

Research Article

Special Issue of Survey Research in the fields of Political Science

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

Frank C. S. Liu^{}, Yu-Sung Su^{***}**

ABSTRACT

Item non-response is endemic to most survey studies, and hinders the researcher in making sensible inferences. One plausible solution to this problem, multiple imputation (MI), is becoming a widely used approach in dealing with the problem of missing data thanks to the development of various software packages. Nonetheless, MI is not a

* The first draft of the paper was presented in AAPOR 2013 Annual Conference, Boston, USA, April 16–19, 2013. This project is supported by financial support from Ministry of Science and Technology (MOST 101-2410-H-110-041-SS3) and the “Aim for the Top University Plan” of the National Sun Yat-sen University, Taiwan. The authors are grateful for comments from Michael Dennis and the two anonymous reviewers of this paper.

The R code script of this study is available online: <http://goo.gl/S5LNxc>.

** Associate Professor, Institute of Political Science, National Sun Yat-Sen University, Taiwan. E-mail: csliu@mail.nsysu.edu.tw. Tel: +886-7-5252000 ext. 5555.

*** Associate Professor, Department of Political Science, Tsinghua University, China. E-mail: suyusung@tsinghua.edu.cn

panacea. Imputing missing data using MI without checking it may further induce biases. This oblivious use of MI arises partly from the conviction that some MI assumptions are simply mathematically unverifiable. Hence, the goal of the paper is twofold: first, it demonstrates how various post-MI diagnostics can be performed with a telephone survey dataset collected in Taiwan in early 2013; secondly, it places greater emphasis on the external validity of MI with a follow-up survey, and compares imputed values to the real ones. This paper concludes that, with a sensible application of MI and accompanying diagnostics, we are able to adjust survey response distribution and, at the same time, elaborate on the inferences in our studies.

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

應用多重插補法評估選民政黨傾向的可行性： 以電話調查中不表態中間選民為例

劉正山* 蘇毓崧**

摘要

電話調查中政黨傾向的「項目無反應」是選舉研究中值得正視的現象。愈來愈多台灣民眾在接受調查訪問時，不願表露本身政黨傾向，或選擇「中立」來回避作答，造成研究人員無法從調查數據中正確地估計選民政黨傾向的分佈，進而誤判選舉結果。多重插補法是解決這類資料缺失問題的統計方法之一。嚴格來說，學者尚無法確定什麼情況下可以放心的使用多重插補法來估計不表態民眾的政黨傾向分佈：原因一是政黨傾向並非隨機遺漏；第二是研究人員尚無法掌握檢驗數據遺漏機制。我們使用 2013 年收集的全國性的電訪資料，展示多重插補法如何有效地解決這個問題。本文首先使用模擬遺漏機制的

* 國立中山大學政治學研究所副教授，通訊作者：csliu@mail.nsysu.edu.tw。

** 中國，清華大學社會科學學院政治學系副教授。

方式，比較多重插補前後數據的差異，指出控制與政黨傾向相關的變數會使政黨傾向的遺漏接近隨機；其次，我們比對插補後的資料及電訪追訪結果，並對中間選民進行深度訪談後，發現經過遺漏值分布和數據遺漏機制檢驗的政黨傾向插補數據具有高度的外部有效性。此方法和檢驗程序亦適用於其他非隨機遺漏的「項目無反應」研究。

關鍵字：多重插補、項目無反應、遺漏值、外部有效性檢證

I. Introduction

Item non-response is endemic to most survey studies and hinders the researcher in making sensible inferences. Scholars have recently become aware that it is inappropriate to make electoral forecasts based on information simply drawn from raw survey and poll data. Respondents who hesitate to disclose their preferences and attitudes can create an item-non-response problem for controversial or sensitive survey questions, such as partisanship or stances on moral issues. Calculating proportions based on the raw data and omitting the non-response data results in biased proportions for interested variables (Bernhagen & Marsh, 2007).

One plausible solution to this issue, multiple imputation (MI), is becoming a widely used approach in dealing with this missing data problem thanks to the development of various software packages, e.g., Amelia (Honaker et al., 2011); mice (Buuren & Groothuis-Oudshoorn, 2011); mi (Su et al., 2011), which make MI more accessible to regular researchers.

Electoral scholars, in particular, have started paying attention to the MI approach and applying it to electoral forecasts. Bernaards et al. (2003) compares the 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 use conventional MI methods to study the relationship between explanatory variables and a chosen response variable, they imply that using MI to study dependent variables is one possible method. That is, scholars can pay more attention to advancing the accuracy of descriptive statistics in dependent variables than explaining their variance. Although this novel focus is absent from Rubin’s treatment of MI methods for nonresponses in surveys (Rubin, 1987), there is little methodological reason to object to the imputing of dependent variables for electoral forecasting. In effect, studying the vote choices of non-voters was proposed at the time when MI was introduced to the discipline (King et al., 2001; Snijders & Bosker, 2011). G. David Garson confirms this perspective in his course website and states “for purposes of univariate analysis (e.g., understanding the frequency distribution for how subjects respond to an opinion item) imputation can reduce bias and often is used for this purpose if data is missing at random.”¹

Nevertheless, MI is not a panacea. Imputing missing data with MI without further verification may induce additional biases. Such oblivious conduct regarding MI arises partly from the conviction that some MI assumptions are simply mathematically uncheckable. Hence, the goal of this paper is twofold: first, it demonstrates how various post-MI diagnostics can be performed with a telephone survey dataset; secondly, it places greater emphasis on the external validity of MI with a follow up survey, and com-

1. See, <http://faculty.chass.ncsu.edu/garson/PA765/missing.htm>

pares imputed values with real ones. Additionally, this paper attempts to apply MI to improve electoral forecasts. It echoes the findings of other similar studies and argues that MI is a cost-efficient and methodologically sound approach for better using raw survey and poll data (e.g., Barzi, 2004).² The paper concludes that, with a sensible application of MI and the diagnostics to verify it, we are able to adjust survey response distribution and, at the same time, enrich our understanding of the data, as well as provide more elaborate inferences for our studies.

II. Multiple Imputation with Diagnostics

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 et al., 2008; Rubin, 2004). 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 with the imputed datasets, the researcher first runs regression models using every dataset. Hereafter, she averages the coefficients and standard errors across these models (Honaker et al., 2001, Snijders & Bosker,

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

2011, Stuart et al., 2009).³ MI is a method commonly used to deal with missing data problems, 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 et al., 1998). This concern regarding biased estimates can be minimized if the loss of cases due to missing data is less than approximately 5%, and if pretest variables can reasonably be included in models as covariates (Graham, 2009).

Two major algorithms are commonly used in the existing MI software packages. One is joint MI (e.g. Amelia), and the other is conditional MI (e.g. mi and mice). Joint MI completes its calculations much more quickly than conditional MI. As King, Honaker, Joseph, and Scheve (2001) argue, EM is a faster and less complex alternative to imputation posterior (IP). Concerned that the EM algorithm ignores estimation of uncertainty, they propose EM is (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 mi and mice when handling computation.

Conditional MI, on the other hand, weights more on assumptions reflected in algorithms than on those focusing on calculation speed. Joint

3. While some scholars may believe 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 et al., 2009: 1134).

MI assumes that the data follows a multivariate normal distribution. It performs variable transformation before the imputation to make the data distribute normal and then uses transformation after the imputation to recover the original format of the data. That joint distribution is multivariate normal seems to be a naive assumption because the data might contain binary, ordinal, (unordered) categorical and other special types of variables, all of which are not of normal distribution. Henceforth, 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 using conditional MI.

He and Ragunathan (2009) conduct a series of experiments and compare the performance of MI using sequential regression (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 MI results can be extremely biased when error distributions possess extremely heavy tails, i.e., when data includes extreme values. Therefore, it is proper to use MI as a tool for avoiding extreme or impossible values and relaxing the joint normal assumption. Conditional MI relaxes the assumption of multivariate normality of the data.

Nevertheless, MI is not a magic algorithm that will recover the missing values of data. There are several assumptions that need to apply in order to ensure the quality of 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), meaning that the chosen missingness indicators are independent of the unobserved data. In other words, conditional on the other observed variables, the missing mecha-

nism does not depend on the unobserved data. Snijders and Bosker (2011) believe it is proper to “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 types of missing mechanism are missing completely at random (MCAR) and not missing at random (NMAR). MCAR means that the missingness indicators are independent of the complete data; NMAR is a situation in which missingness is not at random and will always depend on untestable assumptions. In many cases, researchers ignore the missing data and use complete case analysis. This is equivalent to assuming that their data is MCAR. Clearly, MCAR is a rather strong assumption that is seldom found in most data sets. MAR is a comparatively relaxed assumption about the missing mechanism. However, those who use joint MI and conditional MI are equally at fault if they do not verify their data is MAR.

Secondly, the conditional model should be appropriately specified. Since conditionality is a major component of the imputation procedure, a conditional model that is inappropriate might lessen the accuracy of the prediction. Fortunately, this assumption is less of a problem if a conditional model contains many variables (ignobility can be reached) and if these variables are valid (each conditional family contains the true probability distribution) (Liu et al., 2014).

A sensible MI is a practice of MI with correct procedure plus several checks (diagnostics), which produce results that makes sense to researchers in terms of validity and coherence. In the following sections, we demonstrate the way in which a sensible MI can be achieved after checking the aforementioned assumptions with telephone survey data. Direct checking of the MAR assumption is mathematically impossible. Nonetheless, we show the way in

which the checking of this assumption is still attainable via simulations.

III. The Data and Research Design

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

The target or dependent variable is the political camp with which the respondent identifies—either the Blue camp (the pro-Kuomintang, KMT, camp) or the Green camp (the pro-Democratic Progressive Party, DPP, camp). As Table 1 shows, the missing rate for this variable is 61%; 658 out of 1,078 respondents conceal their partisanship in telephone surveys. Partisanship, measured by the question “Which political party do you support?” has been a “sensitive” question in Taiwan. It has been commonly seen in polls that one-third (in face-to-face surveys) to half (in telephone survey) of Taiwanese respondents refuse to reveal to interviewers their party asso-

ciation. As 61% is even higher than the common cases, we suspect that the number of citizens concealing their partisan preferences in telephone surveys has significantly increased.

Ostensibly, we have no direct clue that the missing mechanism for the data is missing at random. And simply assuming either MCAR or MAR would be erroneous. Nevertheless, if we carefully and correctly choose sufficient auxiliary variables which are strong predictors of the target variable, and which can predict the missingness of such variables well, it is still possible to attain MAR through conditionality. The auxiliary variables chosen for imputation are listed in Table 1. These variables are chosen based on empirically supported evidence proving that Taiwanese national identification is strongly related to partisanship. We then select 18 categorical variables from the dataset and conduct the following analysis using a conditional MI software package, mi.⁴

We are thus going to using the word MI in the following sections to represent the method as well as the algorithm we use in the analysis. The specific algorithm of MI is predicting the missing data by regressing missing outcome with all other variables iteratively with multiple chains until it reaches convergence (the \hat{R} statistics of imputed variables are below 1.05 (Gelman et al., 2003)). We do acknowledge that multiple imputation methods have many variants. Kropko et al. (2013) have done several simulation tests and proved the superiority of mi (Su et al., 2011). Hence to avoid repetition, we limit ourselves to conduct the following analysis with mi.

4. The R package mi takes advantage of existing regression models in dealing with various kinds of variables: it uses 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 et al., 2011).

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 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) |

Table 1 Summary of Variables (Continued)

| Variables | Question Wording | Distribution | Missing (%) |
|-----------|--|---|---------------|
| 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 about 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 have a more restrictive policy on Mainland Chinese tourists? | Strongly disagree: 169; Disagree: 358; Neutral: 19; Agree: 246; Strongly agree: 258 | 28 (3.90) |
| v27 | Some call themselves 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 belong to the 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 unification with mainland China if it becomes a democracy? | Strongly yes: 270; Yes: 229; No: 326; Strongly no: 147 | 106 (9.83) |

Table 1 Summary of Variables (Continued)

| Variables | Question Wording | Distribution | Missing (%) |
|-----------|--|--|--------------|
| v37 | Do you agree that the two sides of Taiwan Straight will eventually become one country? | Strongly disagree: 378; Disagree: 283; Neutral: 22; Agree: 238; Strongly agree: 70 | 94 (8.72) |
| v38 | Have you been to mainland China in the past two years? | 1=yes; 2=no. | 4 (0.37) |
| age | In 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 highest level of education? | 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.

To ensure that, based conditionally on the 18 auxiliary variables, the data is MAR, we use the following steps to experimentally check the missingness mechanism of the data:

1. Impute missing data with MI. Run each MI with 5 chains and iterate the MI until reaching the convergence for which the Rubin-Gelman \hat{R} statistics are smaller than 1.1 (Gelman et al., 2003).
2. Several checks on the fitness of the conditional models in step 1 are per-

formed to ensure the efficacy of the MI. Then we alter the conditional models and rerun the MI again.

3. Obtain imputed datasets once the MI is done. These imputed datasets are now treated as the “true” datasets, or the baseline, for the proceeding comparison.
4. Forge MAR missingness in the imputed datasets and create several copies of the data. Predictive missingness (the forged MAR missingness) of each variable is obtained by regressing its missing indicator on all other variables with logistic regression.
5. Performing the MI again on these newly created datasets and obtaining imputed datasets.
6. Comparing these newly imputed datasets with the imputed datasets from the step 1 and the original data. If there is no significant deviation or difference, we can claim that, by controlling for those auxiliary variables, the data is indeed MAR.

It is impossible to know how well MI does without the true values for the missing data. In the step 2, the imputed datasets are used as the true datasets only for the purpose of checking whether or not the data is indeed MAR. We still lack the “true” data necessary to absolutely verify MI performance because only survey respondents themselves know their true preferences. These true preferences can only be obtained through a follow up survey or interviews with those who failed to identify their political affiliation in the first round. We therefore conduct a follow up telephone survey to obtain answers concerning the missing data.

The external validation for MI is performed by comparing respondents’ answers in the follow up survey with the MI imputed values to determine the accuracy of MI predictions. More importantly, this will provide an under-

standing of why predictions are not accurate, if this is the case.

Because of ethnic and privacy concerns, as well as time and funding limitations, we found it is difficult to reach all 1,078 respondents and ask them again the question they were reluctant to answer the first time. In fact, we could only ask those who consented to be contacted again (484 out of 1,078 agreed). Therefore, we chose an alternative strategy: reaching out to those who did not answer the political party question. We called the 658 respondents (61% of the sample) who we contacted from April 13 to April 15, 2013 and, not to our surprise, only 143 completed the survey in this round.

Given the follow-up telephone survey data set, we contrast our guesses with the answers respondents were forced to provide in the second round (support for the Blue or the pan green camp). The questionnaire for the second round of telephone interviews is shorter than the first one, and is composed of only a few questions, including demographics, two questions to verify answer consistency—whether they had been to mainland China in the past two years, the frequency with which they watched political news, and the question of political camp identity they avoided answering in the first round of telephone surveys.

Next, we focus on those whose answers were predicted incorrectly by MI. Our study then made the third round of calls to those who were willing to be questioned further between April 20 and May 6, 2013 (45 out of 143). Hereafter, we selected 5 willing-to-tell respondents from this pool of 45 people for further face-to-face interviews.

We acknowledge that the representativeness of these follow up samples is an issue that shackles our external validation. Nevertheless, by combining MI analysis with this follow up survey, we garner rich stories from those who

failed to provide an answer in the first interview and, at the same time, partially verify the external validity of the MI in a minimal and imperfect sense.

IV. Results and Findings

Figure 1 displays the missingness pattern of the original data and the five copies of the simulated MAR data. Data sets are grouped together by similar missingness using hierarchical clustering. These five simulated MAR data sets are created from the five imputed datasets. Clearly, none of these five simulated MAR data sets have missingness patterns which are similar

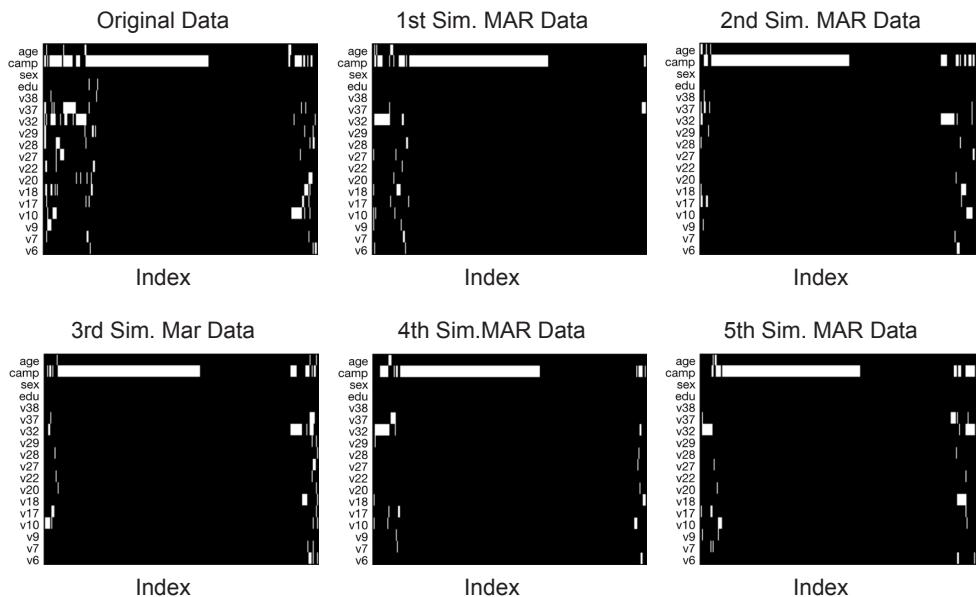


Figure 1 Plot of missingness patterns of the original data against five copies of data with simulated missing at random mechanism on the imputed dataset.

to each other, nor do any of them have a missingness pattern identical to that of the original data. Nevertheless, we are not looking for a perfect match here. As a matter of fact, if these five datasets can pick up partial missingness patterns for the original one after pooling, we might still be able to approximate the true distribution of the original data. Figure 1 complies with this scenario.

Table 2 shows exact binomial test results for the camp variable of the imputed datasets and of the imputed MAR datasets.⁵ For full samples the pooled imputed MAR datasets are not significantly different from the imputed datasets. If we look only at the sample that is missing in the camp variable, once again, pooled imputed MAR datasets are not significantly different from the imputed datasets. This shows that, if we control for these 18 variables, the missing mechanism of the data will be MAR. Henceforth, we can proceed to the MI external validation analysis.

The first wave of telephone surveys provides raw data for the target variable camp and auxiliary variables. In follow up telephone interviews, we asked the 658 respondents who did not answer the camp preference question 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 respond, including 74 who chose the Blue camp and 69

5. We conducted two types of the two-tailed binomial tests. The first test is to compare the distribution of the camp variable as whole of the baseline imputed dataset with that of the imputed MAR dataset. Hence the sample size of the tested (n) is 1078, which is large enough to make binomial test work (see the top panel of full sample of the Table 2). The second test is to compare the distribution of the sample of the missing camp variable of the baseline imputed dataset with those of the imputed MAR dataset. In this test n is 658, which is large enough to make binomial test work (see the bottom panel of sample of missing of the Table 2). The results of the two tests are consistent.

Table 2 Comparison of the camp variable distribution between the original data and imputed datasets.

| | Imputed MAR Data | | Exact Binomial Test p-value |
|------------------------------|------------------|--------------------|-----------------------------|
| Original Data | 0.488 | | |
| <i>Full sample</i> | | | |
| #1 Imputed Data | 0.497 | 0.475 ⁺ | 0.152 |
| #2 Imputed Data | 0.470 | 0.477 ⁺ | 0.626 |
| #3 Imputed Data | 0.493 | 0.463 ⁺ | 0.051 |
| #4 Imputed Data | 0.499 | 0.498 ⁺ | 0.976 |
| #5 Imputed Data | 0.483 | 0.499 ⁺ | 0.300 |
| <i>Sample of the missing</i> | | | |
| #1 Imputed Data | 0.503 | 0.476 ⁺ | 0.161 |
| #2 Imputed Data | 0.459 | 0.470 ⁺ | 0.459 |
| #3 Imputed Data | 0.497 | 0.459 ⁺ | 0.051 |
| #4 Imputed Data | 0.506 | 0.492 ⁺ | 0.483 |
| #5 Imputed Data | 0.480 | 0.489 ⁺ | 0.640 |

Note: ⁺The means reported here are pooled means of 5 chains of MI.

who chose the Green camp.

We perform the exact binomial test again on this sample of 143 to obtain our MI result. The test results show there is no significant difference between the two samples, indicating that these two distributions are not statistically different.

Next, we simulate the estimated answers for these 143 camp variable respondents from the MI results. We arbitrarily slice simulated predictive probabilities into 5 different groups using the cut off points of [0, 0.2, 0.4, 0.6, 0.8, 1]. We label 0 as strong blue, 0.2 as blue, 0.4 as ambivalent blue,

0.6 as ambivalent green, 0.8 as green, and 1 as strong green. Table 3 contrasts the MI camp variable predictions with the real answers obtained from the follow up survey. The accuracy rate of the MI predictions is 75%. Correct predictions were made for 110 of the 143 respondents' political camp choice. Aware of the fact that MI results cannot completely replace individuals' true attitudes, we take this part of analysis as a means of choosing respondents for further interviews. As shown in Table 3, of the 33 respondents for whom predictions are incorrect, 17 have values falling into the categories of ambivalent blue and green. We further interview 5 of them to explore the causes of this ambivalence.

Three patterns can be drawn from our interviews with the 5 respondents who are ambivalent blue or ambivalent green but gave political camp responses inconsistent with our predictions. It is important to note that we did not mention to them our MI predictions of their political orientation until interviews were concluded. The profiles of the five respondents are summarized in Table 4.

The first pattern drawn from the interviews is that they are very politi-

Table 3 Comparison of the MI predictions of the camp variable to the answers of the follow up survey.

| | Blue | Green |
|------------------|------|-------|
| Strong Blue | 30 | 5 |
| Blue | 12 | 4 |
| Ambivalent Blue | 18 | 10 |
| Ambivalent Green | 7 | 12 |
| Green | 3 | 10 |
| Strong Green | 4 | 28 |

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 | AB | 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 | AG | 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 | AG | 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 | AB | 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 | AB | 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. AB stands for ambivalent blue and AG stands for ambivalent green.

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 pro-KMT or “Blue” camp.

cally aware and do not avoid discussing politics with us. Each interview lasted for more than 30 minutes and some respondents even criticized our question wording (for being too narrow in the definition of “Chinese,” for example). We observed that they were active seekers of political information from TV, newspapers, and online news sources. Therefore, they are aware of controversial issues and influenced by impressions of political issues obtained from the news media.

The second pattern is that they do not want to claim support for a political party or camp without being critical. They chose an opposite political camp (contrasted to our prediction) in the second telephone interview primarily because they felt they were forced to answer. Coercion led to an answer provided with only a short-term evaluation of politics. In April of 2013 the Taiwanese people were concerned about a number of domestic policy reforms including nuclear power plant construction, government retirement plan revisions, health care reform, and other issues. Because of this, we found that four Blue Camp supporters (ID 905, 206, 384, and 286) switched their support to the Green camp in the second round interview because of misgivings about the KMT’s leadership in domestic politics and policy reform.

Feeling cross-pressured is a third reason for respondents to conceal their partisan orientation. Respondents 140, 384 and 286 are representative and classic cases of people living in heterogeneous political communication networks. While they could choose either political camp, their answers, when forced, were based on their short-term evaluation of politics. We found that Blue camp identifiers seemed to be influenced by evaluations of policy and concurrent political issues, while Green camp supporters were affected by nationalism.

V. Conclusion and Discussion

One commonly acknowledged challenge in polls or surveys is item non-response, i.e., when a significant proportion of respondents conceal their preferences in particular questions. In this paper, we take two steps in studying the external validity of applying multiple imputation method to the study of “independent” voters who conceal their partisanship. Overall predictions for using selected auxiliary variables perform well. For those who provided answers after being contacted again for further telephone interviews, we find that scores calculated based on MI results help uncover their partisan orientation, including their level of ambivalence. In our follow-up face-to-face interviews with ambivalent respondents, we find that inconsistency in their answers can be explained.

Our study shows that electoral forecasting MI has great potential for solving item non-response problems commonly found in telephone surveys. MI can assist with the reconstruction of the distribution even when the missing rate for the target question “political camp” (the Blue or the Green camp) is over 60%. We believe that, given a set of thoughtfully chosen auxiliary variables, the probability of using MI to make correct guesses about a targeted variable can be higher than 75% in other variables with lower missing rates.

The findings drawn from interviews with these 5 individuals may not represent the whole body of respondents, who are unwilling to be bothered with political questions. But the information they reveal is helpful to us in ensuring that MI procedure scores are not misleading. Rather, we see that scores are quite consistent and representative of the mindset of respondents

caught in the middle.

Panel-like re-contact shows us that so-called independent voters in Taiwan are likely to be those partisans that have difficulty quickly making a choice while the telephone interview is being conducted. They fail to give an answer corresponding to their overall evaluation of the parties but, rather, give a quick answer that reflects their short-term evaluation of politicians or policy issues, or provide an answer based solely on emotions. Independent voters who are actually ambivalent about identifying with a political camp may not be apolitical or indifferent to politics. Instead, they can be “leaners” or “closet partisans,” those partisans who use the excuse of choosing a candidate rather than a party.

Hence, there is good reason for our concerns about using telephone surveys in probing citizens’ partisan orientation and other controversial issues. Respondents, particularly ambivalent ones, are likely to dodge such questions, and give answers that do not correspond with, or are not consistent with, their belief systems. We suggest that survey institutes (1) keep conventional questions about partisanship but not coerce respondents to answer and (2) encourage them to answer sincerely other auxiliary questions that they find less sensitive. By utilizing MI, researchers will be able to use more sincere answers to reconstruct the target variable distribution for high missing values.

We propose four suggestions for future study. First, we need to start thinking about creating more questions and dimensions for auxiliary variables. In the present study, more than 25 questions we asked were found to have a statistically significant relationship with choice of political camp. We chose 18 of them, including demographics. These questions were mostly related to the concept of state and national identification, the two dimensions

that have been found to be empirically related to one's party identification in Taiwan. We suggest that future studies continue to explore other dimensions and concepts, in addition to testing how other measurements and questions contribute to the success rate of predictions.

Second, we did not (and were not able to) interview all of the 1,078 respondents to create a panel. This prohibits us from completely externally validating our guesses. Future studies using a panel that is composed of volunteer respondents will provide a more solid foundation for testing the external validity of this method.

Third, through in-depth interviews, we found that ambivalent respondents were likely to be influenced by their feelings and emotions concerning controversial issues at the time of the survey. Therefore, we suggest researchers consider adding some policy and performance evaluation questions as auxiliary variables.

Fourth, using questions that force respondent to choose one answer may not be the best strategy for extracting “true” answers from defensive respondents. Alternative methods and more skillfully worded questions are needed. Before a method is found to inquire into respondents' party or camp affiliations, we argue that MI is, at present, a cost efficient way to reconstruct the missing information.

REFERENCES

- Allison, Paul D.
2001 *Missing Data* (1st ed.). Thousand Oaks, CA: Sage.
- 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, Coen A., Melissa M. Farmer, Karen Qi, Gareth S. Dulai, Patricia A. Ganz, and Katherine L. Kahn
2003 “Comparison of Two Multiple Imputation Procedures in a Cancer Screening Survey,” *Journal of Data Science* 1(3): 293–312.
- Bernhagen, Patrick and Michael Marsh
2007 “The Partisan Effects of Low Turnout: Analyzing Vote Abstention as a Missing Data Problem,” *Electoral Studies* 26(3): 548–560.
- Buuren, Stef van and Karin Groothuis-Oudshoorn
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, Andrew, Gary King, and Chuan-hai Liu
1998 “Not Asked and Not Answered: Multiple Imputation for Multiple Surveys,” *Journal of the American Statistical Association* 93(443): 846–857.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin
2003 *Bayesian Data Analysis* (2nd ed.). New York: Chapman and Hall/CRC.
- Graham, John W.
2009 “Missing Data Analysis: Making It Work in the Real World,” *Annual Review of Psychology* 60: 549–576.
- He, Yulei and Trivellore E. Raghunathan
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, James, Anne Joseph, Gary King, Kenneth Scheve, and Naunihal Singh
2001 *Amelia: A Program for Missing Data* 2.01 ed., Vol. 2002. retrieved Jan. 5, 2015, from <http://gking.harvard.edu/amelia/>.
- Honaker, James, Gary King, and Matthew Blackwell
2011 “Amelia II: A Program for Missing Data,” *Journal of Statistical Software* 45(7): 1–47.
- King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve
2011 “Analyzing Incomplete Political Science Data: An Alternative Algorithm for

- Multiple Imputation,” *American Political Science Review* 95: 49–69.
- Kropko, Johnathan, Benjamin Goodrich, Andrew Gelman, and Jennifer Hill
2013 “Assessing the Accuracy of Multiple Imputation Techniques for Categorical Variables with Missing Data,” Working Paper, New York: Columbia University.
- Liu, Jingchen, Andrew Gelman, Jennifer Hill, Yu-Sung Su, and Jonathan Kropko
2014 “On the Stationary Distribution of Iterative Imputations,” *Biometrika* 101(1): 155–173.
- Paul, Christopher, William M. Mason, Daniel McCaffrey, and Sarah A. Fox
2008 “A Cautionary Case Study of Approaches to the Treatment of Missing Data,” *Statistical Methods and Applications* 17(3): 351–372.
- Rubin, Donald B.
1987 *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. New York: Wiley.
2004 *Multiple Imputation for Nonresponse in Surveys* (2nd ed.). Hoboken, NJ: Wiley-Interscience.
- Snijders, Tom A. B. and Roel Bosker
2011 *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). London: Sage Publications Ltd.
- Stuart, Elizabeth A., Melissa Azur, Constantine Frangakis, and Philip Leaf
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, Yu-Sung, Andrew Gelman, Jennifer Hill, and Masanao Yajima
2011 “Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box,” *Journal of Statistical Software* 45(2): 1–31.