Attrition in Longitudinal Data and Income Mobility in Chile

Ricardo D. Paredes
José Joaquín Prieto
José Ramón Zubizarreta[1]

Abstract

In this paper we analyze the nature and extent of attrition in Chile's most relevant longitudinal household survey, the Panel CASEN 1996-2001. All studies using this survey have suggested that despite the unequal and persistent income distribution in income, mobility is high. However, none of them have considered the attrition bias problem which might distort mobility results. Through standard methods, we detect attrition and find a significant effect of bias. We correct it through new sample weights and evaluate how this correction affects income mobility studies. We conclude that the attrition bias overestimates mobility and hence, has significant effects on policy prescriptions.

Keywords: longitudinal survey, attrition bias, income mobility.
JEL classification: I32, I38.

1 Introduction

In year 2001 Chile completed its first longitudinal household survey with regional representation, the Panel CASEN 1996-2001; one of the few panel surveys in Least Developed Countries (LDC's).[2] This survey led to a number of studies on income mobility and poverty dynamics, which are of particular interest for a country like Chile, which exhibits one of the most unequal income distributions in the world. These studies show that the persistent income inequality contrasts with a relatively high mobility of all but the richest income decile (e.g., Aguilar 2002, Castro and Kast 2004, Contreras et. al 2004, and Paredes and Zubizarreta 2005). However, neither of these studies analyzed or controlled for attrition, a central problem in longitudinal data, concerning the loss of data and of representativeness of the sample over time.

[1] Paredes, rparedes@ing.puc.cl, Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul; Prieto, jjprieto@uahurtado.cl, Observatorio Social Universidad Alberto Hurtado, Alameda 1869 5th floor, Santiago; Zubizarreta, jzubizar@puc.cl, Escuela de Ingeniería Pontificia Universidad Católica and Observatorio Social Universidad Alberto Hurtado. We thank Ernesto Castillo, Luis Maldonado, Mauricio Rosenblüth, Berta Teitelboim, and the participants of the PUC Industrial and Systems Engineering Department Seminar for their comments. Usual disclaimers apply.
[2] According to Yaqub (2000), following the UNDP classification only 5 out of 44 countries with low human development and 7 out of 66 countries with intermediate human development have this kind of surveys.

An understanding of the nature of this problem is especially important in LDC's, which are starting to implement panel surveys. In particular, it is for Chile, where this persistent and unequal income distribution together with successful macro performances creates a puzzle that proper longitudinal data might clarify. Above and beyond, longitudinal data measures allow a better evaluation of the efficiency of public policies than do cross sectional data.

This paper analyzes the nature and extent of attrition in the Panel CASEN survey, and corrects for this problem. We are especially interested in estimating the effect of attrition on mobility and describing the methodology such that future surveys to be implemented in the region take this problem into account.

The paper has four sections besides this introduction. Section 2 describes the problem and presents some relevant international evidence. Section 3 presents the data and the methodology to detect and correct attrition. Section 4 presents the results and evaluates the policy effects of the correction, and section 5 concludes.

2 The attrition problem

Panel data have a number of advantages over cross sectional data. By following the same units, they measure individual change and make possible an accurate study of transitions between states. They also permit the analysis of the impact of particular policy interventions, the development of models of behavior through the technique of life histories, and to control the effect of non observables, clearly distinguishing the cohort effect (see Rose, 2000). Notwithstanding, the recollection of panel data usually is difficult and expensive, so there is not a single view about the benefits of having this kind of data in LDC's (see, for instance, Ashenfleter *et al.,* 1986).

The quality of a panel depends critically on the participation of the original sample members over time. When some of them are lost, because they move or because they are no longer willing to answer the questionnaire, the quality of the panel lessens since the variances of the sample increases, hence reducing the efficiency of estimates. Furthermore, if those remaining in the sample have systematically different

characteristics than those that are lost, attrition is biased and estimators so. This is the main problem associated with attrition.

High attrition is usual in panel data, including the most recent and relevant ones, like the European Community Household Panel (ECHP), the Household, Income and Labor Dynamics (HILDA) in Australia and the Panel Study of Income Diynamics (PSID) in the USA. Though comparisons between countries must be handled with caution, we define the annual attrition rate as $1-(1-q)^{1/T}$, where q is total attrition and T the time interval between waves 1 and 2. International comparisons suggest that attrition in the Panel CASEN after 5 years is in a lower range (Table 1).

Table 1

Attrition rates in different countries: wave 1 through wave 2

Country	Panel survey	Period between waves 1 and 2	Attrition rate between waves 1 and 2	Attrition rate per year
Australia	HILDA	1 year	0,132	0,132
Belgium	ECHP	1 year	0,087	0,087
Denmark	ECHP	1 year	0,140	0,140
France	ECHP	1 year	0,112	0,112
Germany	ECHP	1 year	0,073	0,073
Germany	GSOEP (Occidental)	1 year	0,124	0,124
Germany	GSOEP (Oriental)	1 year	0,089	0,089
Greece	ECHP	1 year	0,097	0,097
Ireland	ECHP	1 year	0,153	0,153
Italy	ECHP	1 year	0,059	0,059
Luxembourg	ECHP	1 year	0,065	0,065
Netherlands	ECHP	1 year	0,089	0,089
Portugal	ECHP	1 year	0,047	0,047
Spain	ECHP	1 year	0,126	0,126
UK	BHPS	1 year	0,116	0,116
UK	ECHP	1 year	0,238	0,238
USA	PSID (SRC)	1 year	0,145	0,145
USA	PSID (SEO)	1 year	0,091	0,091
Bolivia	PIDI	2 years	0,160	0,083
Kenya	KDICP	2 years	0,410	0,232
South Africa	KIDS	5 years	0,350	0,083
Chile	Panel CASEN	5 years	0,281	0,064

Notes: elaborated from data from Alderman et al. (2001), Peracchi (2000) and Watson (2004).

High attrition doesn't necessarily imply bias. Furthermore, even if there exists bias, depending on specific analysis certain estimates might be unbiased. In a special issue on attrition in panel data of the *Journal of Human Resources* (Spring 1998), a number of studies show that in spite of achieving high attrition rates –as high as 50% for the PSID after 19 years– there were no differences between those households lost and those re-interviewed regarding several important variables. Analyzing the PSID, Fitzgerald *et al.* (1998) show that attrition is highly selective in low socioeconomic households, and in households with unstable marriages and histories of migrations. However, in spite of the high level of attrition there is no evidence suggesting that attrition has strongly distorted

the representativeness of the sample. At the same time they find that the cross sectional representativeness of the sample has remained practically intact.

There is a vast literature on the characteristics of panel members that are lost either because they refused to answer or because they changed their residence. Several demographic characteristics have a close relation with the probability of completing successfully the survey. Lynn *et al.* (2005), in a detailed survey of studies on the subject, find several characteristics repeated through them. In developed countries, regarding refusals, there are more propensities to don't answer in old persons, persons with low incomes or low education, single persons, ethnic minorities, and households with high mobility and from urban zones. In the case of those that weren't contacted, attrition is more probable in old and young men, persons with high incomes, single person households, and again in households with great mobility and urban zones.

Regarding attrition correction, there are basically two ways to face it. First, an *ex ante* way consisting of over representing the observations most likely to be lost. Second, an ex post method that consists in correcting the weights, such that the new weighted sample yields the same average means. Naturally, the second method is the only viable one once the survey has been taken and when lost observations are not random.

3 Methodology and results

The Panel CASEN follows effectively 4,042 households from the III, VII, VIII and Metropolitan Regions, from an initial sample of 5623 households taken from the CASEN 1996 survey.[3] The panel is representative of these four regions which, in turn, represent about 60% of the national population.

3.1 Detection methodology

Once the data have been collected and non random attrition detected, the correction of the weights requires estimating the attrition functional form. Fitzgerald *et al.* (1998) develop an econometric framework for the detection of attrition on observable and non

[3] In a different way to most panel surveys, the initial sample of the Panel CASEN was drawn from another survey, the CASEN 1996, after it was collected. The CASEN's are cross sectional socioeconomic surveys collected every two or three years, and the main instruments to evaluate policies in Chile.

observable variables. From their analysis they conclude that correcting for non observables demands very complex models or the use of census data, which are not always available for the years of interest. Therefore, they recommend the use of other techniques that only need the use of observables.

In Chile there are no surveys having the required statistical similitude or confidence to consider non observables and therefore –as in most attrition studies in longitudinal surveys– the analysis must be restricted to observable variables. Hence, and following Alderman et al. (2001) in their comparative study of attrition in longitudinal surveys in LDCs, we perform three tests to determine the nature and importance of attrition. In first place, a t test to compare means of socioeconomic variables of common interest in the mobility and poverty studies, both for the group followed up in time, as for the group of households that were not surveyed again in the year 2001. In the second place, using a probit model, we estimated what type of variable is more likely to be systematically associated with attrition. Finally, we applied the BGLW test (Becketti, Gould, Lillard and Welch, 1988) which basically tests whether there are structural changes in the parameters obtained from equations estimated with the sample of re-interviewed households and with those we were unable to follow-up (see, e.g., Fitzgerald et al., 1998).

3.2 Correction methodology

Among the methods for correcting the bias associated with non-response, three are the main ones: tracking, sample selection modeling and weighting (McGuigan *et al.*, 1995). The first one is implemented as part of the data collection stage and consists in finding the new location of the households in the sample. Tracking is the ideal solution to non-response, but budget restrictions hardly ever permit it. In an illustrative example, Graham and Donaldson (1993) cite that the cost of tracking is five times the cost of obtaining normal data. However, since it is an *ex-ante* method to the collection of data, our interest lies in the two following methods resorted to for correcting the attrition bias.

It is clear that the construction of the Panel CASEN weights did not considered the effect of attrition. Therefore it was natural to choose the third method, which, in a different way to sample selection modeling, provides a general basis for any analysis to

be conducted from the data. This implied revising 1996 and 2001 weights, and eventually reconstructing and adjusting them. This is something we did considering both the representative of the original Panel sample in 1996 and the respondent sample in 2001 to the whole population.[4]

Regarding the *ex post* correction methods, Kalton and Brick (2000) mention four ways of adjusting the weighs by non response in successive waves of a longitudinal survey: weighting classes, tree algorithms, generalized raking and logistic regressions. For these alternatives, Rizzo *et al.* (1996) present a comparative analysis for the Survey of Income Program Participation in the United States, concluding that there is no a dominant method to reduce the non-response bias. This, together with it being the method most resorted to in recent studies, led us to choose a logistic regression.

Strictly following McGuigan *et al.* (1995), for a dichotomous response variable y which indicates the second wave response ($y = 1$ if there is an answer; 0 otherwise), and for a vector $\{x_i\}$ of household characteristics in the first wave, the estimated response probability can be written as:

$$p = \frac{e^{\Sigma \beta_i x_i}}{1 + e^{\Sigma \beta_i x_i}}$$

where $\Sigma \beta_i x_i$ is obtained from the estimates for a logistic regression. Consequently, the adjustment weight associated with each observation is given by $1/p$, inversely proportional to the propensity to respond given a set of household characteristics in the first wave. Thus, households having characteristics such that they observe a high p, will have an adjustment factor close to 1, while houses with characteristics associated with non-response (low p) will have a higher factor.

Watson (2004), in adjusting weightings to non-response in the second wave of HILDA, the Australian panel survey, adopted a minimum value of $\hat{p}$ equals to 0.3 to avoid

[4] In the Panel CASEN 1996-2001 Survey methodology, there is no direct reference to the construction of weights following the standard steps of any longitudinal survey; that is, on the basis of the inverse of the probability of selection of the households in the sample. The methodology only mentions the construction of population weights such as the ratio between the projected population at 2001 and the households that were actually surveyed.

extreme values. We imputed that value to 118 households that responded in the second wave. In this way, the weights were corrected as follows:

$$w_{wave\ 2} = \frac{w_{wave\ 1}}{\hat{p}}.$$

For subsequent analysis on the Panel CASEN data, the functional form for correcting longitudinal (2001) weights can be found in Appendix 1.

4 Results

4.1 Detecting attrition bias

A first exercise was to compare the mean values of major socioeconomic outcomes from the group of households that were re-interviewed and those lost to follow-up in 2001. This was done on a variable to variable basis for 1996, providing a first idea of the characteristics of attrition. Of course, this method does not allow us to know the impact each variable has on attrition since they covariate. The analysis is completed through regressions models.

Table 2

Differences in means remaining and lost households

	Re-interviewed households		Lost to follow-up households			
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Difference**	**t-test**
Age family head	45,380	17,399	45,872	17,762	-0,492	(-1,935)
Gender family head (1 = female)	0,225	0,418	0,273	0,446	-0,049	(-1,492)
Single parent household (1 = single parent)	0,282	0,450	0,292	0,455	-0,011	(-0,103)
Household size	4,207	1,730	3,657	1,692	0,551**	(11,121)
Number of children 15 years old or older	1,334	1,224	1,038	1,170	0,296**	(8,725)
Schooling Family head	8,821	4,416	10,418	4,902	-1,597**	(-12,546)
Educationd family head						
Complete basic ed. (1 = complete basic ed.)	0,382	0,486	0,273	0,446	0,109**	(8,118)
Complete high ed. (1 = complete medium ed.)	0,417	0,493	0,449	0,498	-0,032**	(-2,521)
Complete technical ed. (1 = complete technical ed.)	0,062	0,240	0,061	0,239	0,001	(-0,488)
Complete universitary ed. (1 = complete universitary ed.)	0,059	0,236	0,157	0,364	-0,098**	(-12,734)
Housing (1 = own housing)	0,749	0,434	0,465	0,499	0,284**	(20,489)
Labor contract (1 = has contract)	0,396	0,489	0,449	0,498	-0,053**	(-4,180)
Proportion of household employed	0,375	0,252	0,439	0,299	-0,064**	(-8,027)
Location of the household						
3rd region (1 = 3rd region)	0,030	0,171	0,020	0,139	0,011**	(2,200)
7th region (1 = 7th region)	0,107	0,310	0,084	0,278	0,023**	(2,576)
8th region (1 = 8th region)	0,242	0,428	0,182	0,386	0,059**	(4,328)
13th region (1 = 13th region)	0,620	0,485	0,714	0,452	-0,093**	(-6,166)
Zone (1 = rural)	0,121	0,327	0,077	0,266	0,045**	(4,584)
Household per capita autonomous income	241867	509144	339953	555403	-98085**	(-7,359)
Observations	4042		1581			

Notes: (1) ** indicates significance at the 5 percent level or less, and * at the 10 percent level. (2) Values of two-sample t-test with unequal variances are given in parenthesis in the last column.

Results in Table 2 suggest there is a direct relation between a better socioeconomic status and attrition. For example, small households, with less children under fifteen years of age and a higher level of schooling of its head –which on average are richer–, tended to be lost on wave 2. The same is observed in households with better labor conditions and, accordingly, with higher incomes. Therefore, poor households seem to be overrepresented and rich ones underrepresented in the panel. The relation between these variables is a first piece of evidence to think that mobility results might be distorted by attrition bias, as a result of loosing of the sample richer households.

To test the effect on attrition of each variable conditional to the rest, probit models were adjusted using a flexible form. This way, we estimated the probability of not re-interviewing a household on several important variables. A third method used to detect attrition was the BGLW test, which provides the inverse approach to the probits. In the BGLW test, it is evaluated whether the coefficients of a regression on a major outcome differ systematically over the group of re-interviewed households and those lost in the second wave. The results of these two tests are shown in Tables 3 and 4, respectively.

Table 3 shows the estimates of the three regressions adjusted. They have different variables to verify how robust the findings are with respect to the specification. The variables “age of household head”, “household size” and “housing tenancy”, the same as “household head schooling” and “income” systematically explain the probability of not re-interviewing a household. This suggests we are in front of attrition bias, but we cannot, on the basis of the form it acquires, have any degree of certainty as to how it affects mobility estimates for instance.

Table 4 shows the BGLW test which distinguishes both the global effect of attrition on an income regression, as the effect of each individual variable on that regression. We employed income regressions since income is a main outcome of the survey. The results show that attrition is significant in explaining global differences in the parameters and that there is a set of independent variables that explain attrition by themselves.

All together, these three tests suggest that households with no tenure of housing, fewer children, higher levels of schooling for its household head, labor contract and higher levels of income were systematically lost in the second wave of the survey. These are

characteristics of wealthier households which, consecutively, tend to remain static in their income position over time. Therefore, selective attrition in the Panel CASEN might be overstating mobility in Chile.

Table 3

Marginal effects for the probability of attrition between waves (probit)

Variable	Reg. 1	Reg. 2	Reg. 3
Age family head	0,004***	0,003***	0,003***
	(4,43)	(4,07)	(3,98)
Gender family head (1 = female)	0,046	0,033	0,033
	(1,50)	(1,07)	(1,09)
Single parent household (1 = single parent)	-0,027	-0,024	-0,018
	(0,89)	(0,82)	(0,60)
Household size	-0,018**	-0,023***	-0,031***
	(1,97)	(2,63)	(3,36)
Number of children 15 years old or older	-0,020	-0,013	-0,008
	(1,39)	(0,90)	(0,57)
Schooling Family head	0,016***	0,012**	0,010*
	(2,80)	(2,14)	(1,81)
Education family head			
Complete high ed. (1 = complete medium ed.)	-0,020	-0,014	-0,018
	(0,52)	(0,37)	(0,46)
Complete technical ed. (1 = complete technical ed.)	-0,037	-0,032	-0,046
	(0,64)	(0,55)	(0,82)
Complete universitary ed. (1 = complete universitary ed.)	0,063	0,079	0,055
	(0,78)	(0,97)	(0,72)
Housing			
Own or paying	-0,299***	0,033	-
	(12,09)	(0,43)	
Sharing	-	0,012	-0,008
		(0,10)	(0,09)
Renting	-	0,488***	0,456***
		(5,66)	(13,89)
Yielded for services	-	0,358***	0,331***
		(3,68)	(6,01)
Yielded by family	-	0,199**	0,174***
		(2,25)	(4,65)
Irregular occupation	-	-	-0,028
			(0,38)
Labor contract (1 = has contract)	0,007	-0,005	0,005
	(0,30)	(0,21)	(0,20)
Proportion of household employed	0,060	0,055	0,003
	(1,39)	(1,27)	(0,07)
Location of the household			
3rd region (1 = 3rd region)	-0,094***	-0,092***	-0,087**
	(2,69)	(2,59)	(2,47)
7th region (1 = 7th region)	-0,034	-0,036	-0,026
	(1,25)	(1,33)	(0,98)
8th region (1 = 8th region)	-0,041*	-0,043*	-0,036
	(1,78)	(1,84)	(1,55)
Zone (1 = rural)	-0,041	-0,015	-0,011
	(1,32)	(0,48)	(0,33)
Household per capita permanent autonomous income	0,000	0,000	-
	(1,18)	(0,92)	
Income quintile			
Quintile 1	-	-	-0,129***
			(3,46)
Quintile 2	-	-	-0,106***
			(2,94)
Quintile 3	-	-	-0,065*
			(1,80)
Quintile 4	-	-	-0,109***
			(3,35)
Observations	5623	5623	5623
Pseudo R2	0,1177	0,1472	0,1525

Notes: (1) *** indicates significance at the 1 percent level or less, and ** and * at the 5 and 10 percent level, respectively. (2) Robust z statistics in parentheses:

Table 4

BGLW test for log income

	Log Income		
Variable	**Re-interviewed**	**Lost**	**F-test differences**
Age family head	0,095***	0,039***	[96,71]***
	(30,44)	(7,68)	
Gender family head (1 = female)	-1,382***	-2,451***	[17,17]***
	(9,01)	(11,26)	
Single parent household (1 = single parent)	-1,821***	0,589***	[85,64]***
	(12,29)	(2,62)	
Household size	-0,119***	-0,111**	[0,02]
	(4,09)	(2,31)	
Schooling Family head	0,083***	0,165***	[3,56]
	(3,45)	(4,32)	
Educationd family head			
Complete high ed. (1 = complete high ed.)	0,665***	0,293	[1,27]
	(3,80)	(0,99)	
Complete technical ed. (1 = complete technical ed.)	1,133***	0,008	[5,08]**
	(4,29)	(0,02)	
Complete universitary ed. (1 = complete universitary ed.)	1,246***	0,660	[0,96]
	(3,49)	(1,31)	
Housing (1 = own housing)	-0,450***	-0,239	[1,21]***
	(3,97)	(1,48)	
Labor contract (1 = has contract)	1,665***	0,977***	[13,62]***
	(15,92)	(6,01)	
Proportion of household employed	1,704***	1,224***	[2,38]
	(8,89)	(4,78)	
Location of the household			
3rd region (1 = 3rd region)	-0,697**	-0,224	[0,70]
	(2,55)	(0,43)	
7th region (1 = 7th region)	-0,199	0,241	[1,96]
	(1,22)	(0,85)	
8th region (1 = 8th region)	-0,051	-0,187	[0,39]
	(0,44)	(0,97)	
Zone (1 = rural)	0,453***	0,492	[0,01]
	(2,84)	(1,59)	
Constant	5,465***	7,650***	
	(17,42)	(16,03)	
Observations	4042	1581	
R2	0,42	0,26	
F-test for attrition			
Difference in all variables without intercept	[15,77]***		
Difference in all variables including intercept	[14,88]***		

Notes: (1) *** indicates significance at the 1 percent level or less, and ** and * at the 5 and 10 percent level, respectively. (2) Absolute value of t statistics are in parentheses. (3) F-values of tests are in brackets.

4.2 Correction for attrition bias and mobility results

The results of the correction for attrition bias are shown in Table 5 in three groups of columns. Each of these shows the mean values and standard deviations of variables of especial interest obtained with the different weights –original and corrected. In the first group of columns, we show the means of the variables at 1996 with the original 1996 weights. In second place, we show the means of the variables at 1996 of the observations that were re-interviewed in 2001 weighted by the original 2001 weights;

Table 5

Differences in means: initial and final sample-original and corrected weights

	Initial sample (1996)		Final sample (2001)				t-tests differences	
			Original weights		Corrected weights		Initial sample - final sample original w.	Initial sample - final sample corrected w.
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**t**	**t**
Age family head	45,543	17,520	45,916	16,964	45,970	17,392	(-1,046)	(-1,186)
Gender family head (1 = female)	0,241	0,428	0,255	0,436	0,232	0,422	(-1,591)	(1,031)
Single parent household (1 = single parent)	0,285	0,452	0,295	0,456	0,281	0,449	(-1,042)	(0,520)
Household size	4,024	1,737	4,139	1,758	4,050	1,711	(-3,202)**	(-0,724)
Number of children 15 years old or older	1,235	1,214	1,288	1,209	1,249	1,201	(-2,119)**	(-0,545)
Schooling Family head	9,367	4,650	8,648	4,391	9,195	4,596	(7,459)**	(1,750)
Education family head								
Complete basic ed. (1 = complete basic ed.)	0,346	0,476	0,397	0,489	0,356	0,479	(-5,197)**	(-1,034)
Complete high ed. (1 = complete medium ed.)	0,428	0,495	0,413	0,492	0,430	0,495	(1,392)	(-0,287)
Complete technical ed. (1 = complete technical ed.)	0,061	0,240	0,054	0,225	0,063	0,244	(1,580)	(-0,408)
Complete universitary ed. (1 = complete universitary e	0,092	0,289	0,058	0,235	0,082	0,274	(6,020)**	(1,716)
Housing (1 = own housing)	0,654	0,476	0,723	0,447	0,673	0,469	(-7,213)**	(-1,901)
Labor contract (1 = has contract)	0,413	0,492	0,397	0,489	0,412	0,492	(1,593)	(0,170)
Proportion of household employed	0,396	0,270	0,379	0,256	0,387	0,262	(3,250)**	(1,638)
Location of the household								
3rd region (1 = 3rd region)	0,027	0,161	0,030	0,170	0,027	0,163	(-0,867)	(-0,130)
7th region (1 = 7th region)	0,100	0,300	0,100	0,300	0,102	0,303	(-0,090)	(-0,408)
8th region (1 = 8th region)	0,222	0,416	0,222	0,415	0,221	0,415	(0,039)	(0,089)
13th region (1 = 13th region)	0,651	0,477	0,648	0,478	0,649	0,477	(0,322)	(0,225)
Zone (1 = rural)	0,107	0,309	0,100	0,300	0,112	0,315	(1,105)	(-0,770)
Household per capita autonomous income	274511	526983	229298	485407	270561	567569	(4,299)**	(0,352)
Observations	5623		4042		4042			

Notes: (1) ** indicates significance at the 5 percent level or less, and * at the 10 percent level. (2) Values of two-sample t-test with unequal variances are given in parenthesis in the last two columns.

and in the third group of columns of the table, we show the same means at 1996 of the observations re-interviewed in 2001, but re-weighted with the corrected for attrition bias weights (i.e. the corrected weights). This should enable us to reconstruct an unbiased sample.

As a matter of fact, for all variables showing attrition, the problem disappears by re-weighting the observations. The results of the mean difference tests are shown in the last pair of columns. As re-weighting is not done variable by variable, but has to be done for the entire set of variables. This way we calculated proper weights for conducting analysis over the survey.

However, the question we are concerned with is whether the existence of attrition has any effect on the policy recommendations. We consider, as an example, two questions concerning the magnitude and nature of mobility we measure through indexes, transition matrices and regressions. In the first place we re-calculate a set of indices associated with income distribution and mobility for the period in question. Table 6 shows these results.

The attrition correction affects both, the Gini coefficient and the Shorrock Rigidity Index, suggesting that mobility, though always high –especially in lower deciles–, is worse than that derived in studies that do not consider the attrition problem. This lower mobility is also reflected in the annex, Tables A1 and A2, for instance in the percentage of families remaining in the decile 10^{th}.

In the second place, we estimate a multinomial regression model as in Paredes and Zubizarreta (2005) which considers transitions between the status of indigence, poverty and no poverty and we present it also comparatively (Tables A4 and A5). In this case, and most certainly in most of the studies that use this type of methodology, the results do differ. Specifically, in the case of the study that we compare, variables such as the quintile on income in which initially is the state of health of the household head and type or level of schooling received, have significantly different impacts.

Table 6

Shorrocks Rigidity Index

	Original weights	Corrected weights
Gini Coeff 1996	0,54	0,54
Gini Coeff 2001	0,53	0,55
Average Gini	0,50	0,52
Mean income 1996 (Ch. $)	91.796	99.120
Mean income 2001 (Ch. $)	118.768	137.991
Shorrocks Rigidity Index	0,93	0,95

5. Conclusions

We analyzed the nature and extent of attrition in Chile's main longitudinal household survey, the Panel CASEN 1996-2001. First we estimated the magnitude of attrition using standard methods, finding a significant effect of bias. Then we corrected this effect by re-weighting the observations, to evaluated its effect on some results of income mobility studies. One of our main conclusions is that correcting for attrition bias changes a number of these results and, consequently, could alter policy recommendations. For instance, without the correction for attrition bias, Chile appears to have a more rigid income distribution, given that wealthier households were systematically lost.

The pertinence of these results for public policy is evident though panel surveys are appropriate tools for policies aimed at the poor and attaining greater equity. However, specific issues such as non-random attrition must be given especial attention. For that reason, in first place it is well worth proceeding rigorously in the initial sampling stages to limit post-survey attrition problems. The evidence in the case of the most relevant panel survey conducted in Chile suggests that part of the problems are a consequence of the fact that it was not designed as a panel from its beginning, therefore potential attritors were not overrepresented. In addition, the quality of a panel also depends on the fact the initial sample be designed appropriately and the investment in maintenance programs of the members of a panel. None of these stages were contemplated in the survey. In second place, the results arrived at should involve an effort to correct the attrition problem which, independently from the selection of the sample and its follow-up, tends to be inevitable.

Ultimately, considering that the implementation of panel surveys is vital to answer questions regarding the most fundamental characteristics of the people and that they shed light on certain aspects which others do not provide with reference to how to attack the problems of poverty, the existence of attrition biases should not discourage the collection of longitudinal data in LCD's. Quite on the contrary, in this work we have shown that these problems can be limited and solved, providing new insight for studies and policy.

References

Alderman, H., J. Behrman, H. Kohler, J. Maluccio, y S. Watkins, 2001, "Attrition in longitudinal household survey data," *Demographic Research*, 5 (4): 80-123.

Aguilar, O., 2002, "Dinámica de la Pobreza: Resultados de la Encuesta PANEL 1996-2001", www.mideplan.cl , Ministry of Planification and Cooperation.

Ashenfleter, O., Deaton, A. y Solon, G., 1986, "Collecting panel data in development countries: does it make sense?" LSMS Working Paper 23, Washington, DC: The World Bank.

Baulch, B. y Hoddinott, J., 2000, "Economic mobility and poverty dynamics in developing countries" *The Journal of Developmente Studies*, August 2000: 1-24.

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

Castro, R. and F. Kast, 2004,, "Movilidad de la Pobreza en Chile. Análisis de la Encuesta PANEL 1996-2001", Serie Informe Social N° 85, Fundación Libertad y Desarrollo.

Contreras, D., R. Cooper, J. Herman, and C. Neilson, 2004, "Dinámica de la Pobreza y Movilidad Social: Chile 1996-2001", Documento de Trabajo, Universidad de Chile, Departamento de Economía.

Fitzgerald, J., P. Gottschalk, y R. Moffitt, 1998, "An analysis of sample attrition in panel data," *The Journal of Human Resources*, 33 (2): 251-299.

Graham, J. W. y Donaldson, S. I., 1993, "Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data," Journal of Applied Psychology, 78 (1): 119-128.

Kalton, G., y M. Brick, 2000, "Weighting in household panel surveys" in Rose, D. (ed.), *Researching Social and Economic Change: the uses of household panel studies*, 96-111, London and New York: Routledge.

Lynn, Peter, Buck, Nicholas, Burton, Jonathan, Jäckle, Annette, and Laurie, Heather (December 2005) 'A Review of Methodological Research Pertinent to Longitudinal Survey Design and Data Collection', Working Papers of the Institute for Social andEconomic Research, paper 2005-29. Colchester: University of Essex.

McGuigan, K. A., P. L. Ellickson, R. D. Hays y R. M. Bell, 1995, "Tracking, weighting, and simple selection modeling to correct for attrition", RAND.

Paredes, R. y J. R. Zubizarreta, 2005, "Focusing on the Extremely Poor: Income Dynamics and Policies in Chile", Documento de Trabajo 183, Departamento de Ingeniería Industrial y de Sistemas, Pontificia Universidad Católica de Chile.

Peracchi, F., 2002, "The European Community Household Panel: A review", *Empirical Economics*, 27, pp. 63-90.

Rizzo, L., G. Kalton y J. M. Brick, 1996, "A comparison of some weighting adjustments methods for panel nonresponse", *Survey Methodology* 22, 43-53.

Rose, D., 2000, "Household panel studies" in Rose, D. (ed.), *Researching Social and Economic Change: the uses of household panel studies*, 3-35, London and New York: Routledge.

Teitelboim, B., A. Oyarzún, y O. Aguilera, 2002, "Metodología Encuesta Panel 1996-2001", Gobierno de Chile, Ministerio de Planificación y Cooperación, División Social.

Watson, N., 2004, "Wave 2 weighting", HILDA *Project Technical Paper Series* NO. 4/04.

Winkels, J., y S. Withers, 2000, "Panel attrition" in Rose, D. (ed.), *Researching Social and Economic Change: the uses of household panel studies*, 79-95, London and New York: Routledge.

Winkels, J. y Davies, S., 2000, "Panel attrition" in Rose, D. (ed.), *Researching Social and Economic Change: the uses of household panel studies*, 79-95, London and New York: Routledge.

Yaqub, S., 2000, "Poverty Dynamics in Developing Countries", Development Bibliography 16, Institute of Development Studies, University of Sussex.

Appendix 1

To correct wave 2 weights for attrition bias we suggest using the following formula

$$w_{wave\ 2} = \frac{w_{wave\ 1}}{\hat{p}},$$

where $\hat{p}$ is the estimated probability of responding in wave 2, given by

$$\hat{p} = \frac{e^{\sum \hat{\beta}_i x_i}}{1 + e^{\sum \hat{\beta}_i x_i}}.$$

The estimated $\hat{\beta}_i$ coefficients, in turn, are given in the following table

Table A1

Variable	Coefficient
Age family head	-0,00097
Square age family head	-0,00015
Gender family head (1 = female)	-0,22776
Single parent household (1 = single parent)	0,13566
Household size	0,14036
Number of children 15 years old or older	0,06448
Schooling family head	0,10000
Square schooling family head	-0,01006
Education family head	
Complete high ed. (1 = complete medium ed.)	0,09643
Complete technical ed. (1 = complete technical ed.)	0,38211
Complete universitary ed. (1 = complete universitary ed.)	0,51702
Housing (1 = own housing)	1,44070
Labor contract (1 = has contract)	-0,08315
Proportion of household employed	0,00273
Location of the household	
3rd region (1 = 3rd region)	0,45210
7th region (1 = 7th region)	0,14587
8th region (1 = 8th region)	0,15638
Zone (1 = rural)	0,25895
Per capita income quintile	
Quintile 1 (1 = quintile 1)	0,85715
Quintile 2 (1 = quintile 2)	0,67436
Quintile 3 (1 = quintile 3)	0,40003
Quintile 4 (1 = quintile 4)	0,57414
Constant	-0,76546

Therefore, on the original Panel CASEN data base called "panel hogar", this correction should be applied truncating $\hat{p}$ to 0.3 if greater than, to avoid extreme values.

Appendix 2

Table A2

Transition matrix by deciles of the income distribution: original weights

	2001 decile										
1996 decile	1	2	3	4	5	6	7	8	9	10	Total
1	2,95	2,08	1,6	0,73	0,55	0,63	0,33	0,21	0,17	0,17	9,42
2	1,47	2,34	1,47	1,13	0,87	0,29	0,18	0,27	0,1	0,02	8,14
3	1,64	1,51	1,59	1,79	0,93	0,82	0,76	0,34	0,07	0,18	9,62
4	1,25	0,95	1,48	1,13	1,29	1,4	1,05	0,55	0,6	0,12	9,82
5	0,58	1,21	0,94	1,53	1,32	1,22	1,28	0,81	0,54	0,12	9,55
6	0,63	0,8	0,94	1,1	1,27	1,47	1,35	1,57	1,22	0,37	10,74
7	0,52	0,39	0,84	0,92	1,2	1,41	1,91	1,97	0,98	0,56	10,69
8	0,45	0,31	0,92	0,65	1,52	1,26	1,51	2,05	1,88	1,15	11,69
9	0,21	0,23	0,31	0,57	0,67	0,92	1,03	1,73	2,07	2,05	9,78
10	0,33	0,23	0,03	0,28	0,4	0,55	0,61	0,54	2,35	5,23	10,54
Total	10,02	10,05	10,13	9,83	10,02	9,96	10	10,05	9,98	9,96	100

Table A2

Transition matrix by deciles of the income distribution: corrected weights

	2001 decile										
1996 decile	1	2	3	4	5	6	7	8	9	10	Total
1	3	2,18	1,39	0,53	0,91	0,3	0,15	0,17	0,18	0,02	8,83
2	1,62	3,14	1,42	1,13	0,54	0,26	0,26	0,16	0,13	0,01	8,67
3	1,41	1,29	1,75	1,89	1,05	0,7	0,5	0,18	0,12	0,17	9,06
4	1,21	0,97	1,23	1,31	1,62	0,99	0,99	0,76	0,33	0,02	9,45
5	0,94	0,53	1,06	1,08	1,19	2,97	0,85	0,56	0,35	0,11	9,64
6	0,45	0,71	1,06	1,09	1,15	1,58	1,08	1,51	1,27	0,21	10,11
7	0,36	0,43	0,78	1	1,51	1,99	1,06	1,35	0,81	0,26	9,56
8	0,41	0,29	0,62	0,87	0,77	1,3	1,52	1,82	1,66	1,32	10,58
9	0,17	0,17	0,61	0,51	1,01	0,65	1,63	2,52	2,61	1,7	11,58
10	0,54	0,2	0,14	0,56	0,4	0,44	0,65	0,87	2,56	6,16	12,51
Total	10,12	9,9	10,06	9,98	10,15	11,19	8,7	9,9	10,02	9,98	100

Appendix 3

Table A4

Multinomial Probit Model: transitions from Indigence, Poverty Not Indigent, and Not Poverty original weights

		From I		From PNI		From NP	
	VARIABLE	**To PNI**	**To NP**	**To I**	**To NP**	**To I**	**To PNI**
Family Composition	Age of the Head of Family	-0,008	0,002	-0,013	0,001	-0,03	-0,026
		(-0,63)	(-0,18)	(-0,86)	(-0,12)	(3,43)***	(3,82)***
	Gender of Family Head (dummy, 1 = female)	0,505	0,714	0,008	-0,241	-1151	-0,663
		(-1,42)	(1,78)*	(-0,02)	(-0,70)	(4,42)***	(3,34)***
	% children < 5 years old	-3932	-5516	2955	0,114	2112	0,652
		(3,24)***	(4,53)***	(2,70)***	(-0,10)	(1,96)**	(-0,84)
	% people > 65 years old	-1303	0,106	1932	1997	-0,234	-1021
		(1,68)*	(-0,11)	(2,21)**	(2,51)**	(-0,63)	(2,64)***
Physical Capital	Housing medium (dummy)	0,789	0,771	-0,438	-0,231	0,224	0,159
		(2,58)***	(2,39)**	(-1,31)	(-0,74)	(-0,79)	(-0,67)
	Housing good (dummy)	1010	1047	0,117	0,266	-0,563	-0,199
		(3,12)***	(3,21)***	(-0,35)	(-0,89)	(2,07)**	(-0,92)
Human Capital	Head of Family schooling	-0,046	-0,015	-0,022	0,001	-0,091	0,017
		(-0,84)	(-0,26)	(-0,37)	(-0,01)	(2,14)**	(-0,53)
	Spouse schooling	0,12	0,086	0,002	-0,018	-0,048	-0,043
		(2,78)***	(1,89)*	(-0,06)	(-0,47)	(1,80)*	(2,20)**
Working Capital	Labor contract (dummy, 1 = has contract)	0,023	-0,244	0,418	0,099	-0,488	-0,207
		(-0,08)	(-0,77)	(-1,63)	(-0,43)	(2,31)**	(-1,17)
	% house employed	1820	4622	-1074	2839	-3624	-3049
		(1,71)*	(4,29)***	(-0,83)	(3,30)***	(5,59)***	(7,63)***
Home Environment	3rd Region (dummy)	0,003	-0,62	0,526	0,089	0,279	0,117
		(0,00)	(-0,94)	(-1,07)	(-0,22)	(-0,73)	(-0,38)
	7th Region (dummy)	0,188	-1052	-0,275	-0,508	0,012	-0,483
		(-0,40)	(2,08)**	(-0,73)	(-1,61)	(-0,04)	(1,94)*
	8th Region (dummy)	-0,309	-1295	0,031	-0,528	0,239	-0,34
		(-0,63)	(2,74)***	(-0,08)	(1,71)*	(-0,94)	(-1,45)
	Urban Zone (dummy)	-0,641	-1733	0,378	-0,211	0,598	0,296
		(2,04)**	(5,57)***	(-1,09)	(-0,80)	(2,18)**	(-1,47)

continues

continuation

	VARIABLE	From I		From PNI		From NP	
		To PNI	**To NP**	**To I**	**To NP**	**To I**	**To PNI**
	% poor houseolds in Municipality	0,253	1858	-1160	-0,122	1504	1169
		(-0,23)	(1,70)*	(-1,12)	(-0,15)	(1,70)*	(-1,52)
Shocks	Household become single parent (dummy)	2126	1864	0,014	-0,43	0,5	0,235
		(2,83)***	(2,58)***	(-0,03)	(-1,31)	(1,67)*	(-1,01)
	Change in % children < 5 years old	-0,046	-0,214	-0,106	-0,149	0,065	0,201
		(-0,58)	(2,69)***	(1,75)*	(2,56)**	(-1,22)	(3,94)***
	Change in % older than 65 years old	-2523	-3065	1561	0,162	0,956	1034
		(2,16)**	(3,03)***	(-1,60)	(-0,17)	(-1,01)	(1,73)*
	Head of Family suffered health problem (dummy)	-0,924	-0,101	-0,682	-0,577	-0,659	-0,496
		(1,97)**	(-0,25)	(-1,56)	(-1,48)	(2,26)**	(1,85)*
	Change in % employed in household	2000	4062	-1526	3168	-4679	-3237
		(2,14)**	(3,49)***	(-1,37)	(5,77)***	(7,42)***	(8,67)***
	Change in province income (Ch$ 000)	0,001	0,002	0,001	-0,005	-0,003	0,002
		(-0,15)	(-0,54)	(-0,18)	(-1,31)	(-1,00)	(-0,57)
Permanent Income	Permanent Income (Ch $ 000)	-0,001	0,003	-0,001	0	0	-0,001
		(-0,40)	(2,34)**	(-1,07)	(-0,11)	(-1,06)	(2,93)***
Adjusted Quintile	Adjusted Quintile 1 (dummy)	-	-	-	-	0,695	1032
		-	-	-	-	(-1,61)	(3,47)***
	Adjusted Quintile 2 (dummy)	-	-	-	-	0,438	0,617
		-	-	-	-	(-0,99)	(2,00)**
	Adjusted Quintile 3 (dummy)	-	-	-	-	0,291	0,148
		-	-	-	-	(-0,70)	(-0,47)
	Adjusted Quintile 4 (dummy)	-	-	-	-	-0,405	-0,13
		-	-	-	-	(-0,94)	(-0,36)
	Constant	-0,391	-1532	0,844	0,767	0,095	0,582
		(-0,43)	(1,72)*	(-0,77)	(-0,95)	(-0,11)	(-0,86)
NUMBER OF OBSERVATIONS		**507**		**825**		**2706**	

Table A5

Multinomial Probit Model: transitions from Indigence, Poverty Not Indigent, and Not Poverty corrected weights

		From I		From PNI		From NP	
	VARIABLE	**To PNI**	**To NP**	**To I**	**To NP**	**To I**	**To PNI**
Family Composition	Age of the Head of Family	0,004	-0,013	-0,023	0,017	-0,039	-0,015
		0,32	1,13	1,53	1,57	4,85***	2,41**
	Gender of Family Head (dummy, 1 = female)	0,641	1,107	-0,134	0,084	-1,309	-0,385
		1,76*	2,48**	0,34	0,28	4,46***	2,02**
	% children < 5 years old	-2,826	-3,445	1,636	0,17	1,126	1,226
		2,63***	3,09***	1,49	0,15	1,18	1,83*
	% people > 65 years old	3,223	5,435	-2,307	0,094	-0,726	-1,023
		2,03**	3,45***	1,53	0,2	1,15	3,02***
Physical Capital	Housing medium (dummy)	0,218	0,668	0,12	0,048	-0,078	0,215
		0,61	1,82*	0,32	0,12	0,29	0,96
	Housing good (dummy)	1,037	0,897	-0,097	-0,14	-0,593	0,102
		3,06***	2,53**	0,24	0,4	2,46**	0,5
Human Capital	Head of Family schooling	-0,046	-0,018	0,018	0,05	-0,05	-0,03
		1	0,38	0,35	1,23	1,59	1,2
	Spouse schooling	0,062	0,088	-0,061	-0,035	-0,054	-0,019
		1,49	2,01**	1,61	1,19	2,36**	1,2
Working Capital	Labor contract (dummy, 1 = has contract)	-0,293	0,414	0,582	0,105	-0,434	0,083
		0,88	1,21	2,00**	0,39	2,11**	0,48
	% house employed	1,593	5,359	-2,813	3,767	-4,98	-3,233
		1,6	5,81***	2,52**	5,11***	7,70***	8,91***
Home Environment	3rd Region (dummy)	-0,597	-1,407	0,306	-0,218	-0,072	-0,009
		0,9	1,89*	0,59	0,54	0,19	0,03
	7th Region (dummy)	-0,421	-1,43	-0,214	-0,643	-0,245	-0,245
		0,86	2,82***	0,64	1,86*	0,73	1,09
	8th Region (dummy)	-0,174	-1,567	-0,377	-0,591	-0,093	-0,208
		0,35	2,94***	1,04	1,6	0,38	1,03
	Urban Zone (dummy 1)	-0,75	-2,04	-0,228	-0,394	0,34	0,583
		1,93*	5,41***	0,66	1,46	1,34	2,99***

continues

continuation

	VARIABLE	From I		From PNI		From NP	
		To PNI	**To NP**	**To I**	**To NP**	**To I**	**To PNI**
	% poor houseolds in Municipality	0,864	3,335	-0,894	0,49	1,431	0,261
		0,56	2,18**	0,73	0,39	1,21	0,32
Shocks	Household become single parent (dummy)	20,231	21,86	1,033	-0,492	0,237	0,357
		21,75***	,	1,97**	0,93	0,42	0,92
	Change in % children < 5 years old	-0,274	-0,36	-0,096	-0,038	0,008	0,103
		3,12***	3,95***	1,25	0,67	0,08	2,08**
	Change in % older than 65 years old	-2,7	-0,823	1,685	1,733	0,668	0,93
		2,03**	0,61	1,2	1,21	0,55	1,18
	Head of Family suffered health problem (dummy)	-0,018	0,109	-0,64	-0,929	0,071	-0,114
		0,05	0,27	1,32	2,67***	0,2	0,64
	Change in % employed in household	1,895	4,716	-4,129	3,204	-4,882	-3,278
		2,43**	5,77***	3,96***	5,51***	7,48***	9,12***
	Change in province income (Ch$ 000)	-0,005	0,007	0,007	0,003	0,003	0,001
		1,02	1,41	1,46	0,91	1,18	0,49
Permanent Income	Permanent Income (Ch $ 000)	-0,001	0,003	-0,001	0	0,001	-0,002
		0,75	2,63***	0,83	0,61	2,05**	4,32***
Adjusted Quintile	Adjusted Quintile 1 (dummy)	-	-	-	-	0,301	0,525
		-	-	-	-	0,74	1,52
	Adjusted Quintile 2 (dummy)	-	-	-	-	-0,42	0,262
		-	-	-	-	1,11	0,78
	Adjusted Quintile 3 (dummy)	-	-	-	-	-0,317	0,177
		-	-	-	-	0,76	0,51
	Adjusted Quintile 4 (dummy)	-	-	-	-	-0,567	0,246
		-	-	-	-	1,2	0,68
	Constant	0,073	-0,573	0,912	-0,663	1,35	0,498
		0,09	0,62	0,97	0,83	1,92*	0,81
	NUMBER OF OBSERVATIONS	**421**		**885**		**2736**	