INTERNATIONAL POLICY CENTER
Gerald R. Ford School of Public Policy
University of Michigan

IPC Working Paper Series Number 36

ANALYSIS OF ATTRITION PATTERNS IN THE TURKISH HOUSEHOLD LABOR FORCE SURVEY, 2000-2002*

İnsan Tunalı

May 4, 2007
*

ANALYSIS OF ATTRITION PATTERNS IN THE TURKISH HOUSEHOLD LABOR FORCE SURVEY, 2000-2002*

İnsan Tunalı
Dept. of Economics
Koç University
İstanbul 34450
Turkey
itunali@ku.edu.tr

May 4, 2007

* Prepared for presentation at the University of Michigan Conference on Labor Markets in Transition and Developing Economies: Emerging Policy and Analytical Issues. Research assistance from Emre Ekinci and funding from TÜBITAK via grant no: 106K160 are gratefully acknowledged. I'd like to thank Dr. Ömer Demir, the president of Turkish Statistical Institute (TURKSTAT), for granting me the permission to use the raw data from the Household Labor Force Survey (HLFS) originally obtained for another project for the purposes of this study. I would like to thank TURKSTAT staff Hasibe Dedeş, Didem Sezer and Enver Taştı for their willingness to respond to my endless questions about the HLFS.

ABSTRACT

This paper offers an analysis of attrition patterns in the "New" Turkish Household Labor Force Survey (HLFS) which has been conduced since 2000. The most important feature of the redesigned survey is its short panel component. We use 12 rounds of micro data collected (on a quarterly basis) over the period 2000-2002 and focus on household level attrition within 3, 12 and 15 months of the initial interview. Attrition is a phenomenon which can be attributed to demographic and economic factors, including conditions in the labor market. If attrition is related to labor force status of individuals, this could result in biases in labor market indicators. We provide strong evidence that household attrition is influenced by the labor force status (outside the labor force, employed, or unemployed) of the household head at the initial survey round and discuss the implications.

Keywords: Attrition, labor force status, labor force dynamics, household survey.
JEL codes: J21, J61, C81.

1. INTRODUCTION

Attrition is recognized as a major issue by users of panel data sets (an early example being Hausman and Wise, 1979). Data collection agencies have methods for adjusting for non-response, but attrition (initial response followed by non-response at a later round of the survey) may cause additional problems which are typically not handled well by standard reweighing schemes (Ridder, 1992). There is a large literature on attrition and its consequences in widely used panel data sets. A representative sample may be found in the Spring 1998 special issue of the *Journal of Human Resources*: see, in particular, Fitzgerald et al. (1998), MaCurdy et al. (1998), van den Berg and Lindeboom (1998), Zabel (1998).

Starting with 2000, the Household Labor Force Surveys (HLFS) administered by the Turkish Statistical Institute (TURKSTAT; formerly the State Institute of Statistics, SIS) have been conducted continuously, using a rotating sample frame designed to yield quarterly estimates (SIS, 2001a). The rotation plan calls for a total of four interviews over a period of six quarters. To be precise, the selected household is interviewed in two subsequent quarters, skipped for the next two, and then interviewed again in two subsequent quarters. Thus it is possible to form estimates of quarterly and annual transitions between labor market states. This is a major breakthrough that could allow tracking of labor market dynamics.[1] However, to date only two papers have addressed the subject (Taşçı and Tansel, 2005a,b).

The sampling frame adopted by the New HLFS is address-based. The survey protocol does not require following households (or individual members, so-called splits) who move to another location. Furthermore, if there is a different household at a previously visited address, the newly arrived household is included in the survey. In essence TURKSTAT deals with attrition in the HLFS by using substitute households in place of attritors when available, and reweighing the cross-section sample so that it is representative of the (projected) population. This could be problematic if attrition and/or substitution probabilities depend on labor market states occupied by members of the respondents. In fact Tunalı and Baltacı (2004) have argued that cross-section estimates of standard measures of labor market outcomes (participation rate, unemployment rate, etc.) formed for the period 2000-2002 are biased, on the grounds that the statistics are influenced by the number of times a household has been interviewed.

[1] TURKSTAT officials do not think so. They argue that many European data collection agencies pay repeat visits to the same address not because they intend to exploit the panel dimension of the data, but because they want to enhance the stability of the sample.

Based on the information in the non-response forms filled by the field staff of TURKSTAT, almost all the attrition takes the form of migration rather than refusal to respond. There is good reason (and ample evidence in the labor economics literature) to believe that individuals, even households, respond to labor market conditions by moving. This certainly was the case in the 60s and 70s in Turkey (Tunalı, 1996). It probably was the case in the period following the February 2001 crisis, when Turkey's economic growth rate (as measured by annual changes in real GNP) swung from –9.5 percent between 2000-01 to +7.9 percent between 2001-2 (World Bank, 2005, p.26).

The objective of this paper is to document the patterns in attrition observed in the HLFS over the period 2000-02. Towards that end I examine the likelihood of attrition within 3, 12 and 15 months of the initial survey by focusing on observed characteristics of the household head. I focus on the household head because standard reweighing schemes (such as those used by TURKSTAT) are designed to match the cross-section distributions of observables such as sex, age and education of the household head with those in the population. Since the links between attrition and labor market outcomes are my main concern, I confine my working sample to prime-age household heads, namely 20-54 years-old individuals identified as the household head in the first round of the survey. For these households the cumulative probability of attrition is about 5 percent by 3 months, 12 percent by 12 months, and 19 percent by 15 months. These large magnitudes call for an investigation of the determinants of attrition, so that their implications for labor market statistics can be drawn. In what follows I show convincingly that the labor market state occupied by the household head in the first round influences attrition probability in subsequent rounds, even when I control for a broader set of household characteristics than those used by TURKSTAT. My results send an important lesson to data collection agencies that insist on simplistic reweighing schemes and policy makers who rely on statistics produced in this manner.

A formal statement of the problem and its consequences are provided in section 2. In section 3 I describe the HLFS data, discuss the data related problems and the solutions I adopted. In section 4, I present the estimation methodology. Section 5 contains the empirical results. The concluding section highlights the key findings and their implications.

2. ATTRITION AND ITS CONSEQUENCES

To illustrate the source of the problem, consider a two-round panel and let y_{ij} = labor market state of individual i at round j, $j = 1,2$; x_i = fixed characteristics of individual i; D_i =

1 if individual is present at both rounds, 0 else. For brevity I ignore the subscript for the individual and define $f(y_1, y_2 \mid x, D)$ as the joint distribution of labor market states conditional on x and D. In general $f(y_1, y_2 \mid x, D = 1) \neq f(y_1, y_2 \mid x)$, a feature which renders the balanced panel problematic for the purposes of drawing inferences on labor market dynamics. The problem can be attributed to the fact that

$$(1) \qquad P(D = 1 \mid y_1, y_2, x) \neq P(D = 1 \mid y_1, x) \neq P(D = 1 \mid x).$$

Equation (1) captures the notion that the attrition process may be influenced by the labor market states occupied by respondents who are observationally identical otherwise (that is, they have the same x). This type of attrition is known as *non-ignorable* attrition (see Rubin, 1976, Little and Rubin, 1987). In this case even cross-section estimates of labor market outcomes could be affected by attrition, because in general

$$(2) \qquad f_2(y_2 \mid x, D = 1) \neq f_2(y_2 \mid x, D = 0) = f_2(y_2 \mid x).$$

Tunalı and Baltacı (2004) provide evidence of non-ignorable attrition in the Turkish HLFS. They focus on three labor market states: not in the labor force, employed, and unemployed. They study the marginal distributions of membership in labor market states $f(y|x)$ in the reweighed cross-section as well as conditional distributions $g(y_2 \mid y_1, x, D = 1)$ which capture the transition probabilities between the three states. They show that all the distributions are influenced by the number of times an individual is observed (controlling for survey round). They also estimate the magnitudes of the biases in the cross-section estimates reported by TURKSTAT by relying on data from individuals who enter the survey sample for the first time, on the assumption that they constitute a 'fresh' sample representative of the population.

The current paper places equation (1) in the limelight. I treat attrition as a choice variable at the household level. I express the attrition probability as a function of household characteristics as well as indicators for the survey round. By including a successively longer list of observables (in x), I illustrate the existence of possible venues for extending the standard reweighing schemes. By including information on the labor market state occupied by the household head in the first round (y_1) as a determinant of attrition, I am able to test for the presence of non-ignorable attrition.

3. DATA AND SUMMARY STATISTICS ON ATTRITION

Household Labor Force Surveys which have nationwide representation have been conducted in Turkey since October 1988. Between 1989 and 1999 the survey was conducted bi-annually, during the months of April and October, with the second full week of the month as the reference week. Reliance on a low sampling frequency and a fixed reference week meant that changes in labor market conditions could not be tracked accurately by the HLFS. The "New" HLFS was designed to respond to this concern and was launched in 2000. It featured a rotating sampling frame (similar to the Current Population Survey conducted in the U.S.A) and a sliding reference week which allows continuous tracking. The design hinges on a total of four visits to the same address, over a period of six quarters. According to the standard pattern, a household is interviewed in two subsequent quarters, allowed to rest for the next two, and returns to the sample for another two. This rotation plan is often abbreviated as "in-in-out-out-in-in" or simply "2-(2)-2." With this rotation plan it is possible to study attrition at three different intervals, namely 3, 12 and 15 months following the initial interview.

There is no question that the switch from the original HLFS to the New HLFS posed challenges for TURKSTAT. Since the surveys had been conducted via Computer Assisted Personal Interviews (CAPI) for some time, the proper infrastructure was already in place. Table A1 in the appendix provides a glimpse of the planning that went into the survey. Each round of the HLFS contains eight subsamples identified by a distinct rotation number. The rotation number determines the number and timing of subsequent visits to the household. In addition, the year and quarter at which each interview round took place is known. Household IDs end with either odd or even numbers, and this assignment is consistent with rotation number and round. Furthermore the visit number is recorded at the time of the survey. With this information in hand, we were able to determine the maximum number of recorded visits as well as the expected number of (total) visits to a given household. A household was classified as an attritor if it did not show up in a subsequent round it was expected to show up at. A similar scheme was used to detect individual household members who attrited.

The survey protocol of the HLFS allows for substitution of an old household by a new one that took residence at the previously visited address between two rounds of the survey. The new household is given a new household ID, but the visit counter is not reset. This practice is consistent with the use of an address-based sampling frame. Since this paper is about attrition patterns, substitute households were excluded from the working sample.[2] It is also possible for a household to leave the address for some time, to return later. We consider them as attritors when they do not show up in the data as scheduled.

Our examination of the raw data revealed that some departures from the survey protocol did take place in the field. In some instances, the ID of a departing household was given to the new household at the old address. In other instances the ID numbers of individual members were messed up as a result of departures from, or new arrivals to, the household. These were identified using a computer program that kept track of changes in the household roster. The former were easy to fix, latter were not. Since the analysis of attrition patterns in the current paper is confined to prime-age household heads, the difficult cases do not concern us here. Finally, there were coding errors in the visit number. Since the rotation number and round determine the visit number, these mistakes were easily corrected.[3]

In this paper I rely on twelve rounds of the HLFS from the period 2000-2002. Each round of the survey includes around 70,000 individuals from 18-20,000 households. The full data set consists of about 900,000 individual records. The rotation plan provides 50 percent overlap in the sample between subsequent quarters and same quarter one year apart. However not all rounds furnish information at all attrition intervals. Firstly, the steady state for the standard rotation plan 2-(2)-2 was not reached until 2001: Q2. By design earlier rounds do not provide information at all attrition intervals. Secondly, I do not have data beyond 2002: Q4. This ushers in censoring. Table A1 in the appendix provides detailed information on the rotation plan. This has implications for the samples (risk sets) we used in our attrition study.

Let $A(m)$ denote the indicator of attrition after m months. We set $A(m) = 1$ if the household is not found at the same address m months after the initial visit, and = 0 otherwise, $m = 3, 12, 15$. By definition, households interviewed for the first time cannot attrit. For notational convenience we set $A(0) = 0$ for everyone. The risk sets $R(m)$, $m = 3, 12, 15$ respectively consist of all households who did not attrit until month m. If household h attrits in month m_h^*, it is excluded from the risk set at higher intervals. That is, household h belongs to risk set $R(m)$ iff $m_h^* > m$. Individual attrition indicators and risk sets are also computed, following the same logic. Table A2 in the appendix shows the risk sets computed in this

[2] Give statistics!!!

[3] My RA Emre Ekinci deserves special credit for writing the STATA code we used for detecting and correcting the errors.

manner, using the information in Table A1, as well as summary statistics on the attrition indicators. Note that technically speaking the pre-steady state rotation plan allows us to study $A(9)$ for a subset of households interviewed in the first quarter of year 2000.

In what follows I study household level attrition patterns for household whose head was 20-54 years old at the time of the first interview. These households form a subset of all households. Households we study are slightly more likely to attrit, but not by much. We focus on the subset because our main objective is to establish the links between labor market status and attrition. Labor market attachment of older household heads is low, and their attrition behavior may have other explanations. As seen in Table A2, incidence of attrition is higher at the individual level, something we expect. However, due to the challenges posed by the coding errors mentioned above, we do not study individual level attrition patterns at this time. From this point on all references to attrition is confined to households with a 20-54 year-old head, and we drop the qualifier and refer to them simply as households for brevity.

My working sample consists of about 50,000 households. The risk sets are respectively 49318 for $A(3)$, 26408 for $A(12)$, and 21855 for $A(15)$. Recall that the rotation plan allows us to compute $A(9)$ as well. However the risk set is considerably smaller (3,329) and unlike the other cases, there is no time-series variation in exposure to attrition risk. Under the circumstances I did not feel confident investigating $A(9)$ behavior. Based on the marginal distributions (see final column of bottom block in Table A2), about 4.7 percent of households attrit by month 3. Subsequent rounds of attrition respectively claim additional 3.1, 5 and 7.7 percent of the survivors. Thus the cumulative probability of attrition is 7.6 percent by 9 months, 12.1 percent by 12 months, and 18.9 percent by 15 months. These magnitudes underscore the importance of our undertaking. By investigating its determinants, we stand to improve our understanding of the implications of attrition for labor market statistics computed on the HLFS data.

The explanatory variables I rely on in the attrition regressions were constructed from a subset of the 56 survey questions which all 12 rounds have in common, using specifications in the literature as a guide. The complete list is given in Table 1 along with some descriptive statistics. All variables are measured at the initial round of the survey.[4] Due to censoring, households (originally) interviewed in 2000 constitute the majority of the households in our

[4] In theory, richer specifications that exploit changes in status (for example, marital status or labor force status) can be estimated. In practice, collinearity is likely to emerge as a serious challenge. We plan to investigate this in a future iteration.

working sample. Those interviewed in 2001 and 2002 respectively account for about 28 and 20 percent of the working sample. Households interviewed in the first quarter have above average, while those interviewed in the fourth quarter have below average representation. Nearly 80 percent of households come from an urban location (defined as having a population of 20,000 or over). In the HLFS sampling frame rural households are underrepresented by design, and this is reflected in our working sample. Arguably labor markets in urban locations are more complex, and the sampling frame strives to capture this.[5]

The average household consists of 4.2 individuals. However there is considerable variation. An overwhelming majority of the households consist of a nuclear family, while 11.6 percent are extended households. More than ninety percent of the household heads are married, and about eight percent are female. Average age of all household heads is about 40, and average education is 7.3 years. At the initial round of the survey 80 percent are employed, while 7 percent are unemployed.

4. METHODOLOGY

In testing for presence of non-ignorable attrition, I follow the approach in Fitzgerald et al. (1998) closely. They conduct two tests, namely the HW and BGLW tests which are in turn attributable to Hausman and Wise (1979) and Becketti et al. (1988). For the HW test, binary outcome equations for attrition status need to be estimated. Under the null hypothesis of ignorable attrition, the coefficient(s) on the lagged value(s) of the labor market state occupied by the individual are zero. For the BGLW test, two binary labor market outcome equations (participate or not, unemployed or not) have to be estimated, as a function of individual and household characteristics, as well as dummies for attrition status in future rounds. Under the null hypothesis of ignorable attrition, the coefficient(s) on attrition dummies should be zero. The current version of the paper is confined to the HW test. I hope to have the BGLW results ready in time for the conference.

Let $A_h(m)$ denote the attrition status of household h as of m months after the initial interview. For m = 3, 12, 15, we estimate Probit models of the form

$$Pr\{A_h(m) = 1 \mid y, x, z; h \in R(m)\} = \Phi[\beta' y(0) + \gamma' x + \delta' z]. \quad (3)$$

All explanatory variables are measured at the initial survey round. Here y is a vector that contains indicators of the labor force status of the household head, x denotes the vector of

[5] I was unable to obtain the sampling weights from TURKSTAT. Since I do not aim to arrive at valid estimates for the population as a whole, this is not a handicap.

other individual and household characteristics, z denotes indicators that identify the survey round, $R(m)$ denotes the risk set, and $\Phi(.)$ denotes the standard normal c.d.f. We estimate the unknown parameters β, γ, δ using maximum likelihood. If the null hypothesis that $\beta = 0$ is rejected, we have evidence that attrition is non-ignorable.

As I argued in the introduction, the survey protocol of HLFS does not call for following movers. Consequently there is strong reason to believe that attrition and migration go hand in hand. Although the results are not published, TURKSTAT officials carefully review the nonresponse forms filled by the field staff. Their impression is that the bulk of attrition is attributable to migration rather than non-response. I shed further light on this issue by estimating models which mimic the specifications used in reduced form migration equations. If determinants of attrition turn out to be the same as the determinants of migration, our expectations will be fulfilled.

5. RESULTS

At each attrition interval, the same set of 5 models was estimated. Each model is nested under the subsequent ones. Model 0 includes year and quarter dummies only. These variables capture the common component of the time-series variation in attrition. In Model 1 a set of household characteristics are added to the model. This list includes variables that TURKSTAT uses for reweighing, such as age, education, sex and location (urban vs. rural), plus indicators of marital status and gender of the household head. In Model 2 two indicators for the labor market status of the household head at the initial round are added. Taking non-participants as the reference category, we explore whether employed and unemployed household heads display different attrition behaviors. In Model 3 a third degree polynomial function of the household size (number of people residing in the household) is added. Arguably large households are less likely to attrit, given the associated costs. In Model 4 we add two dummies for vocational education, respectively at secondary (high school) and tertiary (2 or 3 year university) levels. This model is intended to shed further light to the debate over the value of vocational education in the labor market (see Tunalı, 2005).

Complete results from Probit estimates of attrition probability at 3, 12, and 15 month intervals are reported in Tables A3-A5 collected in the appendix. Here we focus on the key findings using summary tables. The tests of the null hypothesis that attrition is ignorable ($\beta = 0$) against the alternative that it is non-ignorable ($\beta \neq 0$) are based on the fullest specification (Model 4). The results are reported in Table 2. There is very strong evidence that attrition at the 3 and 12 month intervals is systematically linked with labor force status at the time of the initial interview. At the 3 month mark, it is the unemployed individuals who are more likely to attrit. At the 12 month mark, both employed and unemployed individuals are more likely to attrit compared to non-participants. Attrition at the 15 month mark is found to be ignorable.

The magnitudes involved are not negligible. Based on the results reported in Table A3, attrition probabilities of unemployed household heads are 1.1-1.2 percentage points higher than the average at the 3 month mark (4.7 percent). This amounts to a 25 percent increase in attrition probability. Based on the results reported in Table A4, this probability is even higher at the 12 month mark: 2.8 percentage points above the average (5 percent), which translates to a 56 percent increase. Furthermore, employed household heads also have above average attrition probabilities, by a margin of 1.3 percentage points (a 26 percent increase in relative terms). Evidently employed heads also find the need to change their location in response to changes in labor market conditions.

Table 3 contains a summary of the qualitative results based on Model 4 estimated at 3, 12, and 15 month intervals. In this table I report the signs of the statistically significant coefficients taking the 5 percent level as my standard. Zeros in the table mark the non-significant coefficients. At the bottom of the table I also report results from LR tests of the joint significance of the full model. Although all models are statistically significant at the 0.001 level, goodness of fit of the full model deteriorates as the attrition interval increases. As attritors leave the risk set, the attrition process becomes less and less selective (the survivors look more and more similar).

As far as the characteristics of the household head are concerned, the sign patterns in Table 3 are broadly consistent with the notion that attrition and migration go hand in hand. Interestingly neither being young, nor being single (rather than married) render attrition more likely, although they are known to make migration more likely. This could be attributable to the fact that we are studying attrition at the household level, using characteristics of the household head. Arguably closer links between migration and attrition are likely to be present at the individual level.

Our findings from the schooling variable shed further light on this issue. Using the numbers in Table A4 Model 4 for the purposes of illustration, the quadratic form we estimated suggests that the likelihood of household attrition is below average for poorly educated heads, and above average for high school graduates and higher. Attrition probability is lowest when the head has around 6 years of schooling. In fact 5-year primary school graduates actually dominated the labor force in 2000 (SIS, 2001b). However high school and university graduates claim an increasing share of recent cohorts of labor market entrants (Tunalı and Başlevent, 2006). Evidently it is the differences in educational attainment of the young cohorts rather than their age that distinguishes them as attritors/migrants.

Does location matter? If the migration interpretation is invoked, it should. In this paper we rely on a narrow distinction.[6] We find that households residing in urban areas are more likely to attrit. Broadly speaking, this finding is in line with the recent trends in migration, whereby moves between urban areas have come to dominate the internal migration flows.[7] Note, however, that migration studies typically focus on a longer (5-year, 10-year) time horizon than we do. Since new job opportunities are typically located in urban areas, our finding is consistent with job-search arguments. Consistent with a relocation cost based reasoning, small households are more likely to attrit. The cubic polynomial we relied on revealed that attrition probability was higher for below average size households (< 4.2), practically constant in the middle range, and extremely low for very large households.

The only inconsistencies in the sign patterns across attrition intervals are found in the case of year and quarter dummies which mark the timing of the initial survey. There is reason to believe that this is attributable to censoring, which does not allow us to fully capture the time-series variation in attrition likelihood over the 2000-2002 period. Note that the situation is better for $A(3)$: three of the four quarters of 2002 contribute to the risk set. In this case the patterns are easily reconciled with changes in the economic conditions. With year 2000 as the reference, we see that the 3-month attrition probability on average was lower in 2001 (the year of the economic crisis) by about 3 percent, and higher in 2002 (the year following the crisis, when the economy began its rebound) by about 4 percent. With quarter 1 as the reference period, we find that attrition is more likely to take place in the second and third quarters, possibly because individuals respond to seasonal job opportunities. In the second quarter, the average attrition probability is augmented by about 3 percent. In the third the increase is less than 1 percent.

---

[6] To push the migration interpretation further, one would need richer geographic demarcations. Unfortunately TURKSTAT did not include province and regional identifiers in the raw data I was granted access to.

[7] Insert references!

6. CONCLUSIONS

This paper offers a micro-econometric analysis of attrition patterns in the "New" Turkish Household Labor Force Survey (HLFS) which has been conduced since 2000. For this purpose 12 rounds of micro data collected (on a quarterly basis) over the period 2000-2002 were used. In general attrition is a phenomenon which can be attributed to demographic and economic factors, including conditions in the labor market. The purpose of conducting frequent household labor force surveys is to reflect the changing conditions in the labor market. If attrition is related to pre-attrition labor force status of individuals, this could result in bias in the labor market indicators. Our findings confirm a systematic link between labor force status and subsequent attrition. Compared to the average 20-54 year-old household head, those who are unemployed at the time of the initial survey are 25 percent more likely to attrit 3 months later. Conditional on being present 3 months later, they are 56 percent more likely to attrit 12 months after the initial survey. Employed heads who survive the 3 month mark are 26 percent more likely to attrit at the 12 month mark. These are magnitudes which cannot easily be ignored.

Arguably the most important feature of the "New" HLFS which distinguishes it from the older version is its short panel component. The design is similar to the Current Population Survey conducted in the USA and consequently it provides information on changes in the labor force statuses of individuals at quarterly and annual intervals. If we classify individuals of working age as outside the labor force, employed, or unemployed at a given point in time, knowledge of changes in the (quarterly, yearly) transition rates will provide us with extremely important clues on the links between the conditions in the labor market and the broader economics conditions. Unfortunately the Turkish Statistical Institute (TURKSTAT) does not publish predictions based on the panel dimension of the data. This might be attributable to difficulties associated with attrition.

The finding that households which are re-interviewed do not constitute a random subset of the initial sample constitutes strong evidence that attrition results in erosion of the representativeness of the HLFS. In fact TURKSTAT substitutes attriting households by new ones if they move to an original address on their list, and reweighs the cross-section for the purposes of the quarterly (and more recently monthly) indicators it publishes. Construction of weights in the face of attrition is a vigorously debated subject by Survey Statisticians, Applied Econometricians and Labor Economists. Unbiased estimation of cross-section and transition indicators requires full understanding of the demographic and economic determinants of attrition, and suitable corrective measures. This investigation is meant to contribute to this endeavor, so that indicators on transition dynamics could be included among the information published on the basis of the HLFS.

REFERENCES:

Becketti, S., W. Gould, L. Lillard, and F. Welch (1988) "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation." *Journal of Labor Economics*, 6:4, 472-492.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998) "An Analysis of Sample Attrition in Panel Data." *Journal of Human Resources*, 33:2, 251-299.

Hausman, J. A. and D. A. Wise (1979) "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment." *Econometrica*, 47:2, 455-473.

Little, R. and D. Rubin (1987) *Statistical Analysis with Missing Data*. New York: Wiley.

MaCurdy, T., T.Mroz, and R. M. Gritz (1998) "An Evaluation of the NLSY." *Journal of Human Resources*, 33:2, 345-436.

Ridder, G. (1992) "An Empirical Evaluation of Some Models for Non-Random Attrition in Panel Data." *Structural Change and Economic Dynamics*, 3: 2, 337-355.

Rubin, D. (1976) "Inference and Missing Data." *Biometrika*, 63: 581-592.

SIS (2001a) *Hanehalkı İşgücü Anketi: Kavramlar ve Yöntemler* [Household Labor Force Survey: Concepts and Methods]. Devlet İstatistik Enstitüsü Matbaası: Ankara.

SIS (2001b) *Hanehalkı İşgücü Anketi Sonuçları: 2000* [Household Labor Force Survey Results: 2000]. Devlet İstatistik Enstitüsü Matbaası: Ankara.

Taşçı, H. M. and A. Tansel (2005a) "Explaining Unemployment Duration for Men and Women in a Developing Country: The Case of Turkey." Mimeo.

Taşçı, H. M. and A. Tansel (2005b) "Transitions in the Turkish Labor Market: Evidence from Individual Level Data." Economic Research Forum Discussion Paper, Cairo and Bonn; IZA Discussion Paper.

Tunalı, İ. (1996) "Migration and Remigration of Male Household Heads in Turkey, 1963-73." *Economic Development and Cultural Change*, 45:1, 31-67.

Tunalı, İ. (2005) "General vs. Vocational Secondary School Choice and Labor Market Outcomes in Turkey, 1988-1998." Mimeo.

Tunalı, İ. and R. Baltacı (2004) "Attrition in the New HLFS, 2000-2002." Mimeo.

Tunalı, İ. and C. Başlevent (2006) "Female Labor Supply in Turkey." Ch.4 in *The Turkish Economy: The Real Economy, Corporate Governance and Reform and Stabilization Policy*, edited by S. Altuğ and A. Filiztekin, 92-127. Routledge: London.

van den Berg, G. and M. Lindeboom (1998) "Attrition in Panel Survey Data and the Estimation of Multi-State Labor Market Models." *Journal of Human Resources*, 33:2, 458-478.

World Bank (2006) *Turkey Labor Market Study*. Report No. 33254-TR. Poverty Reduction and Economic Management Unit (Europe and Central Asia Region).

Zabel, J. E. (1998) "An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior." *Journal of Human Resources*, 33:2, 479-506.

Table 1. Summary Statistics on the working sample (initial visit)
Household Heads, Age 20-54

Variable *denotes dummy variables	Mean	Std. Dev.	Min	Max
age	39.3	8.19	20	54
age2 (x1/100)	16.1	6.43	4	29.2
age3 (x1/10000)	6.85	3.94	0.8	15.7
sch	7.27	3.78	0	17
sch2 (x1/100)	0.671	0.659	0	2.89
*female	0.076	0.265	0	1
*urban (reference rural)	0.795	0.403	0	1
*emp1 (reference non-participant)	0.793	0.405	0	1
*unemp1 (reference non-participant)	0.069	0.253	0	1
*yr2000 (reference year)	0.517	0.500	0	1
*yr2001	0.285	0.451	0	1
*yr2002	0.199	0.399	0	1
*q1 (reference quarter)	0.333	0.471	0	1
*q2	0.268	0.443	0	1
*q3	0.259	0.438	0	1
*q4	0.140	0.347	0	1
*single	0.026	0.160	0	1
*divorced	0.018	0.131	0	1
*widow	0.037	0.190	0	1
hhsize	4.198	1.75	1	25
hhsize2 (x1/100)	0.2069	0.211	0.01	6.25
hhsize3 (x1/10000)	0.0121	0.029	0.0001	1.56
*voc_highschool	0.0697	0.255	0	1
*voc_college	0.0244	0.154	0	1

No. of observations = 49,318.

Table 2. Tests of the null hypothesis that attrition is ignorable vs. it is not
***p*-values based on the Probit estimates for Model 4 reported in Tables A3-A5**

Labor force status during the initial visit	*A*(3)	*A*(12)	*A*(15)
Employed	0.075	0.004	0.254
Unemployed	0.006	<0.001	0.302
Non-participant (reference)	--	--	--
Joint test	<0.001	<0.001	0.35
Observations	49,318	26,408	21,855

Tests based on the standard Normal and Chi Squared (2) tables.

Table 3. Attrition patterns as of 3, 12, and 15 months, households with 20-54 year-old heads
Qualitative results based on the Probit Estimates for Model 4 reported in Tables 3-5

Variable * denotes the characteristics of the HH head	*A*(3)	*A*(12)	*A*(15)
yr2001	–	+	0
yr2002	+	n.a.	n.a.
q2	+	0	+
q3	+	–	0
q4	0	+	0
*age	0	0	0
*age2	0	0	0
*age3	0	0	0
*sch	0	–	0
*sch2	+	+	0
*female	0	+	0
urban	+	+	+
*single	0	0	0
*divorced	+	0	0
*widow	0	0	0
*emp1	0	+	0
*unemp1	+	+	0
hhsize	–	–	–
hhsize2	+	+	0
hhsize3	–	0	0
*voc_highschool	–	0	–
*voc_college	0	0	0
Observations	49318	26408	21855
Log-likelihood w/o covariates	-9323	-5203	-5919
Log-likelihood w/ full set of covariates	-8603	-5031	-5807
LR test: Incremental Chi-sq (d.f.)	1440 (22)	344 (21)	224 (21)

Reported signs are for statistically significant coefficients at the 5 percent level or lower.

Table A1. Rotation plan of the HLFS, 2000-2002*

	2000				2001				2002			
Rotation number	1	2	3	4	1	2	3	4	1	2	3	4
01	E1x	--	--	--	--	--	--	--	--	--	--	--
02	(E1>)	(E2x)	--	--	--	--	--	--	--	--	--	--
03	O1x	(E1>)	(E2x)	--	--	--	--	--	--	--	--	--
04	(O1>)	(O2x)	(E1>)	(E2x)	--	--	--	--	--	--	--	--
05	[E1>]	(O1>)	(O2x)	[E2>]	[E3x]	--	--	--	--	--	--	--
06	{E1>}	{E2>}	(O1>)	(O2x)	{E3>}	{E4x}	--	--	--	--	--	--
07	[O1>]	{E1>}	{E2>}	[O2>]	[O3x]	{E3>}	{E4x}	--	--	--	--	--
08	{O1>}	{O2>}	{E1>}	{E2>}	{O3>}	{O4x}	{E3>}	{E4x}	--	--	--	--
09		{O1>}	{O2>}	{E1>}	{E2>}	{O3>}	{O4x}	{E3>}	{E4x}	--	--	--
10			{O1>}	{O2>}	{E1>}	{E2>}	{O3>}	{O4x}	{E3>}	{E4x}	--	--
11				{O1>}	{O2>}	{E1>}	{E2>}	{O3>}	{O4x}	{E3>}	{E4x}	--
12					{O1>}	{O2>}	{E1>}	{E2>}	{O3>}	{O4x}	{E3>}	{E4x}
13						{O1>}	{O2>}	[E1>]	[E2>]	{O3>}	{O4x}	[E3c]
14							{O1>}	{O2>}	(E1>)	(E2c)	{O3>}	{O4x}
15								[O1>]	[O2>]	(E1>)	(E2c)	[O3c]
16									(O1>)	(O2c)	(E1>)	(E2c)
17										(O1>)	(O2c)	E1c
18											(O1>)	(O2c)
19												O1c
visit counter	1	1,2	1,2	1,2	1,2,3	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4
max. no. of visits	1	1,2	1,2	1,2	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4
expected no. of visits	1,2,3,4	2,4	2,4	2,3,4	3,4	4	4	3,4	2,3,4	2,4	2,4	1,2,3,4

Source: SIS (2001a) and own calculations (three rows at the bottom).

*Legend: O = odd number; E = even number; > = subsequent visit planned; x = exits from survey; c = censored.
Total number of planned visits: no mark = 1 visit; (parentheses) = 2 visits; [bracket] = 3 visits; {brace} = 4 visits.

Table A2. Risk sets and proportion of attritors [*p*(A = 1)] by *observation unit*, attrition type [A(m)] and survey round

	2000				2001				2002				
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Total**
All Individuals													
A(3): at risk	36,891	24,464	23,566	11,893	12,702	13,394	13,434	13,648	12,897	12,691	11,446	0	**187,026**
***p*(A=1)**	*0.0353*	*0.0739*	*0.0538*	*0.0177*	*0.0228*	*0.0257*	*0.0229*	*0.0381*	*0.0804*	*0.1254*	*0.0916*		***0.0520***
A(9): at risk	12,703	0	0	0	0	0	0	0	0	0	0	0	**12,703**
***p*(A=1)**	*0.0350*												***0,0350***
A(12): at risk	12,175	11,849	11,791	11,683	12,412	13,050	13,127	13,128	0	0	0	0	**99,215**
***p*(A=1)**	*0.0621*	*0.0487*	*0.0466*	*0.0520*	*0.0553*	*0.0613*	*0.0519*	*0.1128*					***0.0619***
A(15): at risk	11,419	11,272	11,241	11,076	11,726	12,250	12,446	0	0	0	0	0	**81,430**
***p*(A=1)**	*0.0849*	*0.0870*	*0.0877*	*0.0797*	*0.0758*	*0.1135*	*0.0752*						***0.0864***
All Households													
A(3): at risk	13,685	9,061	8,765	4,495	4,663	5,088	4,942	5,145	4,805	4,725	4,290	0	**69,664**
***p*(A=1)**	*0.0315*	*0.0673*	*0.0428*	*0.0118*	*0.0148*	*0.0145*	*0.0136*	*0.0292*	*0.0728*	*0.1242*	*0.0795*		***0.0446***
A(9): at risk	4,756	0	0	0	0	0	0	0	0	0	0	0	**4,756**
***p*(A=1)**	*0.0307*												***0.0307***
A(12): at risk	4,553	4,422	4,372	4,442	4,594	5,014	4,875	4,995	0	0	0	0	**37,267**
***p*(A=1)**	*0.0404*	*0.0341*	*0.0348*	*0.0378*	*0.0422*	*0.0469*	*0.0334*	*0.0889*					***0.0454***
A(15): at risk	4,369	4,271	4,220	4,274	4,400	4,779	4,712	0	0	0	0	0	**31,025**
***p*(A=1)**	*0.0712*	*0.0742*	*0.0706*	*0.0679*	*0.0625*	*0.1073*	*0.0545*						***0.0729***
HH heads, 20-54													
A(3): at risk	9,702	6,344	6,238	3,201	3,304	3,571	3,474	3,690	3,410	3,322	3,062	0	**49,318**
***p*(A=1)**	*0.0326*	*0.0695*	*0.0436*	*0.0125*	*0.0154*	*0.0160*	*0.0153*	*0.0301*	*0.0780*	*0.1334*	*0.0846*		***0.0468***
A(9): at risk	3,329	0	0	0	0	0	0	0	0	0	0	0	**3,329**
***p*(A=1)**	*0.0306*												***0.0306***
A(12): at risk	3,264	3,099	3,117	3,161	3,253	3,514	3,421	3,579	0	0	0	0	**26,408**
***p*(A=1)**	*0.0438*	*0.0374*	*0.0372*	*0.0411*	*0.0501*	*0.0492*	*0.0389*	*0.0930*					***0.0495***
A(15): at risk	3,121	2,983	3,001	3,031	3,090	3,341	3,288	0	0	0	0	0	**21,855**
***p*(A=1)**	*0.0750*	*0.0805*	*0.0736*	*0.0680*	*0.0647*	*0.1152*	*0.0584*						***0.0768***

Table A3. Probit estimates of attrition as of 3 months, households with 20-54 year-old heads Dependent variable *A*(3)

Variable * denotes the characteristics of the HH head	Model 0	Model 1	Model 2	Model 3	Model 4
yr2001	-0.332**	-0.337**	-0.342**	-0.343**	-0.342**
	(0.029)	(0.029)	(0.029)	(0.030)	(0.030)
yr2002	0.411**	0.417**	0.409**	0.409**	0.412**
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
q2	0.299**	0.299**	0.304**	0.308**	0.308**
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
q3	0.082**	0.079**	0.084**	0.084**	0.084**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
q4	-0.011	-0.018	-0.016	-0.019	-0.019
	(0.040)	(0.040)	(0.040)	(0.041)	(0.041)
*age		-0.037	-0.038	-0.037	-0.043
		(0.073)	(0.073)	(0.074)	(0.074)
*age2		0.017	0.022	0.055	0.07
		(0.195)	(0.195)	(0.196)	(0.197)
*age3		0.033	0.025	-0.03	-0.041
		(0.169)	(0.170)	(0.171)	(0.171)
*sch		-0.020	-0.019	-0.023*	-0.012
		(0.011)	(0.011)	(0.012)	(0.012)
*sch2		0.262**	0.265**	0.264**	0.217**
		(0.063)	(0.063)	(0.064)	(0.066)
*female		0.02	0	-0.028	-0.027
		(0.051)	(0.054)	(0.054)	(0.054)
urban		0.222**	0.214**	0.206**	0.206**
		(0.029)	(0.029)	(0.029)	(0.029)
*single		0.263**	0.262**	0.095	0.094
		(0.054)	(0.054)	(0.060)	(0.060)
*divorced		0.417**	0.407**	0.281**	0.280**
		(0.070)	(0.070)	(0.072)	(0.072)
*widow		-0.062	-0.065	-0.148	-0.141
		(0.076)	(0.076)	(0.077)	(0.077)
*emp1			-0.059	-0.05	-0.052
			(0.036)	(0.036)	(0.036)
*unemp1			0.111*	0.123*	0.122*
			(0.048)	(0.049)	(0.049)
hhsize				-0.241**	-0.240**
				(0.051)	(0.051)
hhsize2				3.095**	3.075**
				(0.913)	(0.912)
hhsize3				-12.890**	-12.773**
				(4.871)	(4.861)
*voc_highschool					-0.138**
					(0.042)
*voc_college					-0.067
					(0.063)
Constant	-1.838**	-1.118	-1.09	-0.718	-0.675
	(0.020)	(0.894)	(0.894)	(0.900)	(0.900)
Observations	49318	49318	49318	49318	49318
Log-likelihood w/o covariates	-9323	-9323	-9323	-9323	-9323
Log-likelihood	-8853	-8649	-8639	-8609	-8603
LR test: Incremental Chi-sq (d.f.)	940 (5)	408 (10)	20 (2)	60 (3)	12 (2)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**) levels.

Table A4. Probit estimates of attrition as of 12 months, households with 20-54 year-old heads
Dependent variable *A*(12)

Variable * denotes the characteristics of the HH head	Model 0	Model 1	Model 2	Model 3	Model 4
yr2001	0.176**	0.179**	0.176**	0.175**	0.175**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
q2	-0.041	-0.047	-0.048	-0.05	-0.05
	(0.038)	(0.039)	(0.039)	(0.039)	(0.039)
q3	-0.102**	-0.109**	-0.108**	-0.111**	-0.111**
	(0.039)	(0.040)	(0.040)	(0.040)	(0.040)
q4	0.180**	0.182**	0.182**	0.179**	0.179**
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
*age		-0.096	-0.099	-0.089	-0.09
		(0.096)	(0.096)	(0.097)	(0.097)
*age2		0.205	0.204	0.208	0.21
		(0.258)	(0.258)	(0.259)	(0.259)
*age3		-0.16	-0.149	-0.174	-0.176
		(0.224)	(0.225)	(0.226)	(0.226)
*sch		-0.036*	-0.035*	-0.036*	-0.036*
		(0.014)	(0.014)	(0.015)	(0.015)
*sch2		0.312**	0.307**	0.296**	0.296**
		(0.080)	(0.080)	(0.081)	(0.083)
*female		0.142*	0.208**	0.177*	0.178*
		(0.067)	(0.071)	(0.071)	(0.071)
urban		0.179**	0.181**	0.179**	0.179**
		(0.037)	(0.037)	(0.037)	(0.037)
*single		0.265**	0.246**	0.108	0.109
		(0.072)	(0.072)	(0.079)	(0.079)
*divorced		0.127	0.11	0.011	0.01
		(0.110)	(0.110)	(0.112)	(0.112)
*widow		-0.001	0.013	-0.039	-0.04
		(0.093)	(0.094)	(0.094)	(0.094)
*emp1			0.129*	0.134**	0.134**
			(0.050)	(0.050)	(0.050)
*unemp1			0.276**	0.282**	0.281**
			(0.066)	(0.067)	(0.067)
hhsize				-0.198**	-0.198**
				(0.054)	(0.054)
hhsize2				2.336**	2.335**
				(0.858)	(0.858)
hhsize3				-7.595	-7.592
				(3.939)	(3.939)
*voc_highschool					0.00009
					(0.054)
*voc_college					-0.046
					(0.085)
Constant	-1.768**	-0.335	-0.417	-0.214	-0.209
	(0.031)	(1.166)	(1.168)	(1.173)	(1.173)
Observations	26408	26408	26408	26408	26408
Log-likelihood w/o covariates	-5203	-5203	-5203	-5203	-5203
Log-likelihood	-5144	-5051	-5043	-5031	-5031
LR test: Incremental Chi-sq (d.f.)	118 (4)	186 (10)	16 (2)	24 (3)	0 (2)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**) levels.

Table A5. Probit estimates of attrition as of 15 months, households with 20-54 year-old heads
Dependent variable *A*(15)

Variable * denotes the characteristics of the HH head	Model 0	Model 1	Model 2	Model 3	Model 4
yr2001	0.018	0.019	0.018	0.017	0.018
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
q2	0.188**	0.179**	0.181**	0.181**	0.181**
	(0.032)	(0.033)	(0.033)	(0.033)	(0.033)
q3	-0.033	-0.032	-0.031	-0.03	-0.031
	(0.034)	(0.035)	(0.035)	(0.035)	(0.035)
q4	-0.006	-0.011	-0.01	-0.012	-0.014
	(0.044)	(0.045)	(0.045)	(0.045)	(0.045)
*age		-0.046	-0.047	-0.035	-0.036
		(0.094)	(0.094)	(0.094)	(0.094)
*age2		0.084	0.088	0.078	0.077
		(0.250)	(0.250)	(0.251)	(0.251)
*age3		-0.049	-0.055	-0.064	-0.061
		(0.216)	(0.216)	(0.217)	(0.217)
*sch		0.001	0.001	-0.001	0.01
		(0.014)	(0.014)	(0.015)	(0.015)
*sch2		0.124	0.128	0.124	0.073
		(0.080)	(0.080)	(0.081)	(0.083)
*female		0.111	0.096	0.077	0.078
		(0.070)	(0.073)	(0.073)	(0.073)
urban		0.216**	0.213**	0.212**	0.212**
		(0.035)	(0.035)	(0.035)	(0.035)
*single		0.067	0.067	-0.017	-0.02
		(0.081)	(0.081)	(0.085)	(0.085)
*divorced		0.17	0.169	0.104	0.105
		(0.108)	(0.108)	(0.110)	(0.110)
*widow		-0.02	-0.023	-0.061	-0.055
		(0.094)	(0.094)	(0.094)	(0.094)
*emp1			-0.033	-0.029	-0.030
			(0.045)	(0.045)	(0.045)
*unemp1			0.029	0.034	0.034
			(0.065)	(0.065)	(0.065)
hhsize				-0.097*	-0.095*
				(0.044)	(0.044)
hhsize2				0.774	0.748
				(0.617)	(0.617)
hhsize3				-0.79	-0.694
				(2.356)	(2.355)
*voc_highschool					-0.172**
					(0.055)
*voc_college					0.076
					(0.077)
Constant	-1.485**	-0.959	-0.929	-0.919	-0.932
	(0.028)	(1.151)	(1.152)	(1.155)	(1.155)
Observations	21855	21855	21855	21855	21855
Log-likelihood w/o covariates	-5919	-5919	-5919	-5919	-5919
Log-likelihood	-5889	-5823	-5822	-5813	-5807
LR test: Incremental Chi-sq (d.f.)	60 (4)	132 (10)	2 (2)	18 (3)	12 (2)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**) levels.