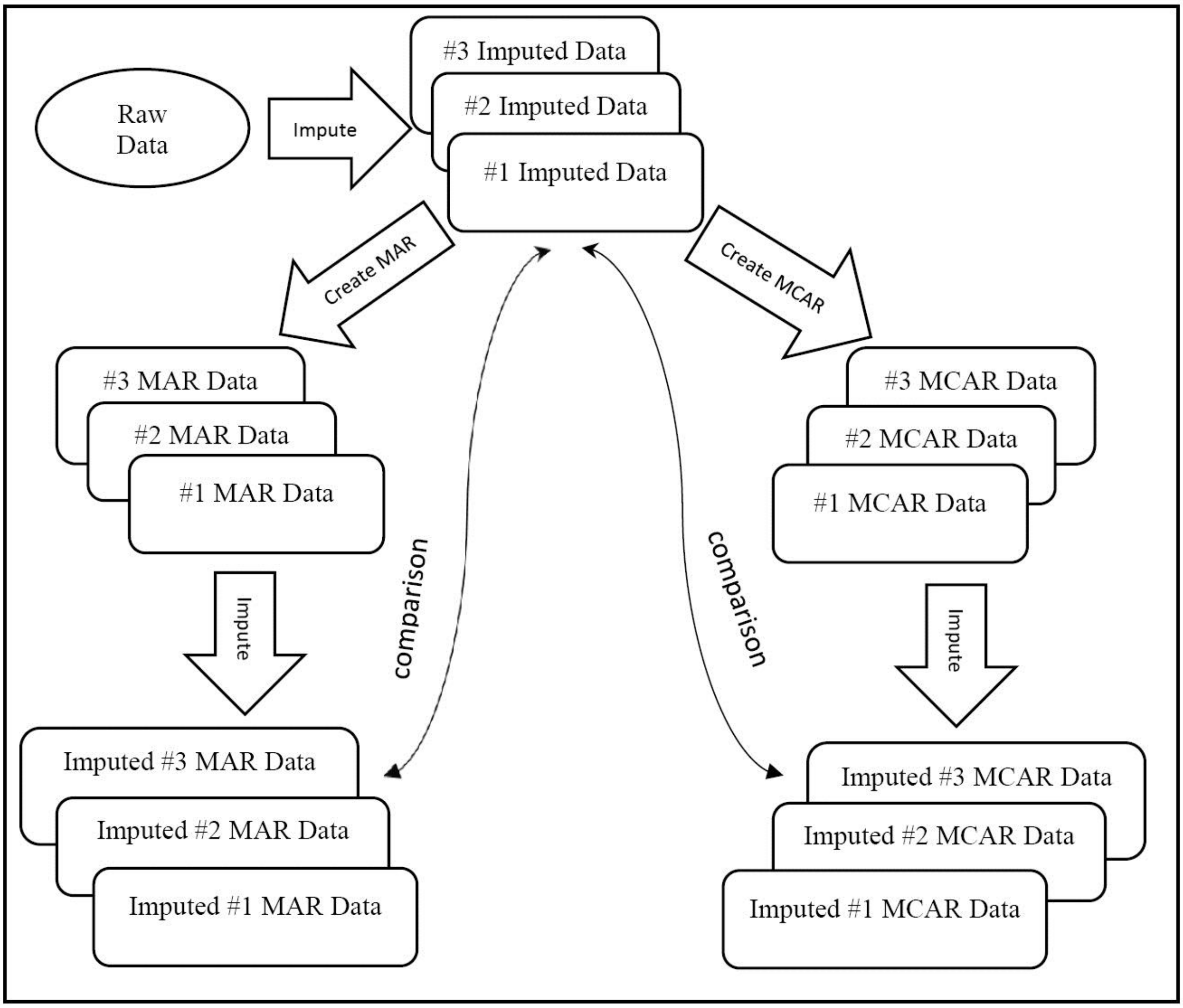


Figure 2: Plot of missingness patterns of the original data against three copies of data with simulated missing completely at random mechanism on the imputed data.

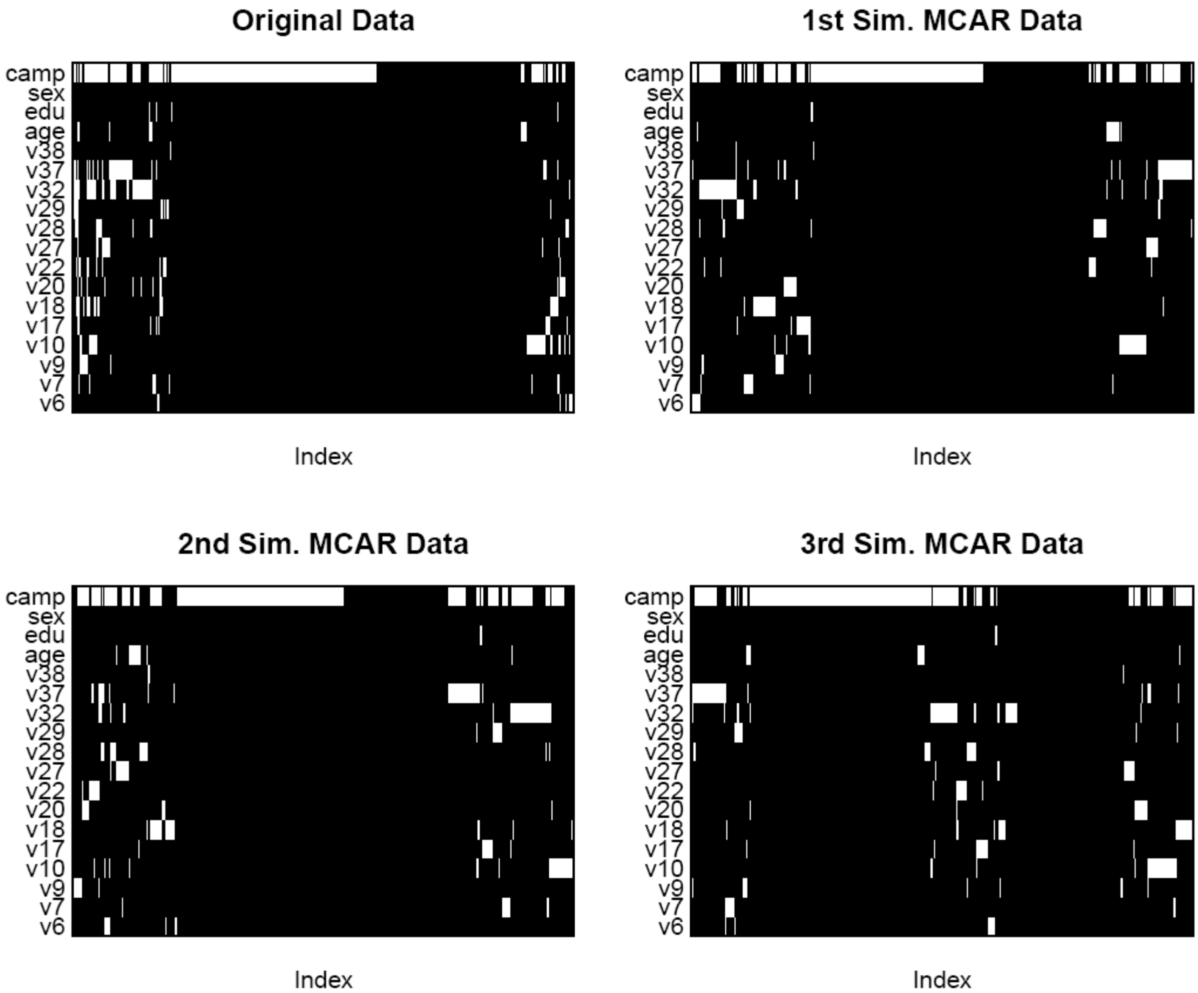
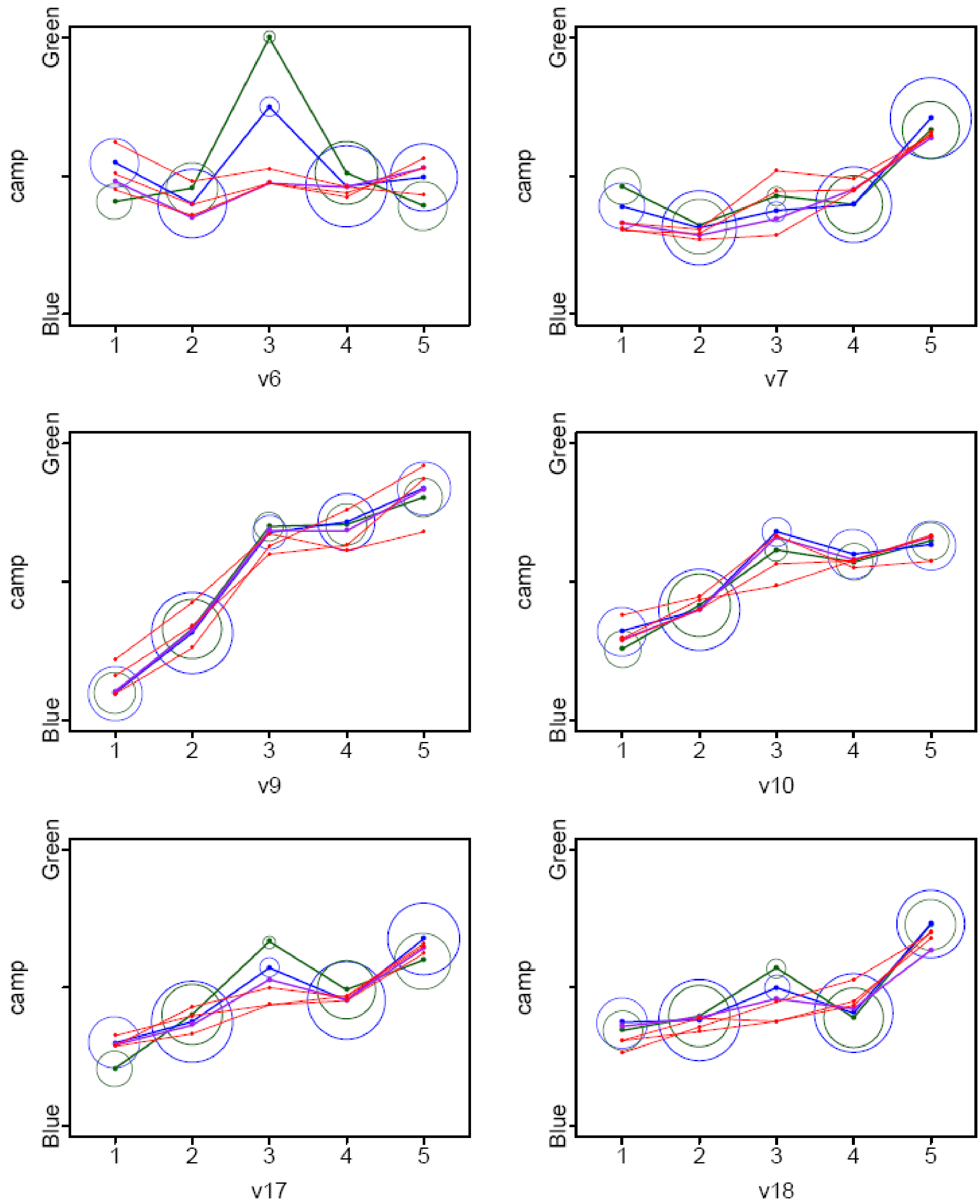


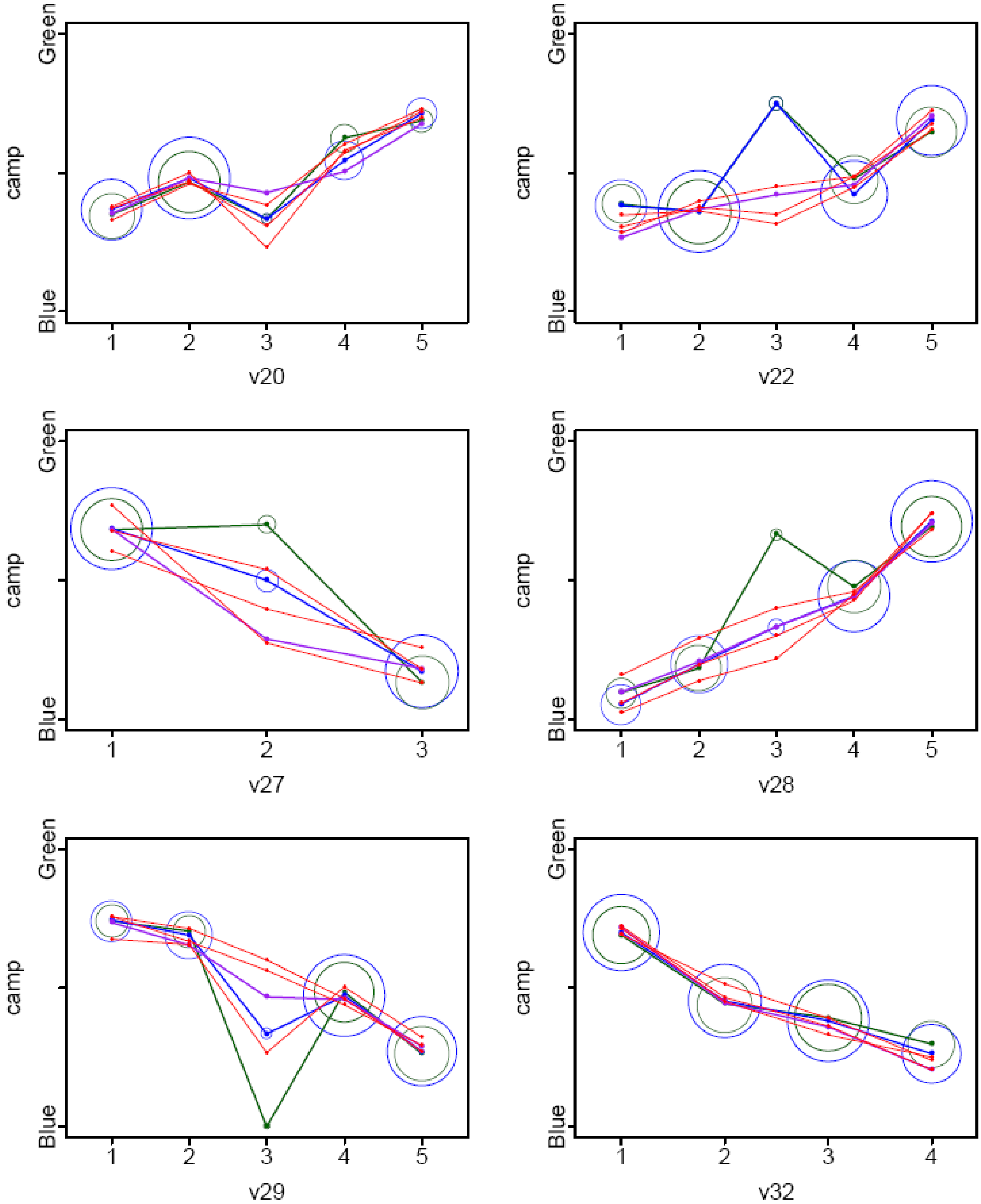
Table 2: Summary of the `camp` variable between the original data and imputed MCAR datasets.

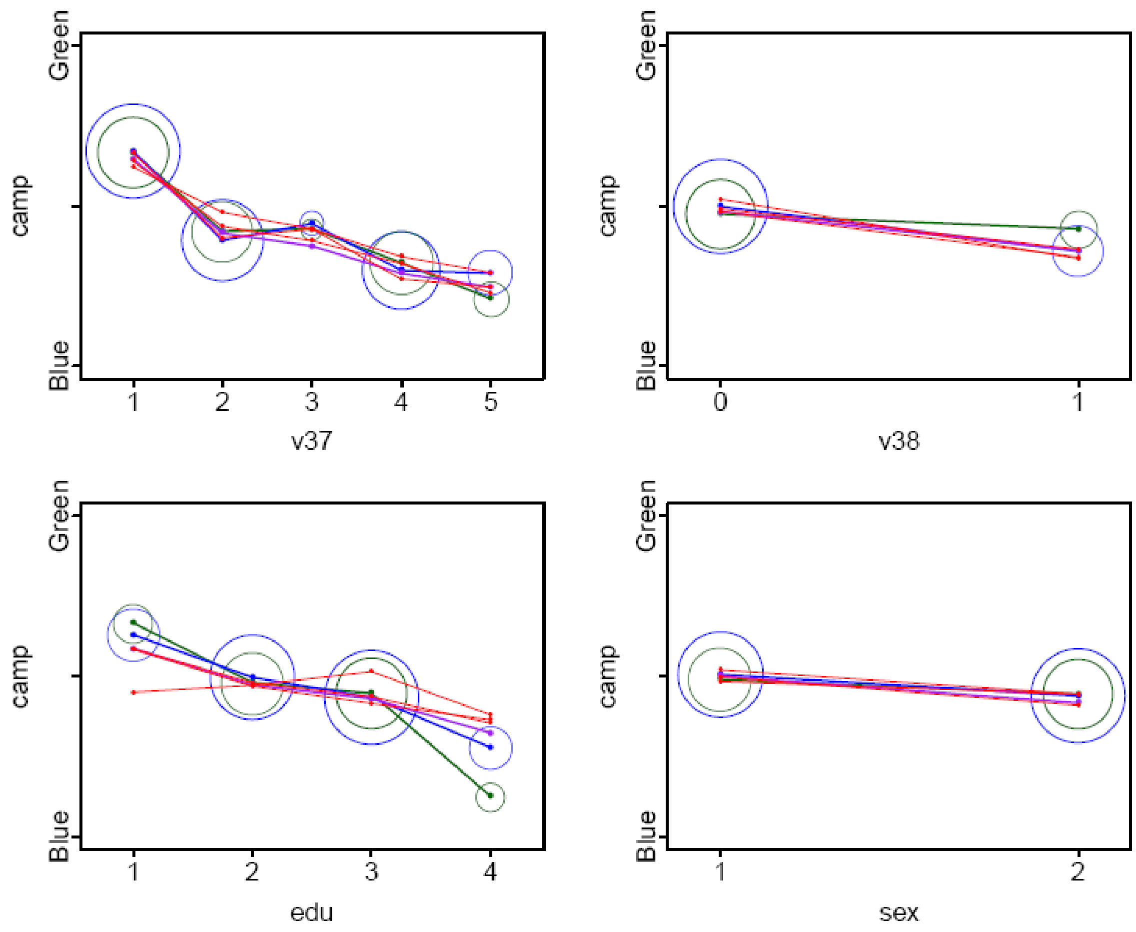
	Mean	SE	Blue	Green
Original Data	0.488	0.500	51.1%	48.8%
Imputed Original Data	0.474$^+$	0.500$^+$	52.6%	47.4%
#1 Imputed MCAR Data	0.464$^+$	0.499$^+$	53.5%	46.5%
#2 Imputed MCAR Data	0.464$^+$	0.500$^+$	53.6%	46.4%
#3 Imputed MCAR Data	0.453$^+$	0.500$^+$	54.7%	45.3%

Note: $^+$ The mean's and SE's reported here are pooled mean and SE's for three chains of MI.

Figure 3: Plots of Camp Variable against Other Variable Using the Imputed MCAR data.




Green
camp
Blue
1 2 3 4 5
v20
Green
camp
Blue
1 2 3 4 5
v22
Green
camp
Blue
1 2 3
v27
Green
camp
Blue
1 2 3 4 5
v28
Green
camp
Blue
1 2 3 4 5
v29
Green
camp
Blue
1 2 3 4
v32



Note: The red lines are the three chains of imputations on MCAR data. The purple lines represent the estimates of the imputed original data. The dark green lines represent the completed cases of the original data. The blue lines represent the available cases of the original data. The size of the circles indicate the size of the observation in each categories.

Figure 4: Plot of Missingness Patterns of the Original Data against Three Copies of Data with Simulated Missing at Random Mechanism on the Imputed data

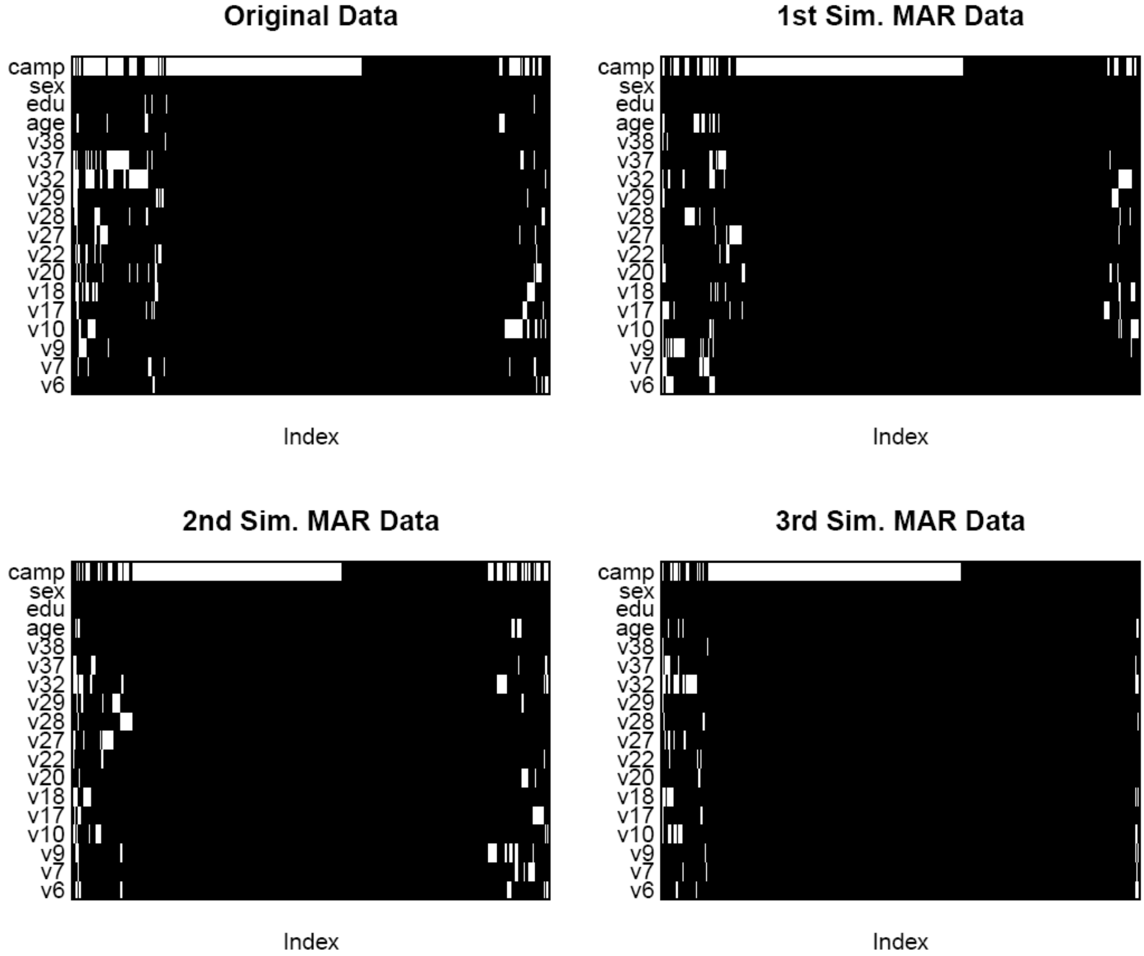
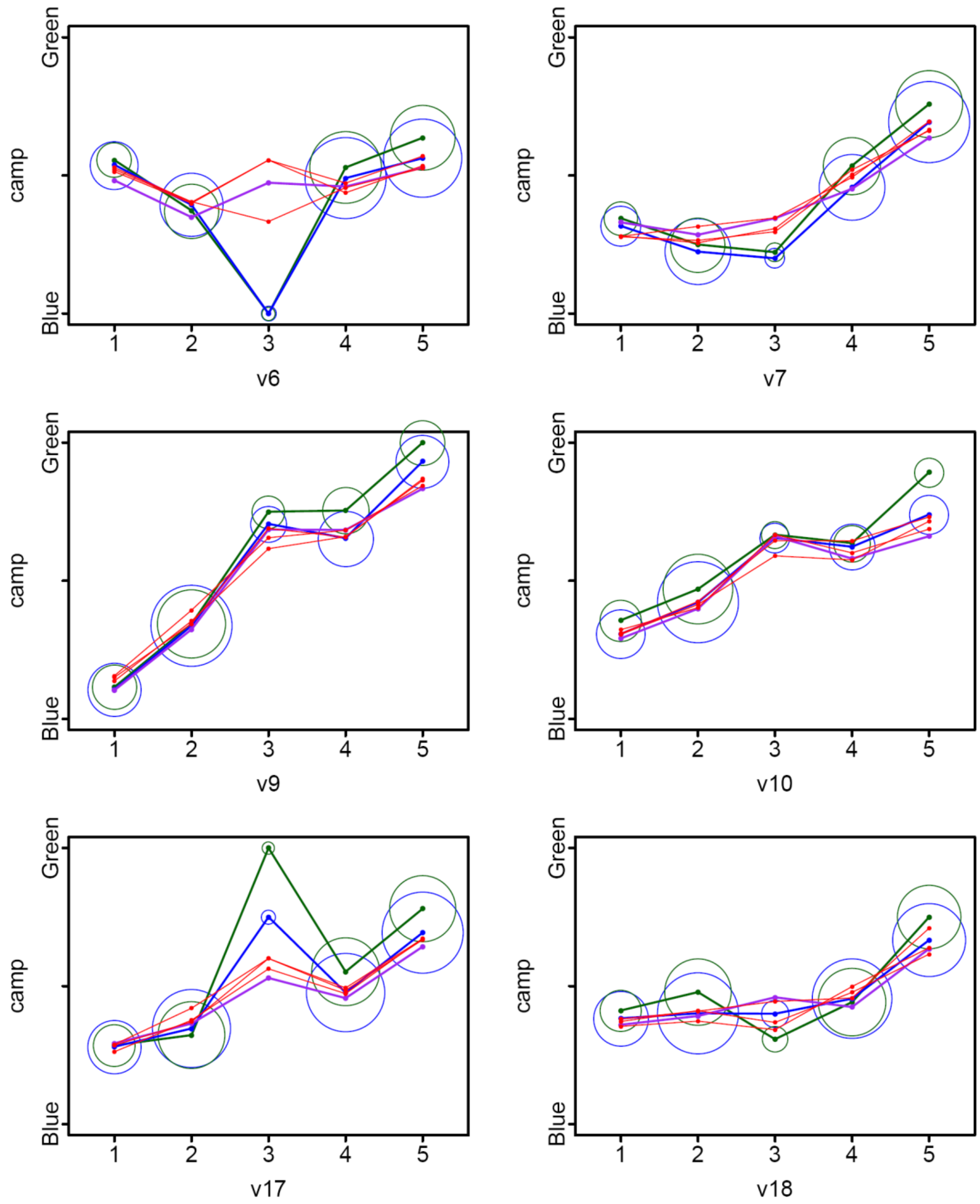


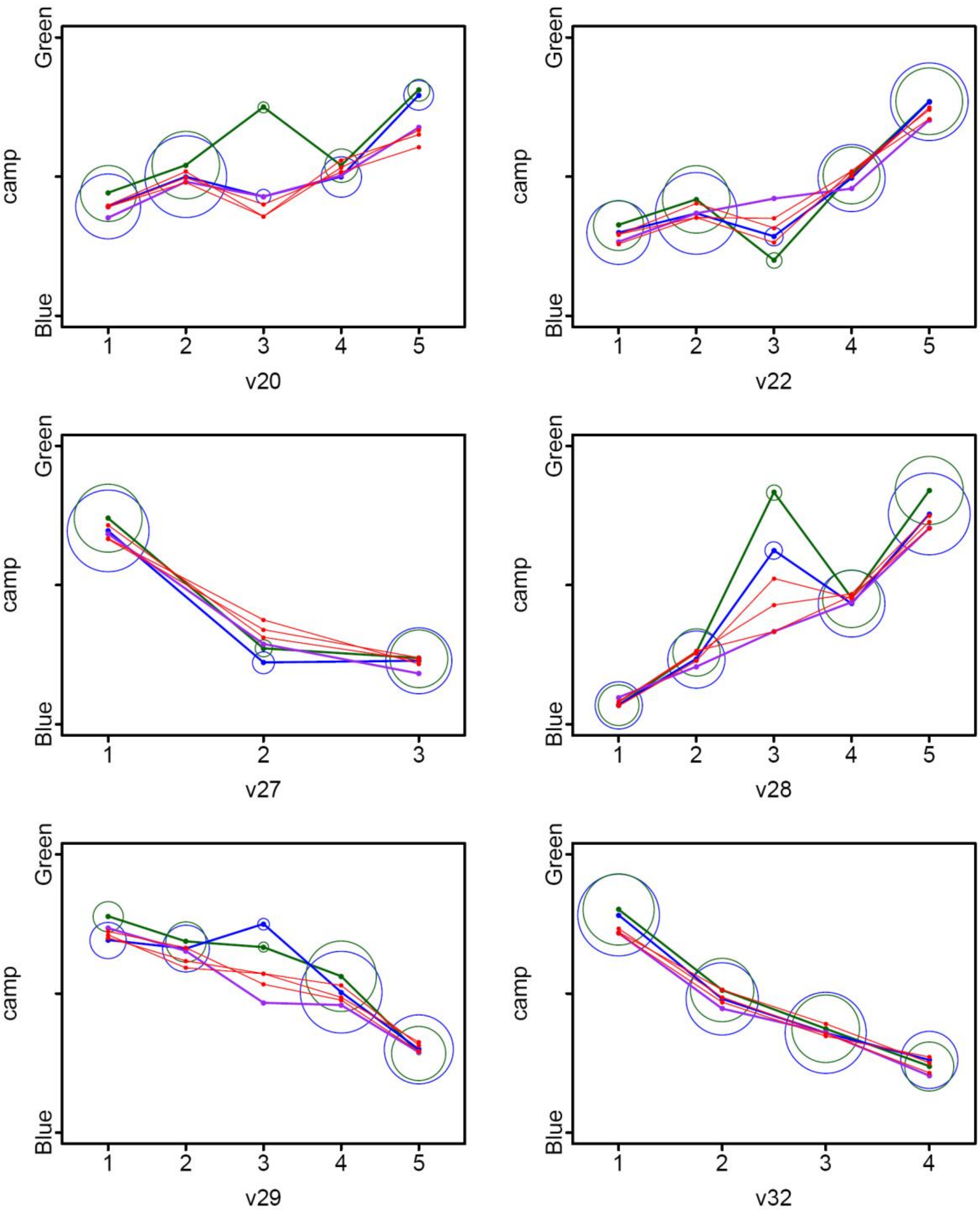
Table 3: Summary of the `camp` variable between the original data and imputed MAR datasets.

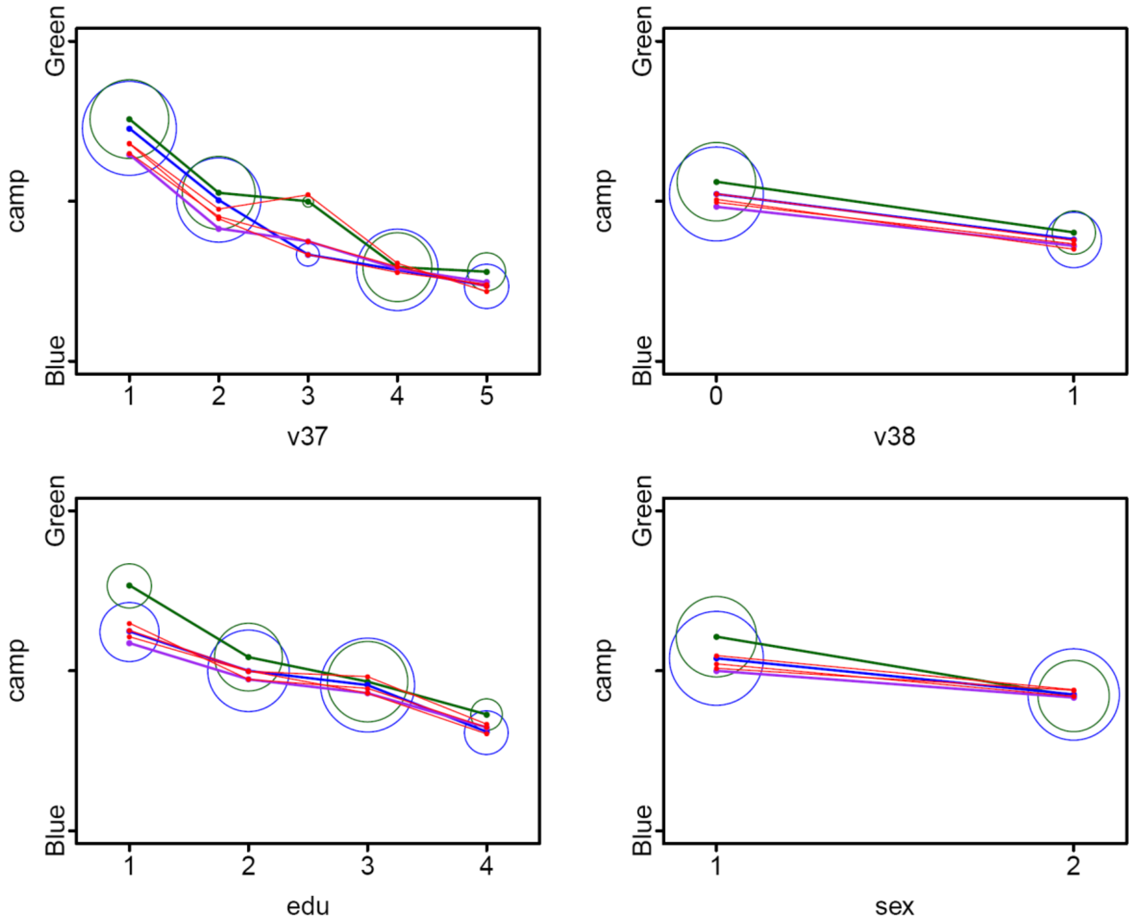
	Mean*	SE*	Blue	Green
Original Data	0.488	0.500	51.1%	48.8%
Imputed Original Data	0.474+	0.500+	52.6%	47.4%
#1 Imputed MAR Data	0.477+	0.500+	50.1%	49.9%
#2 Imputed MAR Data	0.497+	0.500+	50.3%	49.7%
#3 Imputed MAR Data	0.499+	0.500+	50.1%	49.9%

Note: $^{+}$ The mean's and SE's reported here are pooled mean and SE's for three chains of MI.

Figure 5: Plots of Camp Variable against Other Variable Using the Imputed MAR Data.




Green
camp
Blue
1 2 3 4 5
v20
v22
1 2 3
v27
v28
v29
1 2 3 4
v32



Note: The red lines are the three chains of imputations on MAR data. The purple lines represent the estimates of the imputed original data. The dark green lines represent the completed cases of the original data. The blue lines represent the available cases of the original data. The size of the circles indicate the size of the observation in each categories.

Table 4: Summary of In-depth Interviews

ID	Sex	Age	Edu	Date-Time	Place	MI	Camp ID	Causes of Inconsistency
905	F	34	3	2013.4.27 10:00AM	Taipei Main Station, Taipei City	0.42 (B)	G	Disappointed by the KMT's reform on domestic policies. Feeling lost some reasons to keep supporting KMT.
206	M	43	2	2013.4.30 10:00 AM	NSYSU campus, Kaohsiung City	0.50 (B)	B	Disappointed by both political camps but felt more concerned about DPP than KMT regarding DPP's ideology of seeking Taiwan independence.
140	M	29	3	2013.4.30 2:00 PM	A Coffee Shop in Kaohsiung City	0.54 (G)	B	Growing up with KMT supporter mother and have been ok with KMT. Turning to like DPP for a growing Taiwanese national identification.
384	M	25	4	2013.5.4 2:00 PM	Taipei Main Station, Taipei City	0.47 (B)	G	Feeling cross-pressured because parents support for KMT but friends support for DPP. The first vote is for DPP in 2008 presidential election. Feeling disappointed by KMT's performance but not aware the core ideology of DPP.
286	M	37	3	2013.5.4 4:00 PM	Taipei Main Station, Taipei City	0.40 (B)	G	Feeling cross-pressured because his family have been supporting for KMT but wife's family support for DPP. Feeling disappoint about KMT's leadership.

Note:

1. Education level: 01=Under Junior High School; 02=High School; 03=College; 04=Graduate.
2. MI Values close to 1 indicates support for (G) while those close to 0 indicates support for (B). The cutting point of MI is set to 0.5.
3. Camp ID is what respondents gave to a forced choice question in the revisit telephone interview. G denotes support for the pro-DPP or the “Green” political camp while B for the the pro-KMT or the “Blue” camp.

References

Allison, P. D. (2001). *Missing Data* (1st ed.). Sage Publications, Inc.

Barzi, F. (2004). Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology*, *160*(1), 34–45.

Bernaards, C. A., Farmer, M. M., Qi, K., Dulai, G. S., Ganz, P. A., & Kahn, K. L. (2003). Comparison of two multiple imputation procedures in a cancer screening survey. *Journal of Data Science*, *1*(3), 293–312.

Bernhagen, P., & Marsh, M. (2007). The partisan effects of low turnout: Analyzing vote abstention as a missing data problem. *Electoral Studies*, *26*(3), 548–560. doi:10.1016/j.electstud.2006.10.002

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Florez-Lopez, R. (2010). Effects of missing data in credit risk scoring. A comparative analysis of methods to achieve robustness in the absence of sufficient data. *Journal of the Operational Research Society*, *61*(3), 486–501.

Gelman, A., King, G., & Liu, C. (1998). Not asked and not answered: Multiple imputation for multiple surveys. *Journal of the American Statistical Association*, *93*(443), 846–857.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

He, Y., & Raghunathan, T. E. (2009). On the performance of sequential regression multiple imputation methods with non normal error distributions. *Communications in Statistics-Simulation and Computation*, *38*(4), 856–883.

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1–47.

Imai, K., King, G., & Lau, O. (2004). Zelig: Everyone's Statistical Software, http://GKing.Harvard.Edu/zelig. Retrieved from http://GKing.Harvard.Edu/zelig

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, *95*(1), 49–69.

Kropko, J., Goodrich, B., Gelman, A., & Hill, J. (2013). *Assessing the accuracy of multiple imputation techniques for categorical variables with missing data*. Working Paper, New York: Columbia University.

Liu, F. C.-S. (2010). Reconstruct partisan support distribution with multiply imputed survey data: A case study of Taiwan's 2008 presidential election. *Survey Research*, *24*, 135–162.

Liu, J., Gelman, A., Hill, J., & Su, Y.-S. (2012). *On the stationary distribution of iterative imputations*. Working Paper, New York: Columbia University.

Murtagh, F. (1985). *Multidimensional Clustering Algorithms.* (1st ed.). Physica-Verlag HD.

Paul, C., Mason, W. M., McCaffrey, D., & Fox, S. A. (2008). A cautionary case study of approaches to the treatment of missing data. *Statistical Methods and Applications*, *17*(3), 351–372.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York : Wiley.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys* (2nd ed.). Hoboken, N.J. ; Wiley-Interscience.

Snijders, T. A. B., & Bosker, R. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Second Edition.). London: Sage Publications Ltd.

Stuart, E. A., Azur, M., Frangakis, C., & Leaf, P. (2009). Multiple imputation with large data sets: A case study of the children's mental health initiative. *American Journal of Epidemiology*, *169*(9), 1133–1139.

Su, Y.-S., Gelman, A., Hill, J., & Yajima, M. (2011). Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. *Journal of Statistical Software*, *45*(2), 1–31.

Appendix: The Prediction Results of the 43 Respondents Whose Answers Do not Match Prediction

ID	MI	Camp	pred.correct
5	0.161	B	1
8	0.400	B	1
9	0.804	G	1
15	0.728	G	1
24	0.348	B	1
25	0.259	B	1
36	0.300	B	1
44	0.574	G	1
58	0.587	B	0
59	0.730	G	1
69	0.382	B	1
72	0.336	B	1
85	0.637	G	1
91	0.283	B	1
93	0.403	B	1
96	0.545	B	0
99	0.257	B	1
118	0.710	G	1
140	0.540	B	0
141	0.396	G	0
153	0.470	G	0
166	0.268	B	1
169	0.352	B	1
180	0.360	G	0
186	0.477	G	0
187	0.534	G	1
191	0.610	G	1
206	0.500	B	1
207	0.491	G	0
214	0.262	B	1
217	0.667	G	1
223	0.505	B	0
231	0.408	B	1
235	0.190	B	1
243	0.543	B	0
262	0.378	B	1
269	0.444	G	0
274	0.369	B	1
286	0.400	G	0
287	0.620	G	1
299	0.650	B	0
339	0.765	G	1
342	0.664	G	1
352	0.529	G	1
364	0.206	B	1
366	0.481	G	0
370	0.469	G	0
377	0.639	G	1
384	0.466	G	0
396	0.269	B	1

401	0.525	G	1
402	0.332	B	1
409	0.464	B	1
418	0.320	B	1
420	0.636	G	1
421	0.767	G	1
432	0.390	B	1
434	0.561	G	1
435	0.257	B	1
447	0.432	B	1
451	0.412	G	0
460	0.674	B	0
463	0.381	G	0
470	0.499	B	1
471	0.431	B	1
495	0.319	B	1
503	0.582	G	1
524	0.765	G	1
530	0.491	B	1
538	0.443	G	0
543	0.608	G	1
554	0.750	G	1
558	0.255	B	1
564	0.680	G	1
566	0.642	B	0
567	0.404	B	1
575	0.342	B	1
583	0.372	B	1
597	0.279	B	1
631	0.297	B	1
635	0.373	B	1
644	0.305	B	1
645	0.385	G	0
650	0.380	B	1
663	0.465	B	1
670	0.687	G	1
678	0.345	B	1
696	0.778	G	1
703	0.381	B	1
725	0.484	G	0
726	0.343	B	1
730	0.518	B	0
744	0.755	G	1
761	0.713	B	0
762	0.502	G	1
763	0.546	G	1
768	0.380	B	1
778	0.562	B	0
807	0.549	B	0
811	0.576	G	1
821	0.452	G	0
823	0.355	B	1
830	0.294	B	1

831	0.578	G	1
842	0.279	B	1
855	0.490	G	0
874	0.406	B	1
875	0.584	B	0
882	0.657	G	1
885	0.509	G	1
888	0.573	G	1
889	0.576	B	0
899	0.483	B	1
903	0.742	G	1
905	0.423	G	0
913	0.598	G	1
914	0.284	B	1
920	0.274	B	1
922	0.358	G	0
928	0.666	G	1
933	0.437	G	0
948	0.336	G	0
953	0.431	B	1
956	0.764	G	1
958	0.516	G	1
965	0.304	G	0
969	0.707	B	0
974	0.756	G	1
988	0.517	B	0
994	0.348	G	0
995	0.491	B	1
996	0.236	B	1
1000	0.495	G	0
1001	0.424	B	1
1012	0.721	G	1
1016	0.622	G	1
1029	0.470	G	0
1034	0.571	G	1
1053	0.245	B	1
1055	0.600	G	1
1057	0.446	G	0
1059	0.504	B	0
1074	0.240	B	1

Note: The cutting point of MI is set to 0.5.