Title	Residential mobility and panel attrition : using the interview process as identifying instruments
Sub Title	
Author	直井, 道生(Naoi, Michio)
Publisher	Keio Economic Society, Keio University
Publication year	2007
Jtitle	Keio economic studies Vol.44, No.1 (2007.) ,p.37- 47
JaLC DOI	
Abstract	This paper aims to assess the extent of attrition bias in the Keio Household Panel Survey (KHPS) 2004-2006, focusing on household residential mobility. We use the unique dataset of the interview process as identifying instruments for a sample selection model. The results show that sample attrition does lead to statistically significant bias in wave 2, but it seems to be attenuated in the later wave. The results of the bias analysis for each estimated coefficient are also presented.
Notes	
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AA00260492-20070001-0037

KEIO ECONOMIC STUDIES 44(1), 37–47 (2007)

RESIDENTIAL MOBILITY AND PANEL ATTRITION: USING THE INTERVIEW PROCESS AS IDENTIFYING INSTRUMENTS

Michio NAOI

Faculty of Business and Commerce, Keio University, Tokyo, Japan

First version received December 2007; final version accepted January 2008

Abstract: This paper aims to assess the extent of attrition bias in the Keio Household Panel Survey (KHPS) 2004–2006, focusing on household residential mobility. We use the unique dataset of the interview process as identifying instruments for a sample selection model. The results show that sample attrition does lead to statistically significant bias in wave 2, but it seems to be attenuated in the later wave. The results of the bias analysis for each estimated coefficient are also presented.

Keywords: Residential Mobility, Attrition Bias, Sample Selection Model, Interview Process.
JEL classification: C33, C81, R23.

1. INTRODUCTION

Panel data often provides a useful methodological framework for understanding household behavior, which is impossible with cross-sectional information alone. Nonetheless, a typical concern in longitudinal research is that there can be substantial nonrandom attrition which not only undermines the representativeness of the data over time, but also biases the inferences for specific applications.

Previous studies extend the methods of sample selection to account for panel attrition bias (Hausman and Wise, 1979; Baltagi, 2005, chap. 11, and references therein).[1] In applying a sample selection model, the identification of the 'behavioral' coefficients

Acknowledgements. The author is grateful to the 21st Century Center of Excellence Program at Keio University for generously providing us with the data (Keio Household Panel Survey). The author would like to thank the editor and the anonymous referee for their constructive comments. I also thank Miki Seko, Colin McKenzie, Yoshio Higuchi, Kazuo Yamaguchi, Kazuyasu Sakamoto and participants at the 20th Annual Meeting of the Applied Regional Science Conference at Hiroshima University, the 6th Panel Survey Conference at the Institute for Research on Household Economics, and the Japanese Economic Association 2007 Spring Meeting for their helpful comments and suggestions.

E-mail: naoi@2001.jukuin.keio.ac.jp

[1] Alternatively, several studies employ inverse probability weighting (IPW) to account for attrition bias. Some other studies examine the effect of sample attrition by comparing the attributes or behavior of respondents within a single data set (e.g. attriters vs. non-attriters) or across different data sets.

requires an exclusion restriction, i.e., there should be an instrument that affects nonresponse while being independent from the behavior of interest. In practice, however, finding a suitable instrument for unobservable selection is by no means easy in the case of nonresponse. While the situation depends on the specific model under consideration, individual/household characteristics are unlikely to be sources of instruments because most of such characteristics are generally related to individual/household behavior of interest (Fitzgerald et al., 1998).

Alternatively, we use unique information about the interview process, such as contact history and the list of interviewers assigned to each respondent, as identifying instruments which are external to the individual respondents. Since the interview process is closely associated with nonresponse (Hill and Willis, 2001) and, generally, each interviewer is randomly assigned to respondents,[2] this information can be used as a valid instrument for the sample selection model.

The paper examines the attrition bias in the Keio Household Panel Survey (KHPS), 2004–2006. More specifically, it aims to assess the extent of attrition bias for specific empirical example—household residential mobility—using information about the interview process as identifying instruments for a sample selection model. Although respondent mobility is an important source of panel attrition, none of the previous studies have evaluated the impact of attrition in the analysis of residential mobility. Our results demonstrate that attrition leads to statistical bias in wave 2; however, it becomes attenuated in the subsequent wave. We also present the results of the bias analysis for each estimated coefficient to assess the extent of the attrition bias for each coefficient.

The paper is organized as follows. Section 2 briefly introduces the KHPS and summarizes its attrition pattern. Section 3 explains the estimation method and variables. Section 4 summarizes our main empirical results. Section 5 presents the conclusion.

2. SAMPLE ATTRITION IN KHPS

The KHPS, sponsored by the Ministry of Education, Culture, Sports, Science and Technology, is the first comprehensive panel survey of households in Japan, conducted annually by Keio University since 2004. In the following analysis, we use the first three waves of the KHPS, which were conducted in 2004, 2005 and 2006, respectively. In 2004, 13,430 individuals, male and female, aged 20–69 years, were selected by stratified two-stage random sampling as a potential respondent. Out of 13,430 individuals initially approached, 4,005 primary respondents finally participated in the first wave of the survey (response rate = 29.8%). The questionnaire of the KHPS is made of two parts, namely, individual and household related questions. As far as the former is concerned, it contains a wide array of questions with respect to the respondent's demographic characteristics, education and employment activities, among others. If the primary respondent was married at the time of survey, the questionnaire also contains virtually identical

[2] The situation is somewhat different in other surveys. For example, the interviewers of Panel Study of Income Dynamics (PSID) are assigned based on respondent characteristics. In this case, the interview process might be correlated with respondent characteristics, and hence, with the behavior of interest.

questions to be answered by his/her spouse. The standard procedure for the KHPS was to send a pre-survey letter to the respondent and then provide a post-interview payment of 3,000 yen (approximately $25) per household.

By 2006, the KHPS witnessed a sample loss of approximately 28% due to cumulative attrition from its initial 2004 sample. Compared with other longitudinal surveys, the attrition in the KHPS is somewhat heavy; this can possibly be attributed to the fairly long and comprehensive questionnaire used. For example, the cumulative attrition rates for the first three waves are approximately 15% in the Panel Study of Income Dynamics (PSID) and 5.7% in the National Longitudinal Survey of Youth (NLSY). For the European Community Household Panel (ECHP), these figures range from 12.1% (Germany) to 36.5% (Denmark). The heavy attrition in the KHPS suggests that its effect should be examined.[3]

Furthermore, residential mobility is represented as one of the most important factors determining panel attrition. In fact, household residential mobility is strongly and positively related to sample attrition—in wave 2, 36.3% of movers attrited from the survey as compared to a mere 15.9% of non-movers.[4] The same pattern can be observed for wave 3 (the difference between attrition rates are even larger). Basically, sample attrition poses serious problems when it is not independent from the behavior of interest, suggesting that sample attrition is especially important when we try to understand the household residential mobility.

3. EMPIRICAL MODEL AND VARIABLES

Our primary interest is the binary choice of residential mobility, which can be observed only if the respondent remains in the panel. To estimate such a model, we employ the method proposed by Van de Ven and Van Pragg (1981), which extends Heckman's (1979) selection model to the case of a probit model with sample selection. The model of residential mobility is formally given as

$$y_{it}^{*} = x_{it}\beta + \varepsilon_{it}\,, \tag{1}$$

$$y_{it} = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \quad \text{if} \quad y_{it}^{*} \begin{Bmatrix} > \\ \leq \end{Bmatrix} 0\,, \tag{2}$$

where y_{it} takes one if household i moves between wave $t-1$ and t, which is observed if i stays in wave t and x_{it} is the relevant set of explanatory variables. The previous studies on residential mobility suggest that the household life-course events (such as family formation and dissolution) and the housing environment (such as tenure of the dwelling and its location) are the crucial factors for short-distance moves, while the employment reasons predominate for moves of longer distances (Dieleman, 2001; Clark and Lierop, 1986). Based on these empirical findings, x_{it} includes the respondent's age, sex, marital

[3] For an extensive review of attrition problem in the KHPS, see Miyauchi et al. (2006), McKenzie et al. (2007) and Naoi (2007).

[4] Information on residential mobility used here is based on the interviewer's record, which can be obtained even if the respondent is dropped out from the survey. Note that, in the following analysis, we use a self-answered information on mobility which cannot be observed if the respondent is out of the survey.

status, educational attainment, health condition, employment status, changes in household type and composition, housing tenure, and the location of residence.[5]

The survey response equation of the model is

$$s_{it}^{*} = x_{i1}\gamma + z_{i2}\delta + \nu_{it}\,, \tag{3}$$

$$s_{it} = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \quad \text{if} \quad s_{it}^{*} \begin{Bmatrix} > \\ \leq \end{Bmatrix} 0\,, \tag{4}$$

where s_{it} takes one if household i remains in the wave t panel and zero otherwise; x_{i1} is the same set of explanatory variables as in equation (1) but with its value at wave 1; and z_{i2} is the identifying instruments excluded from the main equation (1), as explained in detail later. For the identification of the parameters, we assume that ε_{it} and ν_{it} have unit variances and are joint normally distributed. Setting $\rho = \text{Corr}(\varepsilon_{it}, \nu_{it})$, the null hypothesis of interest is given by $H_0 : \rho = 0$, which indicates that there are no attrition biases. This can be tested either by the Wald or likelihood ratio tests.

The key variable in equation (3) is the identifying instrument (z_{i2}) which has been excluded from equation (1). To construct this variable, we use information about the interview process in wave 2. In wave 2, the supplemental questionnaire of the KHPS provides detailed information to the targeted respondents about (a) the contact history (up to five visits) provided by each interviewer and (b) the complete list of assignments of each interviewer. Using this information, we construct the following four variables for z_{i2}: (1) the month in which the interviewer first contacts the respondent, (2) whether or not the interviewer visits the targeted respondent on weekends (1 if all visits are made on weekdays), (3) the number of respondents (including the targeted respondent) which the assigned interviewer is in charge of, and (4) whether or not the assigned interviewer is the same as the one in wave 1. The first two variables are constructed from the contact history information, while the latter two are from the list of interviewers' assignments.

Given the time constraint on the interview process and that each interviewer must allocate time for the targeted respondents under him/her, the first and third variables capture the interviewer's potential effort towards each targeted respondent. For instance, a slow start to the interview process or handling too many respondents may end up in failure to complete the interview process. The second variable considers the difficulties involved in contacting the targeted respondents. Since it is apparently difficult to contact the targeted respondents on weekdays, particularly for full-time workers, this variable is likely to have a negative impact on the survey responses (Bates, 2004). The last variable, assigning the same interviewer across waves, is included to control for some form of 'familiarity' or 'trust' between the interviewer and respondent, which might reduce the implicit cost of participating the survey (Hill and Willis, 2001).

In the following, we examine the effect of panel attrition on the household residential mobility wave-by-wave. We first estimate the residential mobility function using the cross-section of the wave 2 panel by considering the sample attrition in wave 2 (Model

[5] Since housing tenure and residential region are likely to be determined simultaneously with household mobility (Boehm, 1981), we use one-year lagged values for these two variables.

[1]). Based on this result, we then estimate the mobility function in wave 3 by taking into consideration of the cumulative attrition up to wave 3 (i.e. attrition in either wave 2 or 3) (Model [2]). This enables us to evaluate the evolution of the effect of nonrandom attrition.

4. EMPIRICAL RESULTS

The results of the joint estimation of equations (1) and (3) are provided in Tables 1 and 2. The summary statistics are presented in Table 3.

Table 1 presents the estimated coefficients of the various predictors on the wave 2 and wave 3 participations, respectively. For the explanatory variables, we use respondent/household characteristics at wave 1, and the interview process at wave 2.[6]

First, the result of Wald test demonstrates that the null hypothesis of no attrition bias, i.e. $\rho = 0$, is rejected in both models. However, the likelihood ratio test yields a somewhat different result. While the null of $\rho = 0$ is also rejected in wave 2, this is not so in the wave 3 model. This seems to be consistent with Nawata and McAleer's (2001) results about the finite sample properties of the Wald test in models with sample selection biases. Their Monte Carlo experiments show that the performance of the Wald test is quite poor in finite samples, rejecting the true null hypothesis too frequently. This suggests that the likelihood ratio test is preferable when testing $\rho = 0$.[7] Following this, we conclude that the household residential mobility function is biased at least in the wave 2 panel; however, such a bias seems to fade over time. Such regression-to-the-mean effects are also observed in the PSID (Fitzgerald et al., 1998; Becketti et al., 1988).

All our key variables of the interview process (z_{i2}) have coefficients with appropriate signs. For example, a slow start to the interview process or handling too many respondents is likely to result in failure, weekday visits are negatively associated with response probability, and assigning the same interviewer across waves significantly enhances the survey responses. In the KHPS, given that the observable characteristics of individual respondents explain only a negligible portion of the attrition, the results indicate that the interview process might be an additional and a particularly important source of information explaining nonrandom attrition. Although most of the respondent/household characteristics are insignificant, several variables are comparable to those in the previous studies. The results show that the respondent's age has an inverse U-shaped effect on response probabilities—it is the highest for the middle-aged respondents and the

[6] As can be inferred from equation (1), observations with missing values in wave 2 explanatory variables are dropped from Model [2], while those with missing values in wave 3 variables are dropped from Model [3]. Since item response rates (among those not attrited) are higher in wave 3 than in wave 2, the number of observations for wave 3 participation equation (Model [2]: 3,626) is shown to be slightly larger than that for wave 2 equation (Model [1]: 3,592).

[7] It should be noted, however, that their results are based on a standard sample selection model with linear regression equation, which is not exactly same as our model. However, given the similar structure of the model, their argument on the primary reason for the poor performance of the Wald test—the log-likelihood function is almost flat in the neighborhood of $\rho = 0$—seems to be applicable in this case.

Table 1. Estimation Results for Survey Response

Model	[1]		[2]	
	2005 (wave 2)		2006 (wave 3)	
Survey Response	Probit with Sample Selection		Probit with Sample Selection	
(= 1 if R responds wave t panel)	Coef.	(S.E.)	Coef.	(S.E.)
Respondent Characteristics				
Age				
20–29	—		—	
30–39	0.1341	(0.0918)	0.2069	(0.0813)*
40–49	0.1992	(0.0984)*	0.1945	(0.0864)*
50–59	0.0587	(0.0973)	0.1003	(0.0871)
60 and above	0.0861	(0.1052)	0.1190	(0.0939)
Sex (= 1 if R is female)	0.0617	(0.0603)	0.0298	(0.0534)
Married (= 1 if R is married)	0.0899	(0.0861)	0.1052	(0.0766)
No child (= 1 if R has no child(ren))	−0.1098	(0.0766)	−0.1130	(0.0688)
Health condition (1: Good −5: Bad)	−0.0213	(0.0243)	−0.0286	(0.0219)
Education				
Junior high school	−0.1389	(0.0854)	−0.0892	(0.0772)
High school	—		—	
Junior college	0.2813	(0.0895)**	0.2095	(0.0765)**
4-year college+	0.1159	(0.0704)	0.0971	(0.0619)
Other professional school	0.0779	(0.1318)	0.1154	(0.1163)
Full-time worker (=1 if R works full-time)	−0.0820	(0.0710)	−0.0308	(0.0633)
Retired (=1 if R is retired)	0.3809	(0.1505)*	0.2008	(0.1249)
Household Characteristics				
Changes in household (HH) type and composition				
# of HH members				
Unchanged	—		—	
Increased	0.2322	(0.1834)	0.0560	(0.1509)
Decreased	−0.0343	(0.1029)	−0.0669	(0.0914)
R is newly-wed	0.5401	(0.3736)	0.3355	(0.2730)
R has newborn children	−0.3128	(0.2367)	−0.0655	(0.2020)
R owns house	−0.0351	(0.0646)	−0.0306	(0.0583)
Interview Process Characteristics (wave 2)				
Month of first visit	−0.2362	(0.0507)**	−0.0951	(0.0440)*
All visits on weekdays	−0.1678	(0.0576)**	−0.0907	(0.0515)$^{+}$
# of interviewees in charge	−0.0072	(0.0049)	−0.0099	(0.0043)*
Same interviewer	0.2434	(0.0701)**	0.2178	(0.0645)**
Constant	1.3940	(0.2278)**	0.8234	(0.2012)**
ρ	0.7468	(0.2978)*	−0.6672	(0.3087)*
Likelihood Ratio Test (H_0: $\rho = 0$)	7.45**		1.70	
Log likelihood	−1984.471		−2374.923	
Number of observations	3,592		3,626	
Number of censored observations	601		995	

Notes: **, *, and $^{+}$ indicate that the estimated coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively. A set of dummy variables for regions and firm sizes is also controlled but is omitted from the results. Wave 1 values are used for all explanatory variables except for the interview process characteristics. The corresponding estimation results for residential mobility are shown in Table 2. R denotes respondent. Dependent variables: Model [1]: 1 if R responds in the 2005 survey, and 0 otherwise. Model [2]: 1 if R responds in both the 2005 and 2006 surveys, and 0 otherwise.

lowest for younger and older respondents. Further, at least in wave 2, the survey responses are significantly and positively related to the respondents living in rural areas or those without jobs.

Table 2. Estimation Results for Household Residential Mobility

Model	[1] 2005 (wave 2)				[2] 2006 (wave 3)			
Residential Mobility	Baseline Model		Probit with Sample Selection		Baseline Model		Probit with Sample Selection	
(= 1 if household (HH) moved from $t-1$)	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)	Coef.	(S.E.)
Respondent Characteristics								
Age								
20–29		—		—		—		—
30–39	−0.0250	(0.1292)	−0.0373	(0.1266)	−0.2520	(0.1596)	−0.3468	(0.1451)*
40–49	−0.4884	(0.1544)**	−0.4144	(0.1514)**	−0.2904	(0.1721)+	−0.3542	(0.1554)*
50–59	−0.6388	(0.1711)**	−0.5992	(0.1694)**	−0.6958	(0.2038)**	−0.7038	(0.2141)**
60 and above	−0.3967	(0.1874)*	−0.5492	(0.2004)**	−0.5794	(0.2225)**	−0.6072	(0.2186)**
Sex (= 1 if R is female)	0.1329	(0.1087)	0.1415	(0.1100)	−0.1485	(0.1233)	−0.1167	(0.1137)
Married (= 1 if R is married)	−0.2080	(0.1630)	−0.2048	(0.1670)	0.0792	(0.1771)	0.0315	(0.1618)
No child (= 1 if R has no child(ren))	−0.2197	(0.1570)	−0.2018	(0.1619)	0.1065	(0.1614)	0.2147	(0.1436)
Health condition (1: Good – 5: Bad)	−0.0216	(0.0456)	−0.0182	(0.0454)	−0.0295	(0.0541)	−0.0107	(0.0490)
Education								
Junior high school	−0.2371	(0.1970)	−0.2297	(0.2069)	−0.3757	(0.2499)	−0.2252	(0.2317)
High school		—		—		—		—
Junior college	−0.1096	(0.1419)	−0.0782	(0.1442)	0.0076	(0.1525)	−0.0865	(0.1504)
4-year college+	0.0705	(0.1158)	0.0784	(0.1169)	0.0111	(0.1257)	−0.0076	(0.1172)
Other professional school	−0.0210	(0.2094)	−0.0937	(0.2176)	−0.2749	(0.2858)	−0.3669	(0.2841)
Full-time worker (= 1 if R works full-time)	0.2575	(0.1210)*	0.2965	(0.1227)*	−0.0910	(0.1356)	−0.0412	(0.1228)
Retired (= 1 if R is retired)	−0.1224	(0.3433)	−0.2424	(0.4410)	0.2278	(0.3089)	0.2624	(0.2858)
Household Characteristics								
Changes in HH type and composition								
# of HH members								
Unchanged		—		—		—		—
Increased	0.5230	(0.2255)*	0.3843	(0.2365)	0.6590	(0.2502)**	0.5695	(0.2633)*
Decreased	0.4916	(0.1632)**	0.4253	(0.1670)*	0.1444	(0.2165)	0.0690	(0.1969)
R is newly-wed	1.6637	(0.3081)**	1.5848	(0.3125)**	2.1682	(0.3314)**	1.8166	(0.4510)**
R has newborn children	−0.6909	(0.3108)*	−0.5335	(0.3211)+	−0.2053	(0.3557)	−0.1125	(0.3266)
R owns house	−1.0507	(0.0982)**	−1.0681	(0.1006)**	−0.7825	(0.1076)**	−0.6546	(0.1874)**
Constant	−0.6424	(0.3582)+	−0.6490	(0.3565)+	−1.3729	(0.4616)**	−0.8834	(0.5735)
Log likelihood	−475.154		−1984.471		−363.122		−2374.923	
Number of observation	2,991		3,592		2,631		3,626	
Number of censored observation	—		601		—		995	

Notes: **, *, and + indicates that the estimated coefficient is significant at the 0.01, 0.05, and 0.10 levels, respectively. A set of dummy variables for regions and firm sizes is also controlled but is omitted from the results. The variables for the place of residence and housing tenure are one-year lagged. The corresponding estimation results for the survey response are shown in Table 1. R denotes respondent.

Table 3. Descriptive Statistics

	2004 (wave 1)			2005 (wave 2)			2006 (wave 3)		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Residential mobility									
(= 1 if household (HH) moved from $t-1$)[1)]	—			0.048	0	1	0.035	0	1
Survey response									
(= 1 if R responds wave t panel)[1)]	—			0.833	0	1	0.726	0	1
Respondent Characteristics									
Age	46.372	20	70	47.402	21	71	48.439	22	72
Sex (= 1 if R is female)	0.495	0	1	0.497	0	1	0.491	0	1
Married (= 1 if R is married)	0.736	0	1	0.750	0	1	0.759	0	1
No child (= 1 if R has no children)	0.398	0	1	0.384	0	1	0.384	0	1
Health condition (1: Good – 5: Bad)	1.955	1	5	2.335	1	5	2.344	1	5
Education									
Junior high school	0.116	0	1	0.110	0	1	0.109	0	1
High school	0.490	0	1	0.485	0	1	0.481	0	1
Junior college	0.124	0	1	0.132	0	1	0.132	0	1
4-year college and above	0.226	0	1	0.230	0	1	0.232	0	1
Other professional school	0.043	0	1	0.044	0	1	0.046	0	1
Full-time worker									
(= 1 if R works full-time)	0.378	0	1	0.364	0	1	0.368	0	1
Retired (= 1 if R is retired)	0.044	0	1	0.049	0	1	0.059	0	1
Household Characteristics									
Changes in HH type and composition									
# of HH members									
Increased	0.055	0	1	0.058	0	1	0.043	0	1
Decreased	0.066	0	1	0.068	0	1	0.066	0	1
R is newly-wed	0.009	0	1	0.007	0	1	0.007	0	1
R has newborn children	0.031	0	1	0.026	0	1	0.018	0	1
R owns house	0.765	0	1	0.767	0	1	0.777	0	1
Interviewing Process[2)]									
Month of first visit	—			1.767	1	4	—		
All visits on weekdays	—			0.261	0	1	—		
# of interviewees in charge	—			13.205	1	41	—		
Same interviewer	—			0.853	0	1	—		
Number of observations	3,592			2,991			2,631		

[1)] The number of observations is 3,592 (2005) and 3,626 (2006).
[2)] The number of observations is 3,592.

Table 2 presents the estimated coefficients of respondent/household characteristics on residential mobility. To assess the extent of the attrition bias for each parameter estimate, the baseline model, estimated by a standard probit model using only non-attriters, is presented along with the more appropriate probit model with sample selection. We also estimated the model with the 'interviewer fixed effect' (i.e. including interviewer dummy variables), which qualitatively shows the same result as that in Table 2.

Since we have strong evidence for the attrition bias in wave 2, we limit our discussion primarily to the results of Model [1]. Comparing the results for the probit model with sample selection with those for the uncorrected probit model, we find that the regression coefficients generally show similar signs although there are a number of sizable differences in magnitude and significance. For example, the estimated coefficient

of 'full-time worker' in the uncorrected model (0.2575) is approximately 13% smaller than its corrected value (0.2965). The discrepancies are even greater for the estimated coefficients of household characteristics. Most of the estimated coefficients are found to be 'inflated' in the baseline model, i.e. the absolute values of the estimated coefficients are larger in the uncorrected probit model than in the bias-corrected model. This discrepancy is particularly serious because the estimated coefficients become excessively significant when there is sample attrition bias.

To assess the extent of the attrition bias for each coefficient, separate tests are conducted on a coefficient-by-coefficient basis; this ascertains the difference in the estimated coefficients between the two models (the probit model with sample selection and the uncorrected probit model). Since the variance-covariance matrix for the Hausman test statistic sometimes becomes nonsingular, we instead apply a non-parametric

Table 4. Coefficient Differences between Probit and Probit with Sample Selection

Model	[1] 2005 (wave 2)		[2] 2006 (wave 3)	
Residential Mobility (= 1 if household (HH) moved from $t-1$)	Observed Difference	(S.E.)	Observed Difference	(S.E.)
Respondent Characteristics				
Age				
20–29	—		—	
30–39	0.0123	(0.0197)	0.0948	(0.0650)
40–49	−0.0740	(0.0213)**	0.0638	(0.0728)
50–59	−0.0397	(0.0218)$^{+}$	0.0080	(0.1362)
60 and above	0.1525	(0.0228)**	0.0278	(0.1239)
Sex (= 1 if R is female)	−0.0086	(0.0133)	−0.0318	(0.0529)
Married (= 1 if R is married)	−0.0033	(0.0174)	0.0477	(0.0739)
No child (= 1 if R has no children)	−0.0178	(0.0151)	−0.1082	(0.0494)*
Health condition (1: Good – 5: Bad)	−0.0033	(0.0041)	−0.0187	(0.0172)
Education				
Junior high school	−0.0074	(0.0206)	−0.1506	(0.1244)
High school	—		—	
Junior college	−0.0314	(0.0172)$^{+}$	0.0941	(0.0762)
4-year college+	−0.0079	(0.0148)	0.0187	(0.0549)
Other professional school	0.0727	(0.0256)**	0.0920	(0.1063)
Full-time worker (= 1 if R works full-time)	−0.0390	(0.0132)**	−0.0498	(0.0465)
Retired (= 1 if R is retired)	0.1201	(0.0154)**	−0.0345	(0.1386)
Household Characteristics				
Changes in HH type and composition				
# of HH members				
Unchanged	—		—	
Increased	0.1387	(0.0270)**	0.0895	(0.1667)
Decreased	0.0663	(0.0129)**	0.0755	(0.0785)
R is newly-wed	0.0789	(0.0699)	0.3516	(0.3155)
R has newborn children	−0.1574	(0.0340)**	−0.0928	(0.1239)
R owns house	0.0174	(0.0149)	−0.1279	(0.1782)

Notes:* **, *, and $^{+}$ indicate that the estimated coefficient difference is significant at the 0.01, 0.05, and 0.10 levels, respectively. Standard errors are computed by a non-parametric bootstrap procedure with 5000 replications. A set of dummy variables for regions and firm sizes is also controlled but is omitted from the results. The variables for place of residence and housing tenure are one-year lagged. The original results are shown in Tables 1 and 2. R denotes respondent.

bootstrap procedure to assess the significance of the bias. We draw 5,000 bootstrap replications with replacement to simulate the distribution of the bias. The results are summarized in Table 4.

It is found that in wave 2, the two models yield statistically different coefficient estimates for many of the respondent/household characteristics. Among these, highly significant differences are observed for the coefficients on respondent employment status and household type and composition, which are of central interest in the literature on mobility. However, consistent with our previous results, the differences are shown to be insignificant when we consider the accumulated attrition up to wave 3.

5. CONCLUSION

This paper aims to assess the extent of the attrition bias in the KHPS 2004–2006, focusing on household residential mobility as a specific empirical example. In applying a sample selection model to the attrition problem, we use unique information about the interview process, such as contact history and the list of interviewers assigned to each respondent, as identifying instruments. The results indicate that sample attrition leads to statistically significant bias in the coefficient estimates of household residential mobility functions in wave 2; however, it is attenuated in the subsequent wave. Moreover, the bias analysis suggests that the coefficients on several variables, which are of central interest in the literature on mobility, are significantly biased by nonrandom attrition.

REFERENCES

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3rd ed.). Chichester: John Wiley & Sons.

Bates, N. (2004). Contact histories: A tool for understanding attrition in panel surveys. available from http://www.fcsm.gov/committees/ihsng/ihsng.htm.

Becketti, S., W. Gould, L. Lillard, and F. Welch (1988). The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics 6*(4), 472–492.

Boehm, T. P. (1981). Tenure choice and expected mobility: A synthesis. *Journal of Urban Economics 10*(3), 375–389.

Clark, W. A. V. and W. F. J. V. Lierop (1986). Residential mobility and household location modelling. In P. Nijkamp (Ed.), *Handbook of Regional and Urban Economics*, Volume 1, pp. 97–132. Amsterdam: North-Holland.

Dieleman, F. M. (2001). Modelling residential mobility; a review of recent trends in research. *Journal of Housing and the Build Environment 16*(3), 249–265.

Fitzgerald, J., P. Gottschalk, and R. Moffit (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *Journal of Human Resources 33*(2), 251–299.

Hausman, J. A. and D. A. Wise (1979). Attrition bias in experimental and panel data: The gary income maintenance experiment. *Econometrica 47*(2), 455–474.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–162.

Hill, D. H. and R. J. Willis (2001). Reducing panel attrition: A search for effective policy instruments. *Journal of Human Resources 36*(3), 416–438.

McKenzie, C. R., M. Naoi, T. Miyauchi, and K. Kiso (2007). Attrition and individual behavior in the labor market. In Y. Higuchi and M. Seko (Eds.), *Dynamism of Household Behavior in Japan [III]*, Chapter 1, pp. 13–75. Tokyo: Keio University Press. (in Japanese).

Miyauchi, T., C. R. McKenzie, and M. Kimura (2006). An analysis of sample attrition and survey response behavior in panel data. In Y. Higuchi (Ed.), *Dynamism of Household Behavior in Japan [II]*, Chapter 1, pp. 9–52. Tokyo: Keio University Press. (in Japanese).

Naoi, M. (2007). Household residential mobility and the problem of sample attrition. In Y. Higuchi and M. Seko (Eds.), *Dynamism of Household Behavior in Japan [III]*, Chapter 2, pp. 77–98. Tokyo: Keio University Press. (in Japanese).

Nawata, K. and M. McAleer (2001). Size characteristics of tests for sample selection bias: A monte carlo comparison and empirical example. *Econometric Reviews 20*(1), 105–112.

Van de Ven, W. P. and B. M. Van Pragg (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics 17*(2), 229–252.