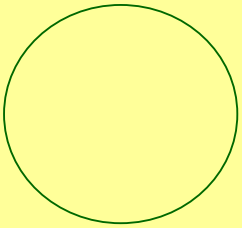
MIMAP TECHNICAL PAPER SERIES
NO. 20

An Analysis of Sample Attrition in the PSES Panel Data

G. M. ARIF
FAIZ BILQUEES



April 2006

PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS
ISLAMABAD, PAKISTAN

MIMAP TECHNICAL PAPER SERIES
NO. 20

An Analysis of Sample Attrition in the PSES Panel Data

G. M. ARIF
Chief of Research

FAIZ BILQUEES
Joint Director
Pakistan Institute of Development Economics

April 2006

PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS
ISLAMABAD, PAKISTAN

MIMAP Technical Paper Series No. 20

This study is a component of the Micro Impact of Macroeconomic Adjustment Policies (MIMAP), Pakistan Phase-II, a project being implemented by the Pakistan Institute of Development Economics, Islamabad. The main aim of this project is to analyse the impact of structural adjustment policies on income distribution and poverty in Pakistan. This work has been carried out with financial assistance from the International Development Research Centre, Ottawa, Canada.

ISBN 969-461-128-8

Pakistan Institute of Development Economics
Quaid-i-Azam University Campus
P.O. Box No. 1091
Islamabad 44000, Pakistan
Tel: 92-51-9206619
Fax: 91-51-9210886
E-mail: publications@pide.org.pk
pide@pide.org.pk
Web: http://www.pide.org.pk

The Pakistan Institute of Development Economics, established by the Government of Pakistan in 1957, is an autonomous research organisation devoted to carrying out theoretical and empirical research on development economics in general and on Pakistan-related economic issues in particular. Besides providing a firm foundation on which economic policy-making can be based, its research also provides a window through which the outside world can see the direction in which economic research in Pakistan is moving. The Institute also provides in-service training in economic analysis, research methods, and project evaluation.

CONTENTS

List of Tables and Figure

1. INTRODUCTION

Household surveys are the major source for socio-economic data in Pakistan; for example, the Household Income and Expenditure Survey (HIES),[1] which was initiated in 1963, generates data suitable to examine the consumption and saving behaviour of households as well as to determine the magnitude of poverty in the country. Labour Force Survey (LFS), which was also initiated in 1963, constitutes a major source of information on labour force participation, employment and unemployment levels. Pakistan Demographic Survey is the other regular series, which gathers demographic data to determine fertility and mortality levels as well as population growth rates. All these surveys are the standard cross-sectional surveys designed by the Federal Bureau of Statistics (FBS) to obtain information from a representative group of households at a given point of time. These datasets are widely used in Pakistan for social sciences analysis, and they are also the major source for monitoring the social sector progress.

However, the longitudinal or panel household surveys, where same households (as well as individuals) are re-interviewed after an interval, which in general varies between one and more than five years, have considerable advantage over the cross-section data. The panel data permits: 'tracing the dynamics of behaviour; identifying the influence of past behaviours on current behaviours; and controlling for unobserved fixed characteristics in the investigation of the effect of time-varying exogenous variables on endogenous behaviours' [Alderman, *et al.* (2000)]. The great attraction of panel data is that they can be used to study dynamics for individual households, including their well-being, child schooling and labour market dynamics. They can also be used to see who benefits and who loses from general economic development, or who gains and loses from a specific shock or policy change [Kurosaki (2002)].

The marginal contribution of such panel data to scientific and policy knowledge is probably extremely high in the developing countries. These are the countries currently undergoing dramatic social, economic and demographic transformation and our understanding of the transitions that people living in those countries are experiencing is sketchy [Thomas, *et al.* (2001)].

Panel household surveys are rare in developing countries including Pakistan, where the IFPRI household food security panel tracked about 700 households from rural-Pakistan in four districts only[2] between 1986 and 1991. They explored the characteristics

---

Authors' Note: The authors are thankful to Dr M. Iqbal, Senior Research Economist, for his help in improving the estimations, and to Mr Masood Ishfaq Ahmad, Ms Nabeela Arshad, and Mr Kamran Khan at PIDE for their assistance in data analysis. We are also thankful to Mr Zaheer Abbas for his valuable assistance in typing the manuscript.

[1]This survey has been renamed as the Household Integrated Economic Survey (HIES). In 1998-99, it was merged with the Pakistan Integrated Household Survey (PIHS). The HIES expenditure module has also been included in 2004-05 Pakistan Socio-economic Living Standard Measurement Survey (PSLM) for the estimation of poverty.

[2]These districts were Attock and Faisalabad in Punjab, Badin in Sindh, and Dir in the NWFP.

which had the greatest influence upon a household being chronically or transitorily poor. After a gap of 10 years, the IFPRI panel households were revisited in 2001 as part of a larger survey; the Pakistan Rural Household Survey carried out by the Pakistan Institute of Development Economics and financed by the World Bank. Kurosaki (2002) employed a two period panel dataset (300 households) collected from North-West Frontier Province of Pakistan to investigate vulnerability to risk as characteristic of dynamic poverty. These surveys provided useful information on individual household dynamics but they were based on rural samples drawn from a few districts of the country, therefore their findings could not be generalised.

Pakistan Institute of Development Economics has also generated a two period panel dataset namely the Pakistan Socio-economic Survey (PSES) based on a large sample of 3564 households carried out in 60 districts, representative at the national level.[3] These households were revisited in 2001, after a gap of about two years. The next round of the PSES is also under consideration. Because of its vast scope and because it is a panel, the PSES has the potential to become an important data source for scholars in all of the social sciences. However, the research community's willingness to use the PSES rests on having confidence in the underlying quality of the panel data.

A legitimate concern in any household panel data involves the extent of sample attrition and the degree to which attrition is non-random. Attrition occurs largely because respondents have moved from their places of original residence, they refuse to participate further in the panel or their households could not be found during a round of the panel. Exit from the panel might be correlated with individual and/or household characteristics in a way that biases estimates of population demographics or behavioural relationships. Similarly, failure to follow movers may yield a panel sample that is seriously deficient for many descriptive and analytical purposes [Thomas, *et al.* (2001)].

However, if exit from a panel sample is completely random, more confidence can be placed in results. The evidence from both developed and developing countries show that for many estimates the coefficients on standard variables in questions are unaffected by attrition. 'Even when attrition is fairly high.attrition apparently is not a general and pervasive problem for obtaining consistent estimates' [Alderman, *et al.* (2000)].

In the PSES panel, approximately 22 percent households could not be followed during the second round for several reasons including household movement, refusal of respondents, and difficulties in finding households in remote rural areas. Moreover, the events following 9/11 attacks in 2001 has also affected the PSES panel. In the two provinces of the country, NWFP and Balochistan, which are adjacent to Afghanistan, where the allied forces took the military action after 9/11, the entry of enumerators into many sampled villages or primary sampling units (PSUs) of these provinces was not possible. The deteriorating law and order situation in Sindh also made the resurvey of some PSUs difficult. So there was no choice but to drop some PSUs completely from Round II of the PSES. Is the attrition in PSES random? What type of households have

[3]The PSES was financed by the International Development Research Centre, Canada, through its Micro Impact of Macro Adjustment Policies (MIMAP) project.

been attrited? What is the impact of this attrition on the representativeness of the PSES? This paper has examines these questions in some detail.

Rest of the paper is organised as follows. Theories of sample attrition and statistical framework for the analysis are given in the next section, followed by a discussion on the PSES sample design in Section 3. The extent and nature of sample attrition are presented in Section 4, followed by an analysis of between the wave attrition, where multivariate and multinomial logit models of attrition have been estimated. The penultimate section presents the results of the BGLW[4] test while the last section presents the conclusions.

2. THEORETICAL CONSIDERATIONS[5]

Attrition in panel surveys is one type of non-response. At a conceptual level, many of the insights regarding the non-response in cross-sections carry over to panels. According to Fitzgerald, *et al*. (1998), attrition bias is associated with models of selection bias. Their statistical framework for the analysis of attrition bias, which has been used by several recent studies [see for example, Alderman, *et al.* (2000); Thomas, *et al.* (2001); Aughinbaugh (2004)], makes distinction between selections on variables observed in the data and variables that are unobserved. 'Given that' there is sample attrition, one determines whether or not there is selection on observables. For this purpose, selection on observables includes selection based on endogenous observables. For this purpose, selection on observables includes prior to attrition (e.g. in the first round of the survey). Even if there is selection on observables, this does not necessarily bias the estimates of interest. Thus, one needs to test for possible attrition bias in the estimates of interest as well' [Alderman, *et al.* (2000)].

Assume that the object of interest is a conditional population density $f(y|x)$ where y is scalar dependent variable and x is a scalar independent variable (for illustration, but in practice the extension to making x a vector is straightforward):

$$y = \beta_0 + \beta_1 + \varepsilon, \; y \text{ observed if } A=0 \quad \ldots \quad (1)$$

where A is an attrition indicator equal to 1 if an observation is missing its value of y because of attrition, and equal to zero if an observation is not missing its value y. Since (1) can be estimated only if $A=0$ that is, one can only determine $g(y|x, (A=0))$, one needs additional information or restrictions to infer $f(.)$ from $g(.)$. These can come from the probability of attrition, $PR(A=0|y, x, z)$, where z is an auxiliary variable (or vector) that is assumed to be observable for all units but not included in x. This implies estimates of the form:

$$A^* = \delta_0 + \delta_1 x + \delta_2 z + V \quad \ldots \quad (2)$$

$$A = 1 \; if \; A^* \geq 0 \quad \ldots \quad (3)$$
$$= 0 \; if \; A^* < 0$$

[4]BGLW test refers to test for the significance of sample attrition put forward by Becketti, Gould, Lillard and Welch (1988).

[5]This section has benefited from and is based on Fitzgerald, *et al.* (1998) and Alderman, *et al.* (2000).

If there is selection on observables, the critical variables is z, a variable that affects attrition propensities and that is also related to the density of y conditional on x. In this sense, z is "endogenous to y". Indeed, a lagged value of y can play the role of z if it is not in the structural relation being estimated is related to attrition. Two sufficient conditions for the absence of attrition bias due to attrition on observables are either (1) z does not affect A or (2) z is independent of y conditional on x. Specification test can be either of these two conditions. One test is simply to determine whether candidates for z (for example, lagged value of y) significantly affect A. Another test is based on Becketti, *et al.* (1988), and is known as BGLW test. It has been applied by Fitzgerald, *et al. (*1998) and Alderman, *et al*. (2000). In the BGLW test, the value of y at the initial wave of the survey (y_0) is regressed on x and on A. This test is closely related to the test based on regressing A and x and y_0 (which is z in this case); in fact, two equations are simply inverses of one another [Fitzgerald, *et al.* (1998)]. Clearly, if there is no evidence of attrition bias from these specification tests, then one has the desired information on $f(y|x)$.

3. THE PSES SAMPLE

The PSES is a panel survey of individuals, households and families that collects information on the lives of the respondents and the environment in which they live. The PSES has particularly been designed to document Pakistan's social and economic transformation through the combination of retrospective data collection and prospective panel. The baseline of the PSES or its Round I was fielded in 1998-99. The universe of the Round I of the PSES consisted of all urban and rural areas of the four provinces of Pakistan defined as such by 1981 population census excluding FATA, military restricted areas, and districts of Kohistan, Chitral, Malakand, and protected areas of NWFP. The population of the excluded areas constitutes about 4 percent of the total population. The village list published by the population census organisation in 1981 was taken as sampling frame for drawing the sample for rural areas. For urban areas the sampling frame developed by the FBS was used. In this frame each city/town has been divided into enumeration blocks of approximately 200 to 250 households. Cities having population of half a million or more such as Karachi, Lahore, Faisalabad, Rawalpindi, Multan, Hyderabad and Peshawar were treated as self-representing cities. Islamabad and Quetta, being federal and provincial capitals respectively, were also considered as the self-representing cities. Each of these cities constituted a separate stratum, which was further sub-stratified according to low, middle and high-income groups. The remaining urban population in each division of all the four provinces was grouped together to form a stratum. For rural sample, each district in Punjab, Sindh and NWFP was grouped together to form a stratum. For Balochistan province a division was treated as a stratum. Two stage stratified sample design was adopted for the 1998-99 PSES. Enumeration blocks in urban domain and Mouzas/Dehs/villages in rural domain were taken as PSUs. Households within the sampled PSUs were taken as secondary sampling units (SSUs). Within a PSU a sample of 8 households from urban domain and 12 households from rural domain was selected. Distribution of the 1998-99 PSES sample by province with rural/urban breakdown is reported Table 1. In total 3564 households were interviewed during the PSES-I.

Table 1

Distribution of the Sample Households, with Their Urban/Rural and Provincial Breakdown, 1998-99 PSES Round I

Province	Total	Rural	Urban
Punjab	1952	1320	632
Sindh	848	456	392
NWFP	508	324	184
Balochistan	256	168	88
Pakistan	3564	2268	1296

Source: Arif, *et al.* (2001).

The second wave of PSES-II was fielded approximately two years later in 2000-01. Some panel surveys in developing countries have revisited the original housing structure and interviewed whoever is there. This is the protocol recommended for longitudinal surveys in the World Bank's Living Standard Measurement Study [Glewwe and Jacoby (2000)]. However, many of the advantages associated with panel data require tracing the same individuals/households through time in order to better understand dynamics over the life course or to control unobserved characteristics that do not change over time [Thomas, *et al*. (2001)]. So the common practice in many existing panel surveys is to trace the same individuals/households for re-interviewing.

However, a key sample design question in these surveys is whether respondents who have moved from the location where they were last interviewed will be tracked and interviewed in their new location. The existing evidence shows that only a small number surveys have included local tracking. Those that have attempted to track longer-distance movers can be counted on one hand [Thomas, *et al.* (2001)].

In the PSES-II, the same households/individuals who were interviewed in the PSES-I of 1998-99 were traced and re-interviewed in 2000-01. The households who moved within a PSU were also tracked and re-interviewed, showing that the local tracking method was adopted for the PSES-II. However, no attempt was made if the sampled households have moved outside their original PSUs. Sample persons, who moved within a PSU, were also pursued even if they leave their original families. Such sample persons are called split offs.[6]

The fieldwork for the PSES-II was divided into two phases. During the first phase of fieldwork teams of interviewer were assigned to visit each household interviewed during the PSES-I. The teams were responsible for finding the households and tracking all members. If in 2000-01 any household had completely moved out from their 1998-99 location, field workers were instructed to obtain information about the current whereabouts from neighbours, relatives, friends, former employers, and local community leaders. Those found to be within the vicinity of the original PSU were treated as local

[6]Households with split-offs in 2000-01 are treated as separate households. In total 128 such households were identified in PSES-II. These households are not included in the present analysis.

tracking cases. Actual interviewing took place during the second phase when questionnaires were administered to respondents including the local movers directly by the field enumerators.[7]

4. EXTENT AND TYPE OF ATTRITION

Many studies in developed as well developing countries have examined the attrition of individuals from the original sample [Fitzgerald, *et al.* (1998); Alderman, *et al.* (2000)]. However, Thomas, *et al.* (2001) have focused on households and have analysed the attrition of sampled households between the first and second waves of Indonesian Family Life Survey. Following the approach of Thomas et al. for this paper to analyse the extent and nature of attrition between PSES-I and PSES-II, attention is focused on whether or not a 1998-99 household was re-interviewed in 2000-01. As noted earlier, the extent of attrition is affected by the design of the panel: whether or not the survey follows individuals who leave the original households, or who move away from the original survey area. Another common reason for attrition is refusal; households that have participated once are sometimes unwilling to do so again. When households from the first round are not traceable in the subsequent round, fewer of the original households remain in the survey. It is the most serious problem in panel datasets.

Table 2 (column 1) shows the attrition rate between the two rounds of the PSES for rural and urban areas as well as for each province. The overall attrition rate is counted as 22.5 percent. There is no marked difference between urban and rural sub-samples in terms of attrition rate, although it is slightly higher in the case of former. However, it varies considerably across the provinces, being lowest, only 15.5 percent in Punjab and highest in NWFP, 33.5 percent. In other two provinces, Sindh and Balochistan, attrition rates were also high, around 30 percent.

Table 2

Attrition Rate and Reasons for Attrition by Rural-Urban Areas and Province

Region/Province	Attrition Rate (%)	Reasons for Attrition (only Attriting Households)				
		All	Dropped from the PSES II Sample	Migration	Household not Found	Others
All Sample	22.2	100	21.6	26.4	22.8	29.2
Rural Areas	21.1	100	32.5	17.1	19.8	30.6
Urban Areas	24.0	100	5.0	40.6	27.4	27.0
Province						
Punjab	15.5	100	10.3	28.6	18.0	43.1
Sindh	28.8	100	33.1	29.3	35.1	2.5
NWFP	33.4	100	21.4	22.0	14.9	41.7
Balochistan	29.1	100	32.9	19.2	21.9	26.0

Source: Computed by the authors from the PSES 1998-99 and 2000-01 rounds.

[7]For the detail on PSES-II sample design, see Bilquees and Arif (forthcoming).

Table 2 also shows the distribution of only attriting households by reasons for attrition; classified into four major categories: 'dropped from the sample', 'migration of entire households', 'not finding households at the original location', and 'others', indicating either premises were locked, respondents were not available or unwillingness/refusal of households to be part of the panel. About one-fifth (21.6 percent) of the total attriting sub-sample (column 3 of Table 2) was dropped or excluded from the round II of PSES because of two main reasons; civil unrest particularly in NWFP and Balochistan after the post-September 11, 2001 US operation in Afghanistan[8], and deterioration in law and order situation in Sindh. These two factors made the access to some of the PSUs dangerous for enumerators. There was no choice but to drop these PSUs from the Round II. Table 2 shows that these PSUs (or households) were located primarily in rural areas.

Attrition because of the entire household mobility (migration) was higher in urban areas than in rural areas (Table 2). Overall more than a quarter of the attriting households moved out of their original PSUs. As noted earlier, in round II of the PSES, households which moved out of the PSUs were not followed. Twenty nine percent of the total attriting households were counted in the 'others' category, and it was the major reason in Punjab and NWFP, while there was no real difference between rural and urban areas in this category. It appears from this simple statistics that households exited the PSES-II for a variety of reasons.

It seems appropriate here to compare the PSES attrition rate with the rates of other panel surveys.[9] The well-known Panel Survey of Income Dynamics (USA) is the longest-running longitudinal household economic survey, initiated in 1968. In the first resurvey, a year later, 88.1 percent of the eligible respondents were re-interviewed; 86 percent were re-interviewed after two years. The China Health and Nutrition Survey has been successful in re-interviewing 95 percent of households during its second wave. The first round in 1989 interviewed 3795 households in eight provinces in China. The second wave which interviewed 95 percent of the households was carried out two years later. The third wave, four years after the baseline, interviewed 91 percent of the original households [Thomas, *et al.* (2001)]. Alderman, *et al.* (2000) report attrition rates of about one-third in panel surveys in Bolivia and Kenya which each had a two-year hiatus between the baseline and first follow up. Movers were not tracked in the Bolivian and Kenyan surveys and the attrition of the vast majority was attributed to migration. Surveys that do not track movers will systematically exclude particular subgroups of population. Attempts have also been made in some panels to trace the movers. For example, the Malaysian Family Life Survey, drew a random sample of 1262 ever-married women in Peninsular Malaysia. The second wave 12 years later, tried to follow movers but re-interviewed only 73 percent of the original primary respondents. However, the

[8]Surveys teams including female enumerators were considered to be part of the Western NGO outfits. Due to cultural considerations and strong anti-west sentiments the teams were not allowed into the samples villages (PSUs), therefore they were dropped from the sample.

[9]For more details, see Alderman, *et al.* (2000) and Thomas, *et al.* (2001).

Indonesian Family Life Survey was successful in tracking the movers and was able to re-interview more than 94 percent of households included in the original sample. The success of PSES in re-interviewing approximately 78 percent of the original PSES households seems to be satisfactory in the context of political and law and order situation prevailing at the time of the second round of PSES.

5. ANALYSIS OF BETWEEN-WAVE ATTRITION

5.1. Descriptive Analysis

This section first examines the observable correlates of attrition in the PSES primarily focusing on characteristics in 1998-99 (PSES-I). Then through the multivariate analyses, attrition probability equations as a function of 1998-99 characteristics have been estimated for the selected variables of interest. Attrition may be selective on many attributes of respondents. However, the focus of analysis, following Thomas, *et al.* (2001), is on the role of household economic status, measured as per capita expenditure (PCE). Household expenditure data is commonly used for poverty estimation, which is one of the main objectives of PSES.

Table 3 shows the mean value of 1998-99 characteristics of households and head of households in 1998-99 by their attrition status as of 2000-01. The first column of the table shows these characteristics for nonattritions or those households that remained in the panel in 2000-01 wave of PSES while the characteristics of the total sub-sample of attritors are presented in the last column. Attritors are further classified in columns 2 to 5 of Table 3 according to main reasons for attrition—'dropped from the sample', 'moved out', 'household not found' and 'others'—as discussed earlier.

Attriting households are slightly more likely than non-attriting households to be headed by females, particularly among those who moved out of the PSUs. Head of the attriting households are more likely to be younger than the head of non-attriting household, but the difference is small. Heads of those households that were dropped from the sample are relatively more illiterate. However, the overall sub-sample of the attritors is not different in levels of education from the non-attritors. In terms of employment, there is no major difference between heads of attriting and non-attriting households.

Family size seems to have an influence on attrition. Attritors are more likely to have small families compared to non-attritors. This is particularly true for households that moved out of the PSUs or those that could not be found. Among the former, 16 percent households consisted of only one or two members. Compared to 66 percent of attriting households, 76 percent of non-attriting households consisted of large families with 5 or more members.

There is also a difference between the non-attriting and attriting households in ownership of dwelling units; 91 percent of the former owned the dwelling units while this percentage was 82 percent in case of the later. The ownership of dwellings was particularly low, only 66 percent, among households who moved out of the PSUs. In

Table 3

1998-99 Characteristics by Attrition Status

1998-99 Characteristics	Non-Attritors (Interviewed in 2001)	Attritors by Type of Attrition				
		Dropped from the Sample	Migrated (Moved Out from the PSU)	Household Not Found	Others	All Attritors
Households Headed by Female (%)	7.6	3.5	12.8	9.9	7.7	8.6
Mean Age of Head of Household (Years)	48.7	47.9	44.9	44.4	44.9	45.4
Education of the Head of Household (%)						
Illiterate	56.5	70.9	50.7	50.0	53.3	55.7
Primary	15.4	13.4	10.0	17.0	12.2	13.0
Middle	10.4	3.5	12.3	8.2	13.1	9.7
Matric	10.5	7.0	15.2	12.6	8.3	10.8
Higher Education	7.2	5.2	11.8	12.1	13.1	10.8
Household Head Employed (%)	79.2	86.0	75.8	74.7	80.3	79.7
Family Size						
1 Member Only	1.2	1.2	7.6	4.9	5.7	5.0
2 Members	5.0	6.4	8.5	9.3	5.2	7.3
3-4 Members	17.4	14.5	18.0	23.1	30.6	22.0
5-6 Members	14.9	12.2	15.6	9.9	12.2	12.6
More than 6	61.4	65.7	50.2	52.7	46.3	53.0
Households Own Dwelling Unit (%)	91.0	93.5	65.9	82.2	87.1	89.7
Dwellings with Electricity (%)	76.1	52.3	81.0	78.0	87.3	75.9
Mean per Capita Expenditure (Rupees)	1026	1024	1411	1162	1012	1155
Average Household Income (Rupees)	2853	2332	3012	2920	3198	2897
Households Receiving Remittances (%)	12.6	13.4	14.2	13.2	11.8	13.1
Households Poor (%)	34.5	45.3	27.5	23.6	47.6	36.3
Urban Households (%)	35.2	9.3	61.1	47.8	37.6	40.1
Province (%)						
Punjab	59.4	18.6	42.2	30.8	58.5	39.2
Sindh	21.9	46.5	33.6	46.7	2.6	30.5
NWFP	12.2	20.9	17.5	13.7	30.6	21.2
Balochistan	6.5	17.5	6.6	8.8	8.3	9.2

Source: Computed by the authors from the PSES 1998-99 and 2000-01 rounds.

rural areas, these are usually the poor landless households without ownership of any dwelling units. In terms of per capita consumption expenditure and poverty it appears that attritors are concentrated in the lower portion of the socio-economic distribution.

Overall, attriting households were more likely to be located in urban areas than non-attriting households. However, more than 90 percent of those households that were dropped from the sample were located in rural areas during the Round I of the PSES. These household were located primarily in Sindh and NWFP. It can be concluded from this descriptive analysis that attritors and non-attritors appear to differ broadly on three counts: household size, ownership of dwelling units and geographical locations.

5.2. Multivariate Analysis

The characteristics of respondents, communities, the survey personnel and budgets all play a role in determining which respondents attrite and which do not [Thomas, *et al.* (2001)]. Although attrition may be selective on many attributes of respondents, we begin our analysis by focusing on the role of household's economic status, measured as per capita expenditure (PCE).[10] Five models have been estimated, where the dependent variable is whether attrition occurred between the survey rounds (1=yes; 0=no). Independent variables used in the models with their definitions are presented in Table 4. The sample included in each of these models consists of all 1998-99 PSES households.

Table 4

Mean and Standard Deviation of Selected Variables

	Attritors		Non-attritors		All	
Variables (1998-99)	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (Head of Household) in Years	45.38	14.54	48.71	14.35	47.96	14.46
Literacy (Head) (1=Literate)	0.43	0.50	0.43	0.52	0.43	0.52
Female Headed Household (1=Female Headed)	0.09	0.28	0.08	0.26	0.08	0.27
Log of Family Size	1.63	0.62	1.80	0.52	1.76	0.55
One-member Household (=1)	0.05	0.22	0.01	0.11	0.02	0.14
Two-members Household (=1)	0.07	0.26	0.05	0.22	0.06	0.23
House Ownership (No=1)	0.18	0.39	0.09	0.29	0.11	0.31
Employed (Head) (Yes=1)	0.79	0.41	0.79	0.41	0.79	0.41
Monthly Household Income (Rs)	7.85	0.91	7.92	0.81	7.91	0.83
Log of per Capita Expenditure (Rs)	6.82	0.61	6.78	0.51	6.78	0.54
Poverty Indicators (Poor=1)	0.36	0.48	0.35	0.48	0.35	0.48
Urban (=1)	0.40	0.49	0.35	0.48	0.36	0.48
Electricity (Yes=1)	0.82	0.78	0.80	0.67	0.80	0.70
Remittance-Receiving Households (Yes=1)	0.14	0.37	0.13	0.35	0.13	0.35

Source: Computed by the authors from the PSES 1998-99 and the 2000-01 rounds.

[10]In most developing countries, including Pakistan, money income measures of well-being are problematic as large numbers of households have limited connection with the formal and paid labour market sector. Consequently, the PSES devoted considerable survey time to a consumption module that collects information on more than 50 groups of major items in the household budget. The value of expenditures, production for own consumption, and transfers are aggregated to calculate household "expenditure".

Following Thomas, *et al.* (2001), the first model of attrition includes the only covariate, ln(PCE), measured at the household level in PSES-I. Table 5 presents coefficient estimates from logit regressions. The first model indicates that there is a statistically significant positive relationship between PCE and the probability of leaving the panel. On average, households with higher economic status were more likely to attrite between the two waves so that without weighting, PSES II will be less representative of higher economic status households than would be a random household survey.

In model 2, two variables, ln(PCE) and ln(family size) have been included. It relaxes the implicit assumption in the per capita measure that the effects of expenditures and family size on attrition are equal in magnitude and opposite in sign. The assumption is strongly rejected. Holding PCE constant, an increase in family size (in 1998-99) is associated with a higher probability that the household was re-interviewed (in 2000-01). PCE does not remain a significant determinant of attrition in model 2.

The third model in Table 5 includes two dummies; (1) if household consists of one member, or (2) if it consists of two members. It demonstrates that nonlinearities exist

Table 5

Determinants of Attrition

Correlates (1998-99)	Model 1	Model 2	Model 3	Model 4	Model 5
Ln per Capita Expenditure	0.148**	–0.093	–0.082	–0.345*	–0.352*
Ln Family Size	–	–0.569*	–0.474*	–0.643*	–0.535*
One Person Household	–	–	0.649*	–	0.753*
Two Person Household	–	–	-0.059	–	–0.053
Female Headed Households	–	–	–	–	–0.051
Age (Head)	–	–	–	–	–0.005
Literacy (Head)	–	–	–	–	–0.012
Employed (Head)	–	–	–	–	–0.106
No Ownership of House	–	–	–	–	0.551*
Electricity	–	–	–	–	0.053
Remittances Receiving	–	–	–	–	0.195
Community Variables	–	–	–	–	–
Mean per Capita Expenditure	–	–	–	0.001*	0.001*
Mean Family Size	–	–	–	0.001	–0.081*
Mean Education (Head)	–	–	–	–	–0.123
Mean Age (Head)	–	–	–	–	–0.039*
Urban	–	–	–	–	0.102
Punjab	–	–	–	–	–
Sindh	–	–	–	–	0.744*
NWFP	–	–	–	–	1.264*
Balochistan	–	–	–	–	0.778*
Constant	–2.127*	0.541	0.272	1.859	3.956*
–2 Log Likelihood Ratio	3767.18	3715.35	3709.218	3702.260	3434.464
Cox and Snell R^2	0.001	0.016	0.017	0.019	0.078
Negelkerke R^2	0.002	0.024	0.026	0.029	0.120
N	3545	3545	3545	3545	3504

Source: Computed by the authors from the PSES 1998-99 and 2000-01 rounds.

* Shows significance at 5 percent level. ** Shows significance at 10 percent level.

in both the numerator and dominator of PCE. There appears to be no significant effect of increasing PCE on attrition. The association between attrition and household size is nonlinear. Attrition rates decrease as household size increases and these effects are strongest for households with one member. As shown earlier, 5 percent of the 1998-99 PSES sample consisted of one-member households. When someone leaves a single person household, the entire household no longer exists in that location and tracking becomes more difficult.

The characteristics of the communities (or PSUs) in which respondents reside may also be important because residential living in Pakistan is likely to be quite segregated along economic lines. In addition, some PSUs are easily accessible, while reaching others poses formidable logistical problems. The fourth model has added measures of PSU-specific mean family size and PCE.[11] Results show that an increase in per capita consumption at the community level increases attrition, an effect that may accelerate among the wealthiest quarter of PSUs (Table 5). Household size in the community did not turn out to be significant. However, including PSU level measures fundamentally alters the individual household level PCE effects.

To assess whether the result is sustained in more comprehensive multivariate models, model 5 includes additional individual, household and community level characteristics. At the individual and household levels, the model includes age and education of the head of household, whether the household is headed by a female, whether the household lives in an owner-occupied dwelling, whether the household has electricity connection and whether it receives remittances. A parallel set of variables is measured at the PSU level: average age and education of household heads in the PSU, proportion of households headed by females, and fraction of owner-occupied dwellings. The model also controls for urban/rural area, and the province in which the sample household is located.

Among the additional household level covariates, the only significant variable is the ownership of home; attrition is higher if the sample lived in a non-owned (rented or gifted) house. Many studies have documented that geographic mobility increases among those who are not home owners. Between-wave mobility is clearly a key reason why some households cannot be found and why an interview is not completed. Controlling these characteristics, single-person households are less likely to be re-interviewed, suggesting that they were proxying for higher mobility households.

Attrition rates are higher in NWFP, Balochistan and Sindh as compared to Punjab. These are the provinces where PSUs were dropped due to civil unrest and adverse law and order situation. The urban dummy does not remain a significant determinant of attrition.

It is evident from the multivariate analyses that there is a very different relation between economic resources and attrition rates when these resources are measured at the household level and when they are measured at the community level. The impact of community level resources is substantially muted by the inclusion of additional controls

[11]PSUs means are based on all PSES-I households, whether or not they were followed up.

in the analysis (model 5) suggesting that resources are a proxy for an array of community characteristics. Controlling the average level of economic well-being in a PSU, attrition is highest among the lowest resource households.

5.3. Types of Attrition in PSES II: A Multinomial Logit Analysis

As noted earlier, a unique feature of the PSES-II is that around one-fifth of the attriting households were not followed or dropped from the sample because of civil unrest and deterioration of law and order situation. Are these households different from other households in terms of socio-economic characteristics? Following Thomas, *et al.* (2001) methodology, Table 6 present the results of two multinomial logit models. In these models, the dependent variable is defined as one of three mutually exclusive panel

Table 6

Multinomial Logistic Models of Types of Attrition. Probability of Households Being Dropped from the Sample and Households Not Found/Refused Relative to Households Interviewed

	Expenditure		Poverty	
Characteristics (1998-99)	Dropped/ Interviewed	Other Attritor/ Interviewed	Dropped/ Interviewed	Other Attritor/ Interviewed
Ln per Capita Expenditure	–0.638*	–0.322*	–	–
Poverty Status	–	–	6.654*	0.427*
Ln family Size	–0.357	0.616*	–0.357	–0.621*
One Person Household	0.091	0.756*	–0.083	0.689*
Two Person Household	–0.096	0.059	–0.203	–0.106
Female Headed Households	–0.550	0.049	–0.597	0.028
Age (Head)	0.004	0.008**	0.004	–0.007**
Literacy (Head)	0.010	0.025	0.093	–0.036
Employed (Head)	0.367	–0.230**	0.372	–0.225
Ownership of House	–0.087	0.622*	–0.056	0.626*
Electricity	–0.169	0.101	–0.186	0.103
Remittances	0.743	0.099	0.746*	0.108
Community Variables	–	–	–	–
Mean per Capita Expenditure	0.001*	0.000	0.001*	0.000
Mean Family Size	0.008	–0.107*	0.001	–0.111*
Mean Education (Head)	–3.697*	0.603*	–3.715*	0.594*
Mean Age (Head)	0.019	–0.050*	0.017	0.051*
Urban	–1.622*	0.378*	–1.622*	0.400*
Sindh	2.430*	0.317*	2.443*	0.341*
NWFP	1.617*	1.194*	1.637*	1.200*
Balochistan	1.515*	0.640*	1.523*	0.651*
Constant	0.187	4.337*	–4.435*	2.128*
–2 Log Likelihood Ratio	3947.191	–	3987.700	–
R^2 (Psendo)	0.155	–	0.158	–
N	3504	–	3504	–

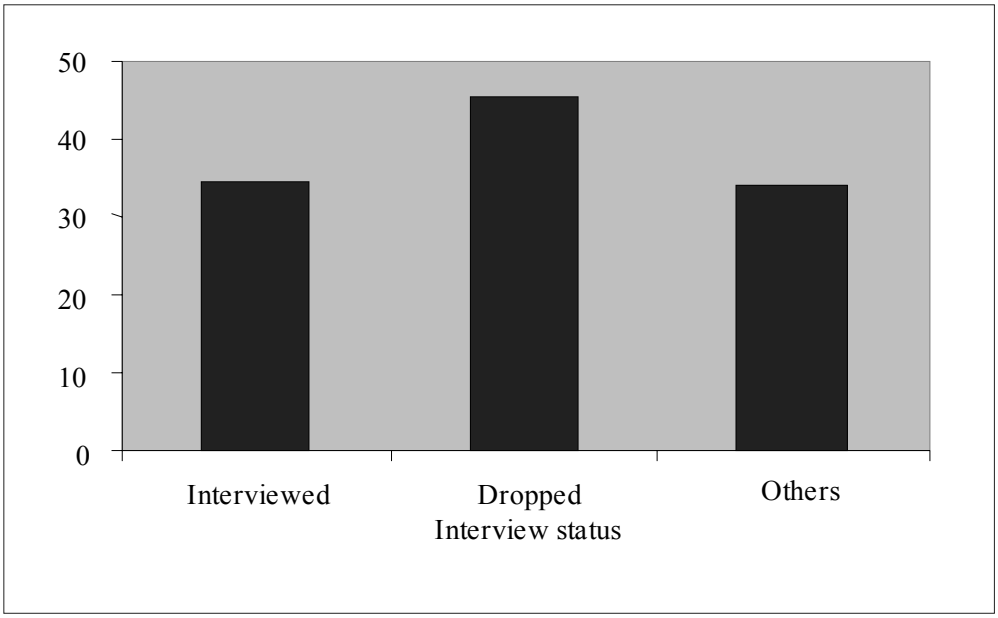
Source: Computed by the authors from the PSES 1998-99 and 2000-01 rounds.

*Shows significance at 5 percent level. ** Shows significance at 10 percent level.

outcomes: completed the PSES-II interview, households dropped from the sample, and all other attriting households. Households who completed the interview in 2000-01 round of PSES are the reference category in the multinomial logit models. All variables used in the logistic regressions previously have also been entered into these models. In model 2, PCE has been replaced by a dummy variable of poverty status.

Findings of the two models are similar. In addition to PCE, poverty status, remittances, and factors associated with geographical locations turned out to be significant in these models. Urban residents were less likely than their rural counterparts to be dropped from the sample. The households excluded from the sample were more likely to be located in NWFP, Balochistan and Sindh than in Punjab. It appears that characteristics related to geographical locations, being rural residents of these provinces, were the key determinants of being excluded from the sample. No doubt, these locations were in general poor. Figure 1 provides information on the incidence of poverty in 1998-99 by their interview status in 2000-01. Households that were dropped from the sample were poorer than the other attriting households or the non-attriting sub-sample.

Fig. 1. Incidence of Poverty in PSES Round I, Households by Interview Status in Round II.



Source: Computed by the authors from the PSES 1998-99 and 2000-01 rounds.

Household characteristics associated with difficulty in re-interviewing respondents or locating the other attriting households seem to be largely related to mobility. Table 6 shows that households that were larger in 1998-99 were easier to re-interview. It is because the probability that all members had moved from the 1998-99 location tends to decline with household size. Similarly, households with older heads, those with employed

heads, and those who were owner-occupiers were all much more likely to be found and all of these characteristics are associated with lower geographic mobility.[12]

Households belonging to Sindh, NWFP and Balochistan were more likely to be in the categories 'not found', 'moved out', or refused. The most plausible explanation has to do with the time costs associated with visiting, and revisiting, these households. Households in some of the PSUs located in these provinces tend to be widely dispersed. If no one was at home on the first contact, the interviewers would return multiple times until contact was made. In these areas, multiple visits involved substantial time costs and so there were not as many visits as in more compact PSUs. There are also several mountainous areas in NWFP and Balochistan. Each of revisits in these areas was expensive in terms of time and transport cost.

6. BGLW TEST

As discussed in Section 2, BGLW test, introduced and used initially by Becketti, *et al.* (1988), is the other method of testing the attrition bias. This test examines whether those who subsequently leave the PSES sample are systematically different from those who stay in terms of their initial behavioural relationships. We examine the consumption (lnPCE) equations as well as poverty equations but distinguishing two successively more restrictive subsets of participants—all 1998-99 households, and those still in the sample in 2000-01, labelled as 'Always in' or non-attritors.

Tables 7 and 8 present estimates of OLS regression for consumption equations and logit estimates for poverty equations respectively. A standard set of household and head of household characteristics, including income, family size, age and education of the head of household, ownership of dwelling unit, remittance-receiving status and place of residence, are the independent variables.

All the results are significant. These estimates indicate a number of associations that are consistent with widely held perceptions about consumption behaviour or poverty. For example, household income is significantly positively associated with consumption and significantly negatively associated with poverty incidence. Age and education of the head of household have a positive impact on consumption while they are negatively associated with poverty. A similar pattern of association was also found for family size. Remittances and residence in urban areas have a positive influence on consumption while their impact, on the incidence of poverty as expected, is negative.

However, there are no significant differences between the set of coefficients for the sub-sample of those lost to follow-up versus the sub-sample of those re-interviewed for indicators of either consumption or poverty. Furthermore the insignificance of the differences of the set of coefficients confirms that the coefficient estimates of standard background variables are not affected by the sample attrition.

---

[12]Single-person households were more difficult to re-interview. When such a respondent moved, he or she may leave little trace. PSES-II is designed as a multi-faceted instrument with a household-level questionnaire targeted at the female head, one targeted at the male head and then an individual-level questionnaire for every household member. In single-person households, the survey burden is large.

Table 7

Household Expenditure: OLS Regression Model

Variables	Full Sample	Always in (Non-attritors)	Difference
Ln Income	0.231*	0.222*	0.009
Age (Head)	0.007*	0.006*	0.001
Education (Head)	0.167*	0.167*	0
Family Size	–0.518*	–0.532*	–0.014
Own House	–0.044**	–0.059*	–0.015
Remittances	–0.198*	0.172*	0.026
Urban	0.168*	0.169*	–0.001
Constant	5.426*	5.550*	–0.126
Adjusted R^2	0.442	0.430	–
F-statistics	309.165*	228.223*	–

Source: Computed by the authors from the PSES 1998-99 and 2000-01 rounds.

* Shows significance at 5 percent level. ** Shows significance at 10 percent level.

Table 8

Correlates of Poverty: Logistic Regression Model

Variables	Full Sample	Always in (Non-attritors)	Difference
Ln(Income)	–0.911*	–0.923*	0.012
Age (Head)	–0.033*	–0.032*	–0.001
Education (Head)	–0.610*	–0.737*	0.127
Ln(Family Size)	2.350*	2.528*	–0.178
Own House	0.253**	0.274	–0.021
Remittances	–1.342*	–1.340*	–0.002
Urban	–0.702*	–0.739*	0.001
Constant	4.237*	3.929*	0.037
–2 Log Likelihood Ratio	2758.740	2101.782	–

Source: Computed by the authors from the PSES 1998-99 and 2000-01 rounds.

* Shows significance at 5 percent level. ** Shows significance at 10 percent level.

7. CONCLUSIONS

The longitudinal or panel household surveys are rare in Pakistan, and a legitimate concern in these surveys involves the extent of sample attrition and the degree to which attrition is non-random. Pakistan Institute of Development Economics generated a two period panel dataset, PSES, with baseline sample of 3564 households in 1998-99. These households were revisited in 2001, after a gap of about two years. Approximately, 22 percent households could not be re-interviewed during the round II of the PSES for several reasons including household movement, refusal of respondents, and difficulties in finding

households in remote rural areas. The 9/11 event of 2001 also affected adversely the tracking of the PSES panel in NWFP and Balochistan. The deteriorating law and order situation in Sindh made the resurvey of some PSUs very difficult. This paper has first examined the extent of attrition in the PSES; and has then differentiated the characteristics of attriting households from the non-attriting households. It has also examined whether this attrition has affected the coefficients of standard background variables.

The analysis shows that attrition was selective on many attributes of respondents. This selectivity, however, differs across the types of attrition. Exclusion of some PSUs was locality-specific. Although the households excluded from the sample were relatively poor, they were not different from non-attriting households in other social and economic attributes. Factors associated with mobility such as small family size and non-ownership of residential houses were associated with attrition due to migration or not finding households at their places of origin.

After controlling for community variables, attrition was higher among the lowest resource households, as measured in terms of per capita expenditure. However, no significant differences could be found between the set of coefficients for attritors versus non-attritors for indicators of interest, consumption and poverty. The present analysis therefore suggests that the coefficient estimates of standard background variables are not affected by sample attrition. Attrition of more than 22 percent sample of the PSES is not a pervasive problem for obtaining consistent estimates. Several recent studies on health, education and poverty have used the PSES panel data or its any single round.[13] The present analysis on the extent and nature of PSES panel attrition suggests that researchers may proceed with greater confidence in their attempts using this panel dataset to explore dynamic relationships in Pakistan. It is also worth investing in PSES panel for its next round.

REFERENCES

Alderman, H., J. R. Behrman, H. Kohler, J. A. Mauccio, and S. Watkins (2000) Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing Country Samples. The World Bank, Development Research Group Rural Development. (Policy Research Working Paper 2447.)

Ali, S. Mubashir, and M. F. K. Kiani (2003) Ageing and Poverty in Pakistan. Pakistan Institute of Development Economics, Islamabad. (MIMAP Technical Paper Series No. 18.)

Arif, G. M. (2003) Poverty Dynamics in Pakistan: Evidence from the Two Panel Household Survey. In Pakistan Human Condition Report 2003. Centre for Research on Poverty Reduction and Income Distribution, Islamabad.

Arif, G. M., and Faiz Bilquees (2006) Chronic and Transitory Poverty in Pakistan: Evidence from a Longitudinal Household Survey. Pakistan Institute of Development Economics, Islamabad. (MIMAP Technical Paper Series No. 19.)

[13]See, for example, Qureshi and Arif (2001), Nasir (2001), Qureshi, *et al.* (2001), Arif (2003), Ali and Kiani (2003).

Arif, G. M., S. Mubashir Ali, Z. M. Nasir, and Nabeela Arshad (2001) An Introduction to the 1998-99 Pakistan Socio-economic Survey (PSES). Pakistan Institute of Development Economics, Islamabad. (MIMAP Technical Paper Series No. 4.)

Aughinbaugh, A. (2004) The Impact of Attrition on the Children of the NLSY97. *The Journal of Human Resources* 39:2.

Becketti, Sean, William Gould, Lee Lillard, and Finis Welch (1988) The Panel Study of Income Dynamics after Fourteen Years: An Evaluation. *Journal of Labor Economics* 6, 472–92.

Fitzgerald, J., P. Gottschalk, R. Moffit (1998) An Analysis of Sample Attrition in Panel Data. *The Journal of Human Resources* 33:2.

Glewwe, P., and H. Jacoby (2000) Recommendations for Collecting Panel Data. In Margaret Grosh and Paul Glewwe (eds.) *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Vol. 2. Washington, D. C: The World Bank.

Kurosaki, Takashi (2002) Consumption Vulnerability and Dynamic Poverty in the North-West Frontier Province, Pakistan. www.ier.hit-u.ac.jp/~kurosaki

Nasir, Z. M. (2001) Poverty and Labour Market Linkages in Pakistan. Pakistan Institute of Development Economics, Islamabad. (MIMAP Technical Paper Series No. 7.)

Qureshi, S. K., and G. M. Arif (2001) The Poverty Profile of Pakistan. Pakistan Institute of Development Economics, Islamabad. (MIMAP Technical Paper Series No. 5.)

Qureshi, Sarfraz Khan, Hina Nazli, and Ghulam Yasin Soomro (2001) Nutritional Status in Pakistan. Pakistan Institute of Development Economics, Islamabad. (MIMAP Technical Paper Series No. 8).

Thomas, D., E. Frankenberg, and J. P. Smith (2001) Lost but not Forgotten: Attrition in the Indonesian Family Life Cycle Survey. *The Journal of Human Resources* 36:3, 556–592.

ABSTRACT

Panel household surveys study the dynamics for individual households, including their well-being, child schooling, and labour market dynamics, and they can also be used to observe who benefits or loses from general economic development. The MIMAP panel survey, covering 300 households in 60 districts, is the largest panel survey carried out in Pakistan. This paper addresses two very legitimate concerns regarding the panel data—the extent of attrition and the degree to which attrition is non-random. The paper describes in detail the factors responsible for attrition in this panel at the provincial as well as the urban-rural levels. The paper first examines the extent of attrition and then differentiates the characteristics of the attriting household from the non-attriting household, and how this attrition affects the coefficients of standard background variables. It shows that attrition was higher among the poorest households as measured by the per capita expenditures. However, no significant difference could be found between the set of coefficients for attritors versus non-attritors for indicators of interest, i.e., consumption and poverty. This shows that coefficient estimates of standard background variables are not affected by sample attrition, implying that an attrition of more than 22 percent sample of PSES is not a serious problem for obtaining consistent estimates. This analysis of the extent and nature of PSES panel attrition suggests that researchers can use this panel data set with confidence to explore the dynamic relationship in Pakistan. It also indicates clearly that it is worth investing in another round of the PSES panel.