Study of Income Nonresponse in a Survey of Japanese Elderly

Chad Stecher, Colby College

A large amount of economic research has been done evaluating the characteristics of individuals who refuse to participate in a survey, unit nonrespondents, and the effects these missing individuals have on the survey sample's respresentiveness of the target population (cf. Hill and Willis 2001; Horowitz and Manski 1998). While the nature of unit nonresponse has been thoroughly examined, the study of individuals who refuse to answer specific questions within a survey, item nonrespondents, has been largely neglected. This negligence is surprising because the loss of information and sample representativeness that results from item nonresponse can be just as significant as the loss of information resulting from unit nonresponse. Moreover, past studies have found that survey questions that ask for a respondent's personal income are particularly prone to item nonresponse (cf. Gronhaug et al. 1988 or Sheth et al. 1980). Since income is one of the most commonly used variables in economic research, the identification of a nonresponse bias from item nonresponse to an income measure has important implications for past and future economic research.

This paper aims to identify significant predictors of item nonresponse to a question about personal income in a sample of elderly Japanese using a logistic regression framework. In this sample, nonrespondents can be grouped into two types, either refusals (called income nonrespondents) or respondents who are physically or mentally unable to answer the survey and require the help of a proxy. The results of maximum-likelihood logistic regression analysis find that gender, being self-employed, and owning a house all have a significant effect on the odds of income nonresponse. Further analyses find that gender, age, years of education, family size, and health status have a significant effect on the probability that a respondent will require a proxy. The presence of significant determinants of nonresponse indicates that a nonresponse bias exists in the income data. Based on the results of these analyses, researchers will be able to adjust future survey procedures to better target potential nonrespondents or, alternatively, to sign biases to missing data.

This paper is organized as follows. Section I provides a more detailed description of the existing research on item and income nonresponse. Section II describes the data and presents summary statistics for the main variables of interest. Section III outlines the specification of the analytic model, and Section IV presents the empirical results from those analyses. Lastly, Section V offers concluding remarks.

I. Review of Item and Income Nonresponse

Ferber (1966) examined item nonresponse in a mail survey on annual consumption habits for a nationwide sample of 40,000 US adults. A linear regression model was used to identify variables that had a statistically significant effect on the number of questions a respondent omitted. OLS results showed that item nonresponse varied significantly across three main respondent characteristics. First, age was found to have a positive correlation with the number of questions a respondent omitted. In contrast, years of education was shown to negatively affect the number of questions a respondent omitted. Finally, results showed that on average, females omitted a greater number of questions than males. All of these findings were found to be statistically significant at the 1% level. Based on these results, Ferber concluded that the item nonresponse rate was the lowest for young, highly educated males.

These results served as the basis for future studies of item nonresponse in national survey data (cf. Houston and Ford 1976; Burchell and Marsh 1992; Craig and McCann 1978). The results of these later studies supported Ferber's findings, and expanded the study of item nonresponse to explore factors such as survey length and question form. Additional research has been conducted on item nonresponse in nationwide samples within European countries. Sousa-Poza and Henneberger (2000) showed that item nonresponse rates are higher for respondents that own a house and for respondents that are self-employed in Switzerland. Additionally, their results lend further support to the findings of Ferber (1966), and suggest that similar cultural influences of item nonresponse exist across Western countries.

In recent years the study of item nonresponse has focused on nonresponse for questions asking for potentially sensitive information (cf. Kupek 1998). Existing research has found that item nonresponse for questions asking respondents for their personal income is one of the largest sources of item nonresponse (cf. Sheth, LeClaire, and Wachspress 1980). Gronhaug et al. (1988) showed that the odds of income nonresponse are greater for older respondents with lower levels of education. However, the results did not find a significant relationship between a respondent's gender or race and the odds of income nonresponse. Other predictors of income nonresponse, such as household size, were explored by Riphahn and Serfling (2005) in a national sample of German adults. Their results demonstrated that respondents living in larger households have significantly lower income nonresponse rates. Schräpler (2006) found evidence to support the findings of Gronhaug et al. (1988) in a nationwide sample of adults in Britain. The results of Riphahn and Serfling (2005) and Schräpler (2006) suggest that similar culture influences of income nonresponse also exist within Western countries.

By combining the results of the existing research over the past forty years, several key variables can be identified as strong predictors of both item and income nonresponse including age, gender, and education (Ferber 1966; Gronhaug et al. 1988). Additional studies on item and income nonresponse have also used family size, whether a respondent is self-employed, and whether a respondent owns a home to characterize nonrespondents (Sousa-Poza and Henneberger 2000 and Riphahn and Serfling 2005). This study will attempt to extrapolate the results of the existing literature on item and income nonresponse to non-Western countries, by applying these commonly used explanatory variables to a sample of Japanese adults. Since there is no existing evidence to suggest that cultural differences affect survey nonresponse, we expect to observe similar effects on income nonresponse for each of these explanatory variables in our sample of Japanese elderly as was found in other countries.

II. Data

The data for this study come from the National Survey of the Japanese Elderly (NSJE). The NSJE is a joint project between the University of Michigan and Tokyo Metropolitan Institute of Gerontology designed to create a panel dataset to be used in analyses of aging in Japan. The data used by this study come from the second wave of the NSJE, collected in November 1990. The sample consists of 2,163 non-institutionalized Japanese adults aged 60 and older. A two-staged stratified sampling design was used to collect the data. In the first stage, census enumeration districts were stratified by geographic region and city size, and 192 sampling units were selected. In the second stage, 15 potential respondents were randomly selected from within each primary sampling unit using an equal interval method. Personal information for each potential respondent, such as name, address, sex, and birth date, was obtained from the Japanese Resident Registries. A personal interview was administered in which detailed information about

a respondent's family structure, physical and mental health, life satisfaction, and financial status was recorded. By focusing on the elderly population, this study will be able to examine the relative effect of an increase in age on survey nonresponse for adults over 60, providing a greater understanding to the commonly accepted positive relationship between age and survey nonresponse.

One problem with the data, however, is the presence of unit nonresponse bias. Interviewers were given a reserve, or backup, list of potential respondents to use in the event that a sampled respondent could not be reached due to death, refusal, or their address was not found or had changed. While this method of substitution is useful for collecting a desired sample size, a bias in the data arises when the characteristics of the respondents substituted into the sample differ from those of the nonrespondents. Jay et al. (1993) studied the characteristics of nonrespondents in the first wave of data collection (November 1987), and found that gender, age, and geographic characteristics had a statistically significant impact on the probability that a potential respondent would not respond to the survey. Based on these results, the presence of unit nonresponse bias in the NSJE data should be recognized.

The goal of this study is to characterize item nonresponse for a question about personal income. Unfortunately, the NSJE survey form does not include a question about a respondent's individual income, but instead asks for the combined income of the respondent and his or her spouse. As a result, this study will analyze the question about the combined income of the respondent and his or her spouse as if it were a question about the respondent's personal income. The subtle difference in the question's wording should be noted, however, and will be taken into consideration when the logistic regression models are being specified.

As noted above, there are two types of item nonrespondents to the question about personal income: refusals (called income nonrespondents) and respondents who were either physically or mentally unable to answer the survey and required a proxy to complete the survey form. Existing research has found that item nonresponse rates are the highest for questions that ask for potentially sensitive information, such as personal income. For example, Sheth et al. (1980) found that in a mail survey the item nonresponse rate for a question about income was roughly 9.1%, while the item nonresponse rate for a question about the respondent's age was only 1.9%. Additionally, Gronhaug et al. (1988) observed that the item nonresponse rate was roughly 14% for a question about personal income, while the item nonresponse rate was only 0.36% for a question about the respondent's age. Table (1) presents the item nonresponse rates and the survey questions for the main variables of interest in this study. The item nonresponse rate for the personal income question is 3.63%, compared to 1.14% for the question about the respondent's present health status and 0.70% for the question about homeownership. While the income nonresponse rate is considerably lower in this dataset than those found by Sheth et al. (1980) and Gronhaug et al. (1988), there are a total of 73 nonrespondents for this analysis. Furthermore, the item nonresponse rates are lower for this survey because the data were collected through personal interviews instead of mail surveys, which was the survey method used by Sheth et al. (1980) and Gronhaug et al. (1988) (cf. Riphahn and Serfling 2005). Item nonresponse rates may also be lower for samples of the elderly because the non-working population likely has a lower opportunity cost of time and would be more inclined to complete the entire survey.

Other household surveys within East Asian countries have observed even lower item nonresponse rates. In the Gansu Survey of Children and Families, a survey conducted in a northwestern province of China, the item nonresponse rate for a question about personal income

was zero. The income nonresponse rate was also found to be zero in the Chinese Healthy Longevity Survey, which was administered to a sample of Chinese adults aged 60 and older (cf. Zeng et al. 2002). These variations support the conclusion that cultural differences lead to lower item nonresponse rates in East Asian countries compared to the observed item nonresponse rates in Western countries.

Table (2) presents summary statistics for the entire sample of 1,957 adults, as well as for the subsamples of respondents and nonrespondents to the personal income question. The average age of the entire sample is 68.72 years, and the average number of years of schooling for the entire sample is approximately 8.9 years. Roughly 26.19% of respondents reported a personal income of less that ¥1,200,000, and 37.86% of respondents reported a personal income of between ¥1,200,000 and ¥3,000,000. The subsample of nonrespondents is 33.85% male, and 68.75% of nonrespondents own a house. Both of these statistics are lower than the percentages found in the subsample of respondents, where roughly 45.10% of respondents are male and 84.70% of respondents own a house. A more thorough analysis of the characteristics of nonrespondents to the question on personal income will be presented in the following sections.

In the event that a respondent was physically or mentally unable to answer the survey, a proxy survey was administered in which data for the respondent were collected from a family member or other household member. The use of a proxy survey allows researchers to collect at least some data for all respondents; however, the proxy survey was shorter and only collected information about a respondent's family structure and health. As a result, the use of abbreviated proxy surveys causes a greater item nonresponse bias on questions that are not included in the proxy survey. Summary statistics for the 154 respondents who used a proxy as well as for the entire sample of 2,160 adults are presented in Table (3). The average number of years of schooling for the entire sample is 8.82 years, and the average number of family members for the entire sample is approximately 3.6 persons. The average age for the subsample of respondents who used a proxy is 72.88 years old, and the subsample is roughly 57.14% male. Both of these statistics are greater than those for the entire sample, where the average age is only 69.01 years old and the entire sample is 45.4% male. These results indicate that the respondents who needed a proxy are not evenly distributed across the sample, and illustrate that a potential item nonresponse bias exists due to the use of the proxy surveys. The significance of these differences will be more thoroughly analyzed in the following sections.

One of the important explanatory variables for this analysis is a measure of a respondent's health status. The question in the respondent survey form asks respondents to rank their present health as either "excellent," "very good," "good," "fair," or "poor." The same question in the proxy survey form asks proxies to rank the respondent's health as either "perfectly healthy," "fairly healthy," "okay," "not very healthy," or "not at all healthy." The responses from these two survey forms are combined in the data under the labels found in the respondent survey. Since the different survey forms present two different sets of possible answers, the combined data may include considerable measurement error. Unfortunately, the difference in survey forms cannot be corrected, so this potential bias must be factored into the analysis.

III. Empirical Specification

The following logistic regression model was used to describe the variation in the probability of a nonresponse to the question of personal income:

$$\log \frac{\Pr(INR=1)}{1-\Pr(INR=1)} = \beta_0 + \beta_1 MALE + \beta_2 AGE + \beta_3 EDUC \qquad \text{Eq. (1)}$$

$$+ \beta_4 FAMSIZE + \beta_5 MARRIED + \beta_6 HEAD + \beta_7 HOUSE + \beta_8 SELFEMPLOY + \varepsilon$$

MALE is a dummy variable for male survey respondents; AGE is a continuous variable measuring a respondent's age in years; EDUC is a continuous variable measuring a respondent's total years of education; FAMSIZE is a continuous variable measuring the number of people living in a respondent's household; MARRIED is a dummy variable for respondents who are married; HEAD is a dummy variable for respondents who are the head of their household; HOUSE is a dummy variable for respondents who reported owning the house he or she lives in; SELFEMPLOY is a dummy variable for respondents who reported being self-employed. The event of income nonresponse is abbreviated INR, so $\Pr(INR=1)$ represents the probability that a respondent does not respond to the question about personal income. Since a question asking for the combined income of a respondent and his or her spouse is used in this analysis, the MARRIED and HEAD dummy variables were included in the logistic model in order to control for additional factors that might be associated with a nonresponse to a question asking about a spouse's income. To model variations in the odds that a respondent needs a proxy, the following logistic regression is used:

$$\log \frac{\Pr(PROXY=1)}{1-\Pr(PROXY=1)} = \beta_0 + \beta_1 MALE + \beta_2 AGE + \beta_3 EXHEALTH \qquad \text{Eq. (2)}$$

$$+ \beta_4 VGHEALTH + \beta_5 GOODHEALTH + \beta_6 FAIRHEALTH + \beta_7 EDUC + \beta_8 FAMSIZE + \varepsilon$$

EXHEALTH is a dummy variable for respondents who were recorded as having "excellent" health; VGHEALTH is a dummy variable for respondents who were recorded as having "very good" health; GOODHEALTH is a dummy variable for respondents who were recorded has having "good" health; FAIRHEALTH is a dummy variable for respondents who were recorded as having "fair" health. The coefficients for the four health dummy variables included in the model can be interpreted as the odds of a respondent needing a proxy relative to respondents who were recorded as having "poor" health. The event that a respondent requires a proxy is abbreviated PROXY, so $\Pr(PROXY=1)$ represents the probability that a respondent requires a proxy to answer the survey.

Since there are few existing studies on the characteristics of respondents who need a proxy, there is little theoretical support for the specification of this logistic regression model. Instead, explanatory variables were chosen based on intuitive assumptions about the relationships between respondent characteristics and the odds of needing a proxy. *A priori*, it is assumed that males and older respondents will have greater odds of needing a proxy because older males are at a greater risk of developing physical disabilities. Similarly, it is assumed that health, education, and family size are negatively correlated with the odds of needing a proxy. A negative coefficient is expected for education because higher levels of education reduce the likelihood of developing severe disabilities. Additionally, larger families are more adept at caring for one another which reduces the severity of any existing disabilities. Finally, lower overall

levels of health often lead to, or accompany, the physical or mental disabilities that require the use of a proxy.

IV. Results

The logistic models specified by Equations (1) and (2) are estimated using maximum-likelihood estimation and the results are presented in Tables (4) and (5), respectively. In both tables, odds ratios and standard errors are presented for each explanatory variable.

Table 4. Logistic Regression Coefficients for Income Nonresponse

Predictor	Nonresponse	Either Nonresponse or Don't Know
Age	0.983	1.016
	(-0.82)	(-1.37)
Education	1.012	0.978
	(0.22)	(-0.74)
Family Size	1.029	1.068
	(0.44)	(1.61)
Married	1.367	0.895
	(0.99)	(-0.60)
Head	0.964	0.640**
	(-0.08)	(-2.09)
Gender	0.511*	0.519***
	(-2.33)	(-2.72)
House	0.362***	0.803
	(-3.59)	(-1.06)
Self-Employed	1.707**	1.320
	(1.69)	(1.54)
Pseudo R^2	0.033	0.043
Observations	1957	1957

Odds ratios reported
Z-scores in parentheses
Heteroskedasticity robust standard errors are used
*** significant at 1%; ** significant at 5%; * significant at 10%

Table (4) shows that many of the common explanatory variables used to describe both item and income nonresponse in the literature are not statistically significant for this sample. Respondent's age, which was found to be one of the most important predictors of item and income nonresponse in previous research (cf. Ferber 1966; Gronhaug et al. 1988), is found to be statistically insignificant. Both years of education and the number of family members are also found to be statistically insignificant. Additionally, the dummy variables for respondents who are married and for respondents who are the head of the household are also not statistically significant.

Even with these insignificant variables, the logistic regression model presented in Equation (1) identifies several important predictors for income nonresponse within our sample of elderly Japanese adults. First, gender is a statistically significant predictor of income nonresponse. Table (4) confirms that the odds of a nonresponse to the personal income question are 48.9% lower for males, which is statistically significant at the 10% level. This supports the findings of both Ferber (1966) and Riphahn and Serfling (2005) that women have higher levels of income nonresponse than males. Second, the results show that whether a respondent owns a home and whether a respondent is self-employed are statistically significant predictors of income nonresponse at the 1% and 5% levels, respectively. The odds of income nonresponse are 63.8% lower for respondents who own a home and 70.7% greater for respondents who are self-employed. Sousa-Poza and Henneberger (2000) also found that the probability of item nonresponse is higher among self-employed respondents. However, the lower odds ratio for homeowners is contrary to Sousa-Poza and Henneberger (2000) who found that the odds of item nonresponse are greater for homeowners.

The logistic model specified in Equation (1) is also estimated for the combined subsample of the 73 nonrespondents and the 189 respondents who answered "don't know" to the question about personal income. Theoretically, a "don't know" response is intended to capture the respondents who do not remember or never knew his or her personal income, while nonrespondents are only the respondents who do not feel comfortable answering the question (Shoemaker, Eichholz, and Skewes 2002). Existing research has found that the loss of information from "don't know" responses can be just as significant as the loss of information from nonresponses (cf. Shoemaker, Eichholz, and Skewes 2002 or DeMaio 1980). The results of this analysis are presented in Table (4), and show that homeownership, being self-employed, and the number of family members are all statistically insignificant. The results do demonstrate, however, that the odds of not answering the personal income question are 36% lower for heads of household, statistically significant at the 5% level. The results also illustrate that the odds of not answering the personal income question are 48.1% lower for males than females, statistically significant at the 1% level.

These findings come from the logistic regression model presented in Equation (1), which describes the variation in the probability of income nonresponse. Since item nonresponse in the NSJE is almost negligible for all other questions, a robustness check within the dataset used by this study could not be performed. It is important that tests of robustness be made in future research to confirm the significance of these results.

The results of maximum-likelihood estimation for the logistic regression model illustrated in Equation (2) are presented in Table (5). Almost all of the explanatory variables are statistically significant predictors of the odds that a respondent will require a proxy. First, the odds of needing a proxy are 68.3% higher for males than for females, statistically significant at the 1% level. Second, the results confirm that a one year increase in a respondent's age results in a 5.2% increase in the odds of a respondent needing a proxy, also statistically significant at the 1% level. Additionally, the odds of needing a proxy decrease for higher levels of overall health. Table (5) shows that the odds of needing a proxy are 94.3% lower for respondents with "excellent" health, while the odds of needing a proxy are only 93.6% lower for respondents who reported having "very good" health relative to respondents with "poor" health. This trend continues with the odds of needing a proxy for respondents who reported having "good" health at 91.8% and the odds for respondents who reported having "fair" health at 77.5%, relative to respondents with "poor" health. All of the odds ratios for the health status question are found to

be statistically significant at the 1% level. Finally, the results illustrate that an additional member in the household increases the odds of needing a proxy by 11.5%, which is statistically significant at the 5% level.

Table 5. Logistic Regression Coefficients For Proxy Survey

Predictor	Entire Sample	Male	Female
Gender	1.683***		
	(2.85)		
Age	0.520***	0.451***	0.627***
	(-3.59)	(-3.37)	(-3.62)
Excellent Health	0.057***	0.049***	0.065***
	(-9.00)	(-7.23)	(-5.34)
Very Good Health	0.064***	0.063***	0.061***
	(-8.72)	(-6.35)	(-6.02)
Good Health	0.082***	0.070***	0.093***
	(-8.66)	(-6.44)	(-5.66)
Fair Health	0.225***	0.187***	0.265***
	(-5.12)	(-4.12)	(-3.11)
Education	0.958	1.010	0.883**
	(-1.15)	(0.23)	(-1.92)
Family Size	1.115**	1.157**	1.065
	(2.40)	(2.41)	(0.93)
Pseudo R^2	0.165	0.176	0.144
Observations	2160	979	1181

Odds ratios reported
Z-scores in parentheses
Heteroskedasticity robust standard errors are used
*** significant at 1%; ** significant at 5%; * significant at 10%

It was anticipated that there would be a significant relationship between a respondent's education and the odds of needing a proxy. However, the initial logistic regression model does not detect such a relationship. Existing literature suggests that results can vary across gender (cf. Brown and Tierney 2007), so the logistic regression model is estimated again for the two subsamples consisting of just males and females; the results are presented in Table (5). The results of this secondary analysis find that a respondent's education is statistically significant at the 5% level for the subsample of females. For females, a one year increase in total education results in a 11.7% decrease in the odds of needing a proxy. Additionally, the results find that family size does not have a significant effect on the odds of needing a proxy for females, but does have a significant effect for males; an additional member in the household increases the odds of a male respondent needing a proxy by 15.7%.

V. Discussion and Conclusion

The goal of this study is to characterize the nonrespondents to the question about personal income in the NSJE dataset. Nonresponses to the personal income question arise either when a respondent refuses to respond or when a respondent is physically or mentally unable to respond. To characterize the nonrespondents in each of these groups, two separate logistic regression models were used. The coefficients in each of these models were obtained through maximum-likelihood estimation.

The first logistic regression model describes the variation in the probability of income nonresponse. The results of this model found that income nonresponse is most likely for respondents who are female, do not own a home, and are self-employed. Since Japanese women during the middle of the 20th Century were predominantly homemakers and responsible for their children's education (cf. Iwao 1993; Tomida and Daniels 2005), one possible explanation for the higher odds of income nonresponse for females is that the females in this dataset are less likely to have earned an income and feel uncomfortable providing information about their husband's income. The lower odds of income nonresponse for homeowners may be due to the fact that homeowners feel more secure with their financial status. As a result, a homeowner may be more willing to divulge information about his or her personal income. One possible explanation for the higher odds of income nonresponse for respondents who are self-employed is that self-employed individuals are by nature more private and less willing to reveal personal information; their choice of occupation is driven in part by this same desire for greater privacy. Another possibility is that self-employed respondents are more apprehensive about providing income information that might contradict information contained in their tax reports to the government. Regardless of the reasons behind these observed relationships, this logistic regression model identified three statistically significant predictors of income nonresponse among Japanese elderly: gender, homeownership, and being self-employed.

The results also show that several common explanatory variables for income nonresponse were statistically insignificant in this population, which is contrary to the existing research on income nonresponse in other populations (cf. Gronhaug et al. 1988; Riphahn and Serfling 2005). While most of the existing research on item and income nonresponse has been conducted within the United States and other Western countries, this study uses data from Japan. The statistical insignificance of explanatory variables such as age, education, and family size suggest that there exist significant culture differences between Western and non-Western countries which lead to different patterns of item nonresponse. Furthermore, the result that age is not a significant predictor of income nonresponse for adults aged 60 and above raises questions about the way age is modeled in other studies. In the existing research, the elderly are rarely broken out as a separate group (cf. Jay 1993, DeMaio 1980, or Weaver, Holmes, and Glenn 1975). The data employed by this study allow for more detailed analyses of the patterns of nonresponse within the elderly. The lack of statistical significance for age, education, and family size indicates that different factors affect income nonresponse during old age.

The logistic regression model presented in Equation (1) is also estimated for the combined sample of nonrespondents and respondents who answered "don't know" to the personal income question. The results show that the odds of not answering the income question are statistically lower for respondents who are the head of the household. This can be explained by the fact that the household head is usually responsible for the household's budget, and as a result, would be more likely to know their personal income. The results also show that females are more likely to not provide an answer to the personal income question than males. These

results are contrary to what we might expect, since women typically play the role of the household head in Japan (cf. Iwao 1993; Tomida and Daniels 2005). These results suggest that additional cultural factors might influence a Japanese woman's ability or willingness to answer questions about personal income. Overall, these findings serve to describe additional characteristics of the nonrespondents to the personal income question.

The second logistic regression model describes the variation in the odds of a respondent needing a proxy. The results show that older males with lower levels of health and larger families are the most likely to need a proxy. These results support our *a priori* hypotheses about the direction of each relationship. The only contradictory result is the significant positive relationship between family size and the odds of needing a proxy. *A priori,* it was assumed that a larger family would be more able to take care of a respondent, preventing him or her from developing a severe physical or mental illness or reducing the severity of any existing disabilities. However, this result suggests that a larger family might in fact limit the resources available to a respondent, increasing the likelihood that the respondent develops a severe disability. It is also found that education does not have a statistically significant effect on the odds of needing a proxy. While it was assumed that higher education levels would reduce the likelihood of a respondent developing mental disabilities, the results indicate that this effect does not exist within the elderly population of Japan. When the logistic regression is estimated for the subsample containing only females, a statistically significant negative effect is observed for education on the odds of needing a proxy. This suggests that different factors influence mental deterioration for males versus females. Regardless of the reasons behind these observed relationships, the results of this logistic regression reveal additional characteristics of the nonrespondents to the personal income question.

Based on the results from these logistic regression models, it is clear that the nonrespondents to the personal income question are not randomly distributed within the sample population. Thus, in addition to the known tendency of respondents to exaggerate their income, there exists a bias in the income data from item nonresponse. Researchers using income data from the NSJE as either a response or predictor variable should acknowledge this bias and make appropriate corrections to their analyses. The results of these logistic regression models can be used by researchers to derive weighting adjustments for the NSJE data. Additionally, these results will enable researchers to identify procedural problems and aid in the design and prevention of nonresponse in future studies. Based on these results, researchers will be able to identify individuals at high risk for nonresponse to a personal income question, which can help researchers to modify interviewer training and field procedures to increase response rates among these high risk groups. Finally, by characterizing the nonrespondents to the personal income questions researchers will be able to estimate the sign of the nonresponse bias present in this dataset.

In addition to the impact of these results on economic research incorporating the NSJE dataset, this study expands the existing knowledge on item and income nonresponse. Since many of the common explanatory variables found to have a significant effect on item and income nonresponse in Western countries were not found to significantly effect income nonresponse in this sample of Japanese adults, and the overall percent of item and income nonresponse was lower in this sample than was observed in national samples collected in Western countries, this study indicates that cultural differences have a significant effect on item and income nonresponse. Additionally, the insignificance of the effect of age on income nonresponse for this elderly sample, suggests that the commonly accepted positive relationship between age and survey

nonresponse diminishes for older respondents. While these developments were not the primary research focus of this study, they are important and unique findings within the existing body of work on item and income nonresponse and should be more closely examined in future research.

VI. REFERENCES

Brown, P.H. and Tierney, B. 2007. "Religion and Subjective Well-Being Among the Elderly in China." *Working Paper, Goldfarb Center for Public Affairs and Civic Engagement*.

Buchell, B. and Marsh, C. 1992. "The Effect of Questionnaire Length On Survey Responses." *Quality & Quantity* 26: 3, 233-244.

Craig, C. and McCann, J. 1978. "Item Nonresponse in Mail Surveys: Extent and Correlates." *Journal of Marketing Research* 15: 2, 285-289.

DeMaio, T. 1980. "Refusals: Who, Where, Why?" *Public Opinion Quarterly* 44: 223-33.

Ferber, R. 1966. "Item Nonresponse in a Consumer Survey." *Public Opinion Quarterly* 3: 399-415.

Gronhaug, K., Gilly, M., and Enis, B. 1988. "Exploring Income Nonresponse: A Logit Model Analysis." *Journal of the Market Research Society* 30: 3, 371-78.

Hill, D. and Willis, R. 2001. "Reducing Panel Attrition: A Search for Effective Policy Instruments." *Journal of Human Resources* 36: 3, 416-38.

Horowitz, J. and Manski, C. 1998. "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations." *Journal of Econometrics* 84: 37-58.

Houston, M. and Ford, N. 1976. "Broadening the Scope of Methodological Research on Mail Surveys." *Journal of Marketing Research* 13: 4, 397-403.

Iwao, S. 1993. "The Japanese Woman: Traditional Image and Changing Reality." The Free Press. Macmillan, Inc. New York, NY.

Jay, G. M. 1993. "Patterns of nonresponse in a national survey of elderly Japanese." *Journal of Gerontology* 48: 3, 143-52.

Kupek, E. 1998. "Determinants of Item Nonresponse in a Large National Sex Survey." *Archives of Sexual Behavior* 27: 6, 581-94.

Riphahn, R. and Serfling, O. 2005. "Item Nonresponse on Income and Wealth Questions." *Empirical Economics* 30: 2, 521-38.

Schräpler, J. 2006. "Explaining Income Nonresponse – A Case Study By Means of the British Household Panel Study (BHPS)." *Quality & Quantity* 40, 1013-1036.

Sheth, J., LeClaire, A., and Wachspress, D. 1980. "Impact of Asking Race Information in Mail Surveys." *Journal of Marketing* 44: 1, 67-70.

Shoemaker, P., Eichholz, M., and Skewes, E. 2002. "Item Nonresponse: Distinguishing Between Don't Knows and Refuse." *International Journal of Public Opinion Research* 14: 2, 193-201.

Sousa-Poza, A. and Henneberger, F. 2000. "Wage Data Collection by Telephone Interviews: An Empirical Analysis of the Item Nonresponse Problem and its Implications for the Estimation of Wage Functions." *Swiss Journal of Economics and Statistics* 136: 1, 79-98.

Tomida, H. and Daniels, G. 2005. "Japanese Women: Emerging From Subservience, 1868-1945." Global Oriental; Kent, CT.

Weaver, C., Holmes, S., and Glenn, N. 1975. "Some Characteristics of Inaccessible Respondents in a Telephone Survey." *Journal of Applied Psychology* 60: 260-62.

Zeng, Y, Vaupel, J., Zhenyu, X., Chunyuan, Z., and Yuzhi, L. 2002. "Sociodemographic and Health Profiles of the Oldest Old in China." *Population and Development Review* 28: 2, 251-273.

Table 1. Item Nonresponse By Question

Variable	Survey Question	Nonrespondents N	Potential Respondents N	Percent Item Nonresponse
Income	Which of the following fits you and your spouse's joint income? If you or your spouse are receiving welfare, please include it in your annual income.	73	2009	3.63%
Age	Could you please tell me the date and year of your birth? How old are you?	1	2009	0.05%
Gender	Sex of respondent.	0	2009	0.00%
Present Health Status	On the whole, how is your health at the present time?	23	2009	1.14%
Years of Education	Altogether, how many years did you attend school?	0	2009	0.00%
Number of Family Members	At present, including yourself, how many persons are living with you?	0	2009	0.00%
Married	Are you married at present? Or are you divorced or separated?	0	2009	0.00%
Homeowner	Do you own the house you live in?	14	2009	0.70%
Self-Employed	Are you self-employed?	0	2009	0.00%

Table 2. Sample Characteristics for Income Nonresponse

Variable	Entire Sample (N = 1957)	Income Nonrespondents (n = 65)	Income Respondents (n = 1892)
Age	68.72 (7.238)	68.03 (9.289)	68.74 (7.159)
% Personal Income			
Less than ¥1,200,000			26.19 (0.439)
¥1,200,000 – ¥3,000,000			37.86 (0.485)
¥3,000,000 – ¥5,000,000			15.03 (0.357)
Over 5,000,000 RMB			11.15 (0.315)
% Present Health Status			
Excellent	26.49 (0.441)	23.08 (0.425)	26.61 (0.442)
Very Good	25.03 (0.433)	16.92 (0.378)	25.30 (0.435)
Good	33.18 (0.471)	47.69 (0.503)	32.69 (0.469)
Fair	12.77 (0.334)	9.23 (0.292)	12.89 (0.335)
Poor	1.37 (0.116)	0 (0)	1.41 (0.118)
Years of Education	8.89 (2.564)	8.98 (2.427)	8.87 (2.569)
Number of Family Members	3.57 (1.945)	3.35 (1.807)	3.57 (1.950)
% Male	44.73 (0.497)	33.85 (0.477)	45.10 (0.450)
% Homeowners	84.18 (0.365)	68.75 (0.467)	84.70 (0.360)
% Self-Employed	19.91 (0.399)	26.15 (0.443)	19.70 (0.399)
% Married	65.85 (0.474)	67.12 (0.473)	65.81 (0.474)
% Head of Household	54.31 (0.248)	50.68 (0.503)	54.44 (0.498)

Standard errors in parentheses

Table 3. Sample Characteristics For Proxy Survey

Variable	Entire Sample (N = 2160)	Proxy Survey (n = 154)	Respondent Survey (n = 2006)
Age	69.01 (7.471)	72.88 (9.252)	68.72 (7.234)
% Present Health Status			
Excellent	25.61 (0.436)	12.99 (0.337)	26.58 (0.441)
Very Good	24.23 (0.428)	14.29 (0.351)	24.99 (0.433)
Good	32.50 (0.468)	23.38 (0.424)	33.20 (0.471)
Fair	13.59 (0.342)	24.68 (0.432)	12.74 (0.333)
Poor	3.01 (0.170)	24.68 (0.432)	1.34 (0.115)
% Male	45.40 (0.498)	57.14 (0.496)	44.50 (0.497)
Years of Education	8.82 (2.597)	8.26 (2.963)	8.867 (2.563)
Number of Family Members	3.60 (1.955)	4.05 (2.015)	3.57 (1.947)

Standard errors in parentheses