ECONOMIC RESEARCH FORUM

منتدى البحوث الاقتصادية

2008

working paper series

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

Insan Tunali

Working Paper No. 393

Analysis of Attrition Patterns in the Turkish Household Labor Force Survey, 2000–2002

Insan Tunali

Working Paper 393

April 2008

Earlier versions were presented at the Conference on "Labor Markets in Transition and Developing Economies: Emerging Policy and Analytical Issues," Ann Arbor, May 25–27, 2007, the 14th Annual Conference of the Economic Research Forum, Cairo, December 28–30, 2007, and workshops at Bilkent, Bilgi, and Koç Universities. Research assistance from Emre Ekinci and Berk Yavuzoğlu, and funding from TÜBİTAK via grant no: 106K160 are gratefully acknowledged. I would like to thank Dr. Ömer Demir, the president of Turkish Statistical Institute (TURKSTAT), for granting me access to the raw data from the Household Labor Force Survey (HLFS) and TURKSTAT staff Hasibe Dedeş, Didem Sezer and Enver Taştı for their willingness to respond to my endless questions about the HLFS. Finally, I am grateful to Aysıt Tansel, Jackie Wahba, Berk Yavuzoğlu, conference and workshop participants for their comments on earlier versions.

Insan Tunali, Dept. of Economics, Koç University, Istanbul, Turkey
Email : itunali@ku.edu.tr

Abstract

This paper offers an analysis of attrition patterns in the "new" Turkish Household Labor Force Survey (HLFS) which has been conducted since 2000. A key feature of the redesigned survey is the short panel component obtained from the rotating sampling frame. I exploit the information in 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 the labor force status of individuals, this could result in biases in labor market indicators. I 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.

ملخص

تقوم هذه الورقة بتحليل نماذج للاستنزاف الذي تتعرض له الأسرة في مسح القوي العاملة للأسرة التركية والذي اجري منذ عام 2000. ويتسم المسح المعاد تصميمه بقصر مدي نتائج البحث نظرا لتناوب الإطار الذي تؤخذ منه العينة. وقد استثمرت تلك المعلومات في اثنتي عشرة جولة من البيانات الدقيقة التي يتم تجميعها فصليا (كل ثلاثة اشهر) خلال الفترة من 2000 إلى 2002 مع التركيز علي ذلك الاستنزاف علي مستوي الأسرة خلال ثلاثة اشهر, واثني عشر شهرا وخمسة عشر شهرا من أول مقابلة. واستنزاف الأسرة ظاهرة يمكن أن نعزيها إلى عوامل سكانية واقتصادية منها ظروف سوق العمل. فإن تعلق استنزاف الأسرة بمكانة الأفراد في القوي العاملة فقد يؤدي ذلك إلى وجود انحرافات في مؤشرات سوق العمل. وسأسوق أدلة قوية علي أن استنزاف الأسرة يتأثر بوضع رب الأسرة في القوي العاملة (خارج نطاق هذه القوة العاملة سواء وجدت عملا أم ظلت بلا عمل) في الجولة الاستقصائية الأولية.

1. Introduction

Attrition (initial response followed by non–response at a later round of a survey) 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 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 Verbeek and Nijman (1996), and 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) and Zabel (1998).

Starting in 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 allows tracking of labor market dynamics.[1] However, to date, only two papers have addressed the subject (Taşçı and Tansel, 2005; Tansel and Taşçı, 2006).

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 2000/02 are biased, on the grounds that the statistics are influenced by the number of times a household has been interviewed.

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.

[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.

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 (around 47,000) households headed by prime–age (20–54 years–old) individuals, identified as the household head in the first round of the survey. For the subset of (23,790) households which were designated for four interviews, the cumulative probability of attrition is 8 percent by 3 months, 18.3 percent by 12 months, and 24.7 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 understood. In what follows I show convincingly that the labor market state occupied by the household head in the first round influences and is influenced by attrition 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 which insist on simplistic reweighing schemes and policy makers who rely on statistics produced in this manner.

A short formal statement of the problem and its consequences are provided in section 2. Sections 3 and 4 are devoted to data related issues: I discuss the HLFS survey, data problems and the solutions I adopted. I then define attrition formally and give some summary statistics. The estimation and testing methodology is the subject of section 5. Sections 6 and 7 contain the empirical results. The concluding section 8 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, but 0 otherwise. 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

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

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

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

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 steady–state marginal distributions of membership in labor market states $f(y|x)$ in the reweighed cross–section as well as steady–state 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 the 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 Measurement Issues

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.

In this paper I rely on twelve rounds of the HLFS for the period 2000/02. Each round of the survey includes around 70,000 individuals from 18,000 to 20,000 households. The full data set consists of about 890,000 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 did 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 and the observation plan to which I turn next.

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 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, it is possible to determine the maximum number of recorded visits as well as the expected number of (total) visits to a given household. Consistency checks on the raw data exposed coding errors in the visit number. These mistakes were easily corrected using information on the rotation number and round which uniquely determine the visit number (see Table A1 in the Appendix).

A household was classified as an "attritor" if a scheduled interview did not take place in a subsequent round. A similar scheme was used to detect individual household members who attrited. There were a total of 66,467 households (headed by someone of age 15 or older) and 184,339 individuals (of age 15 or older) who were eligible for the analysis, although not all were subjected to the full rotation plan. About 26 percent of eligible households and 31.6 percent of eligible individuals attrited sometime during the observation window.[2] The survey protocol of the HLFS allows for substitution of a departing 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 in the original sample, the departing households were classified as attritors and substitute households were excluded from the working sample. It is also possible for a household to leave the address for some time, only to return later. The returnee households were classified as attritors when they did not show up in the data as scheduled and excluded from subsequent analysis.

Investigation of substitution provides additional clues about the magnitude of attrition. To begin with, the number of substitute households is less than the number of attriting households (8,492 vs. 11,618). One might argue that substitution may correct for the distortions introduced by attrition, however not all attriting households were replaced. Furthermore a significant share of substitute households (1,001) attrited themselves and some of these were replaced by other substitute households as part of the protocol. Since the total includes multiple substitutions for the same household, double counting is present. Unfortunately the exact number cannot be determined without a marker for the address.

Secondly, there might be initial non-response followed by response. That is, some households classified as substitutes may actually be returnees who were absent in the initial round of the survey. Given the fact that a total of 3,688 attritors (amounting to 31.7 percent of all attriting households) returned to their original addresses during the period under study, there is good reason to believe that a sizeable subset of those classified as substitutes may be returnees themselves. Since the HLFS sample frame is address based, it is impossible to distinguish returnees from substitutes in the absence of information about migration history. Conveniently this paper is about attrition patterns in the sample subjected to the first round of the interview, so the approach adopted below serves us well.

Further examination of the data revealed that some departures from the survey protocol did take place in the field. In some rare instances, the ID of a departing household appears to have been 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. Since the analysis of attrition patterns in the current paper is confined to prime–age household heads, the latter group does not concern us here. As for the former, in theory the problem cases can be identified using a computer program that keeps track of changes in the household roster. However, legitimate life cycle events and coding errors that are part and parcel of panel data proved difficult to distinguish.[3] In the end I decided not to do any corrections.

---

[2] These figures were obtained by summing the fraction of attritors after 3, 12 and 15 months given in the first two panels of Table A2. The third panel is associated with the working samples of this paper and is analysed in some detail in Section 4.

[3] Examination revealed that in 14.7 percent of eligible households, in some subsequent round the age of the household head differed from the average by more than two years. In 4.1 percent of the eligible households the sex of the household head changed somewhere along the line. Inspection of a random sample of individual

4. Definitions and Summary Statistics on Attrition

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 mh*, it is excluded from the risk set at higher intervals. That is, household h belongs to risk set R(m) if mh* > 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 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 (71.3 percent) of all households (see Table A2). Households I study are slightly more likely to attrit, but not by much. I focus on the subset because my main objective is to establish the links between labor market status and attrition at a time when the head is likely to be economically active. Labor market attachment of older household heads is low, and attrition behavior of their households may have other explanations. Households whose heads are younger (age<20) are extremely rare (114 in 66,467), and these are typically single person households who are attrition–prone. As seen in Table A2, incidence of attrition is higher at the individual level, which was expected. However, due to the challenges posed by the coding errors mentioned above, analysis of individual level attrition is undertaken as a separate project. From this point on all references to attrition is confined to households with a 20–54 year–old head. For brevity, I drop the qualifier and refer to them as households.

My working sample consists of 47,373 distinct households. Given the rotation plan and the observation window 2000/02, different subsets of households qualify for studying attrition behavior at different intervals. The risk sets are respectively 47,373 for studying *A*(3), 25,324 for studying *A*(12) and 19,437 for studying *A*(15). Recall that (pre–steady–state segment of) the rotation plan allows us to compute *A*(9) as well. However the risk set is considerably smaller (3,414) 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 totals (see final column of bottom block in Table A2), about 8.8 percent of the households attrit before the second round (by month 3). Conditional on survival until the second round and inclusion in the sample frame, an additional 11 percent attrit before the third round (by month 12), that is, six months after the second interview. Conditional on survival until the third round and inclusion in the sample frame, an additional 7.8 percent of the households attrit before the fourth and final round (by month 15). 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 on which I rely 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 labor economics literature as a guide. The complete list is given in Table 1 along with some descriptive statistics on the working samples. All variables are measured at the initial

---

records revealed that most of these were legitimate changes. For a summary discussion of these issues in other contexts, see Deaton (1997: 37–39).

round of the survey.[4] Due to censoring, households (originally) interviewed in 2000 constitute close to one–half of the households in our working sample for $A(3)$. Those interviewed in 2001 and 2002 respectively account for 30 and 21 percent of the working sample respectively. In the $A(3)$ sample, households interviewed in the 1st through the 3rd quarter have above average representations (27.5 percent or more), while those interviewed in the 4th quarter have below average representation (15 percent). Observations in the $A(12)$ sample are almost evenly split between 2000 and 2001, and the four quarters of the year. In the $A(15)$ sample households interviewed in 2001 are underrepresented, because those interviewed in the last quarter are not eligible for analysis.

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 samples.[5] To provide a picture of the households, we focus on the $A(3)$ working sample. 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 12.1 percent are extended households. More than ninety percent of the household heads are married, and about eight percent are female. Average age of household head is about 39 years, and average education is approximately 7.3. Since I study prime age heads, labor market attachment is very high in my working sample. At the initial round of the survey 79 percent are employed, while 6.9 percent are unemployed.

5. Econometric Methodology

In testing for presence of non–ignorable attrition, I follow the approach in Fitzgerald et al. (1998) closely. They conduct two tests, which I term FGM and BGLW, which are in turn attributable to Fitzgerald, Gottschalk and Moffat (1998) and Becketti, Gould, Lillard and Welch (1988). For the FGM 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.

Let $A_h(m)$ denote the attrition status of household h as of m months after the initial interview. To implement the FGM test, I estimate Probit models of the form

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

for m = 3, 12, 15. All explanatory variables are measured at the initial survey round. Here $y(0)$ is a vector that contains indicators of the labor force status of the household head, x denotes the vector of 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

[4] In theory, richer specifications that use information on 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.

[5] Arguably labor markets in urban locations are more complex, and the sampling frame strives to capture this. 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.

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.[6]

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 non-response 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.

To implement the BGLW test, I focus on two binary labor market outcomes $y_h(0)$ recorded for the household head at the initial survey round: labor force participation (LFP) and unemployment (Unemp) conditional on LFP =1. I estimate Probit models of the form

$$(Pr\{y_h(0) = 1 \mid A(k), k \leq m, x, z; h \in R(m)\} = \Phi[\Sigma_{k \leq m} \alpha_k A(k) + \theta' x + \lambda' z] \qquad (4)$$

for m = 3, 12, 15. Here $A(k)$'s denote the binary indicators of (future) attrition status. All other explanatory variables are measured at the initial survey round. As before x denotes the vector of 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 α, θ, λ using maximum likelihood. If the null hypothesis that $\alpha = 0$ is rejected, we have evidence that attrition is non–ignorable.

As Fitzgerald et al. (1998, p.263) underscore, FGM and BGLW tests are related: "the BGLW method is an indirect test of the same restriction as the direct method of estimating the attrition function itself."[7] The BGLW version of the test is attractive because the estimated equation is a standard equation from the labor economics literature, augmented by indicators of future attrition status. As such it can be used routinely to check for presence of attrition bias in cross–section estimates. If the ultimate aim is to correct the bias in cross–section (or panel) estimates, the FGM methodology is preferable, because it provides the weights needed for the correction (see Fitzgerald at al. 1998).

6. Results from the FGM Tests (Based on Fitzgerald, et al. 1998)

At each attrition interval, the same set of 4 models was estimated. Each model is nested under the subsequent ones. Model 1 (baseline) includes year and quarter dummies only. These variables capture the common component of the time–series variation in attrition. In Model 2 a set of household characteristics are added to the model. This list includes variables that TURKSTAT uses for reweighing, such as age and gender of the household head, and location (urban vs. rural). In Model 3 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 households in which heads are employed (unemployed) display different attrition behavior. If the answer is yes, we have evidence of non–ignorable attrition. In

---

[6] Technically speaking it is possible to include additional lagged values of y (labor force status) when we study attrition over longer intervals (12, 15 moths). Fitzgerald et al. (1998) refer to expanded specifications involving the lagged terms as "dynamic attrition" models. These could be useful when constructing weights. Since I focus on whether attrition is ignorable, I rely on the simpler version.

[7] Using the terminology which was fashionable in the early 1980s, equation (4) is the "reverse regression" counterpart of equation (3). Reverse regressions were used to conduct an alternative test of discrimination until Goldberger (1984) established the strong distributional requirements needed for that interpretation. Goldberger's criticisms do not apply to the BGLW test, because no attempt is made to link the parameter (vector) α in equation (4) with the parameters in equation (3).

Model 4 information on schooling and marital status of the household head and household size (number of people residing in the household) are included. The functional forms of multiple valued variables (age, schooling and household size) are initially specified as third degree polynomials. After independent tests on Model 4 results, the functional forms are simplified and the resulting specification is reported as Model 5.

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. The sample sizes are the respective risk sets identified in the third panel of Table A2. At the bottom of each table incremental Likelihood Ratio test statistics that exploit the nesting properties of the subsequent models are reported. In Models 1–4 this statistic is used to test whether added variables are jointly statistically significant. In Model 5 the same statistic is used to test whether excluded variables are jointly statistically significant. Here we focus on the key findings using two summary tables. The FGM 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 final specification (Model 5). The results are reported in Table 2. There is very strong evidence that attrition is systematically linked with labor force status at the time of the initial interview. With non–participants as the reference category, at the 3 month mark employed individuals are less, and unemployed individuals are more likely to attrit. At the 12 month mark, unemployed individuals are more likely to attrit. Attrition at the 15 month mark is ignorable, possibly because response to labor–market draws has already taken place. Recall, however, that the samples are not fully nested, and investigations of longer interval behavior are carried out on considerably smaller samples.

The magnitudes involved are not negligible.[8] Based on the results reported in Table A3 for Model 5, attrition probabilities of unemployed household heads are 2.1 percentage points higher than the average (= 8.8 percent) at the 3 month mark. This amounts to a 24 percent increase in attrition probability. Based on the results reported in Table A4, this probability is even higher at the 12 month mark: 4.9 percentage points above the average (= 11 percent), which translates to a 45 percent increase. Arguably incidence of unemployment triggers mobility, so that job search can be extended beyond the local labor market. Notably employed household heads have below average attrition probabilities at the 3 month mark, by a margin of 1 percentage points (an 11 percent decrease in relative terms).

Table 3 contains a summary of the broader qualitative results based on Model 5 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. Thus the attrition process becomes less and less selective (the survivors look more and more similar) as attriting households leave the risk set.

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. Being young and being single (rather than married) render attrition more likely. Consistent with a relocation cost based reasoning, small households are more likely to attrit. The cubic polynomial we relied on in Table A3 revealed that attrition probability was higher for below

[8] The derivatives were calculated by multiplying the reported slopes by the values of the standard normal density at the average attrition probability for that round (yielding 0.16, 0.19 and 0.15 respectively).

average size households (< 4.2 members), practically constant in the middle range (5–10 members), and dropped sharply for very large households.

Holding age of household head constant, differences in educational attainment distinguish attritors (migrants). Using the numbers in Table A4 Model 5 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 5 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).

Does location matter? If the migration interpretation is invoked, it should. In this paper we rely on a narrow distinction.[9] 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.[10] 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.

The signs of the quarter dummies – which mark the timing of the initial survey – are consistent in identifying the round associated with the 3rd quarter as the time during which attrition is highest. With the 1st quarter as the reference period, when the initial interview takes place in the 2nd quarter, we see that the average 3–month attrition probability is augmented by 3.3 percentage points and the average 15–month probability is augmented by 2.7 percentage points. We also see that when the initial interview takes place in the 3rd quarter, the average 12–month move probability is augmented by 1.7 percentage points. This pattern is attributable to the fact that the employment level is highest in the third quarter.[11] The year dummies also help us establish a connection between employment prospects and attrition. With year 2000 as the reference, we see that the 3–month attrition probability on average was the same in 2001 (the year of the economic crisis), but higher in 2002 (the year following the crisis, when the economy began its rebound but employment continued to decline) by about 1.3 percentage points. Consistent with this finding, we see that the 12–month attrition probability is 1.4 percentage points higher in the case of households interviewed in 2001.

7. Results from the BGLW Tests (Based on Becketti at al. 1988)

The results from the BGLW test are collected in Tables A6 and A7 of the Appendix. The rotation design yields three different samples for each outcome under study. Sample sizes are

---

[9] 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.

[10] Gross flows data from the 2000 General Census which are available on the TURKSTAT web site (www.tuik.gov.tr) support this conclusion.

[11] For the period 2000/06, regression of total quarterly employment (millions) on year and quarter dummies yields the following estimated equation:

TOTEMP = 20 – .0075 YR01 – .075 YR02 – .25 YR03 + .17 YR04 + .54 YR05

+ . 81 YR06 + 2.2 Q2 + 2.7 Q3 + 1.4 Q4.

Regression of employment–population ratio on the same set of regressors yields:

EMPPOPRATIO = .44 – .0094 YR01 – .019 YR02 – .031 YR03 – .030 YR04

– .031 YR05 – .034 YR06 + .042 Q2 + .052 Q3 + .023 Q4.

determined by the expected number of visits to the address (see Table A1). The largest sample on which the participation decision is studied turns out to be the same as the risk set used for the attrition Probits at the 3 month mark (Table A3). The samples associated with expected number of visits 3 and 4 are larger than the corresponding risk sets (for attrition Probits at the 12 and 15 month marks). This is because longer interval attrition probabilities are estimated *conditional* on survival in the previous round. The restriction does not apply to the participation probability, where attrition dummies mark if and when 'future' attrition took place. The respective samples for the unemployment outcome are smaller, because only participants contribute to them.

On each sample, two models were estimated. Model I serves as the baseline and includes attrition indicators along with dummies that identify the initial survey round. As with the attrition models, not all survey rounds contribute to the sample. The extended model augments the baseline model with the usual list of explanatory variables used in reduced form specifications motivated by the standard formulations of the participation decision and the unemployment outcome. As in the FGM version, independent tests were used to simplify the polynomial specifications. Model II contains the simplified version of the extended model, which leads to the same substantive conclusions as the excluded full version. Since the focal point of the paper is attrition, in what follows I concentrate on the top panel of Tables A6 and A7 and refrain from detailed discussion of the determinants of participation and unemployment in the extended model. Suffice it to say that the patterns are as expected, and the models fit very well.

Summary findings are collected in Table 4. Recall that the steady state rotation plan involves four visits to each address. The top panel informs us about the consequences of attrition when we examine the subset of the sample which is supposed to deliver to this requirement. Using the *p*–values as our guide, in the participation equation we do not see any evidence of selective (non–ignorable) attrition. From the unemployment equation we learn that attrition 3 months as well as 12 months after the initial survey is selective of participants who were unemployed at the initial round. There is no evidence of additional selectivity in the final round. Viewed together, the patterns from the two equations corroborate the findings from the FGM test: There is a strong association between being unemployed at the time of the initial survey and attrition 3 or 12 months later.

The remaining panels of Table 4 use information from a broader set of households. When we include households which should have been visited three times, impacts of attrition behavior at the 3 and 12 months can be studied. When we include all households slated for a repeat visit, only the impact of attrition at the 3 month mark can be investigated. The results from the broader samples reinforce the conclusions drawn from the smallest sample.

8. Conclusions

This paper offers a micro–econometric analysis of attrition patterns in the "new" Turkish Household Labor Force Survey (HLFS) which has been conducted 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 household headed by a 20–54 year–old, those headed by an unemployed individual at the time of the initial survey are 24 percent more likely to attrit 3 months later. Conditional on being present during the second interview that takes place at the 3 month mark, households headed by an unemployed individual are 45 percent more likely to attrit 12 months after the initial survey. These large magnitudes underscore the strong links between attrition behavior and adverse experiences in the labor market. Although the effect is milder, there is evidence that good draws make attrition less likely: households headed by an employed individual are 11 percent less likely to attrit at the 3 month mark.

Arguably the most important feature of the "new" HLFS, which distinguishes it from the older version, is its short panel component. The rotation design is similar to those used in better–known surveys, such as the Current Population Survey conducted in the USA. 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 economic 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 attrition process removes individuals (households) from the sample on the basis of their (household head's) labor market status and may render the remaining HLFS sample unrepresentative. 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.

Deaton, A. (1997). "The Analysis of Household Surveys: A Microeconomic Approach to Development Policy." Published for the World Bank, Johns Hopkins University Press: Baltimore and London.

Goldberger, A.S. (1984). "Reverse Regression and Salary Discrimination." *The Journal of Human Resources*, 19: 3, 293–318.

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.

Tansel, A. (2002). "General vs. Vocational High Schools and Labor Outcomes in Turkey." Ch. 11 in *Human Capital: Population Economics in the Middle East*, edited by I. Sirageldin. Economic Research Forum and American University in Cairo Press: Cairo.

Tansel, A. and H. M. Taşçı (2006). "Explaining Unemployment Duration for Men and Women in a Developing Country: The Case of Turkey." Mimeo.

Taşçı, H. M. and A. Tansel (2005). "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.

Verbeek, M. and T. Nijman (1996). "Incomplete Panels and Selection Bias." In *The Econometrics of Panel Data*, edited by L. Matyas and P. Sevestre. Dordrecth: Kluwer.

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 Samples (initial visit)

	Household Heads, Age 20–54											
	***A*(3)**				***A*(12)**				***A*(15)**			
Variable *denotes dummy variables	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
*yr2000 (reference year)	0.49	0.50	0	1	0.48	0.50	0	1	0.56	0.50	0	1
*yr2001	0.3	0.46	0	1	0.52	0.50	0	1	0.44	0.50	0	1
*yr2002	0.21	0.41	0	1	–	–	–	–	–	–	–	–
*q1 (reference quarter)	0.29	0.45	0	1	0.25	0.43	0	1	0.29	0.45	0	1
*q2	0.29	0.45	0	1	0.25	0.43	0	1	0.29	0.45	0	1
*q3	0.28	0.45	0	1	0.25	0.43	0	1	0.28	0.45	0	1
*q4	0.15	0.36	0	1	0.26	0.44	0	1	0.14	0.35	0	1
age	39.06	8.31	20	54	39.11	8.27	20	54	39.30	8.20	20	54
sch	7.34	3.78	0	17	7.24	3.74	0	17	7.12	3.68	0	17
*female (reference male)	0.078	0.27	0	1	0.074	0.26	0	1	0.070	0.26	0	1
*urban (reference rural)	0.80	0.40	0	1	0.79	0.41	0	1	0.78	0.41	0	1
*non–participant1 (reference)	0.142	0.35	0	1	0.140	0.35	0	1	0.139	0.35	0	1
*emp1	0.790	0.41	0	1	0.796	0.40	0	1	0.803	0.40	0	1
*unemp1	0.069	0.25	0	1	0.065	0.25	0	1	0.058	0.23	0	1
*single	0.035	0.18	0	1	0.030	0.17	0	1	0.026	0.16	0	1
*married (reference)	0.91	0.28	0	1	0.92	0.27	0	1	0.92	0.26	0	1
*divorced	0.017	0.13	0	1	0.014	0.12	0	1	0.013	0.11	0	1
*widow	0.037	0.19	0	1	0.037	0.19	0	1	0.037	0.19	0	1
hhsize	4.18	1.76	1	25	4.22	1.76	1	23	4.27	1.76	1	23
No. of observations	47,373				25,324				19,437			

Table 2: FGM Tests of the Null Hypothesis That Attrition is Ignorable vs. It Is Not

Signs (*p*–values*) from the Probit Estimates for Model 5 Reported in Tables A3–A5

Labor force status during the initial visit	***A*(3)**	***A*(12)**	***A*(15)**
employed	−	0	0
	(0.037)	(0.476)	(0.729)
unemployed	+	+	0
	(0.001)	(<0.001)	(0.565)
Reference: non–participant			
Joint test *p*–value	<0.001	<0.001	0.60
Observations	47,373	25,324	19,437

*Based on the Standard Normal and the Chi Squared c.d.f.

Table 3. Attrition Patterns as of 3, 12, and 15 Months, Households with 20–54 Year–old Heads

Qualitative Results from the Probit Estimates for Model 5 Reported in Tables A3–A5

Variable *** denotes the characteristics of the HH head**	***A*(3)**	***A*(12)**	***A*(15)**
yr2001	0	+	0
yr2002	+	n.a.	n.a.
q2	+	0	+
q3	0	+	0
q4	0	0	0
*age	−	−	−
*age2	0	n.a.	n.a.
*age3	n.a.	n.a.	n.a.
*female	−	0	0
Urban	+	+	+
*emp1	−	0	0
*unemp1	+	+	0
*sch	−	–	+
*sch2	+	+	n.a.
*sch3	n.a.	n.a.	n.a.
*single	+	+	0
*divorced	+	0	0
*widow	0	0	0
Hhsize	–	–	–
hhsize2	+	+	+
hhsize3	–	n.a.	n.a.
Observations	47373	25324	19437
Log–likelihood w/o covariates	−14089	−8797	−5338
Log–likelihood w/ full set of covariates	−13652	−8587	−5235
LR test: Chi–sq (d.f.)	874 (19)	420 (16)	206 (15)

Reported signs (“+” or “–“) denote the signs for statistically significant coefficients at the 5 percent level or lower while “0” denotes non–significance; n.a. = not available (excluded).

Table 4. BGLW Tests of the Null Hypothesis That Attrition is Ignorable vs. It Is Not Signs (*p*–values*) from the Probit Estimates for Model II Reported in Tables A6–A7

	Outcome at initial round	
	Participant	**Unemployed**
Future attrition status (expected visits = 4)		
Attritor at 3 months	0	+
	(0.22)	(<0.001)
Attritor at 12 months	0	+
	(0.40)	(<0.001)
Attritor at 15 months	0	0
	(0.23)	(0.30)
Joint test *p*–value	0.23	<0.001
Observations	23,790	20,523
Future attrition status (expected visits = 3,4)		
Attritor at 3 months	0	+
	(0.38)	(<0.001)
Attritor at 12 months	0	+
	(0.21)	(<0.001)
Joint test *p*–value	0.27	<0.001
Observations	27,578	23,726
Future attrition status (expected visits = 2,3,4)		
Attritor at 3 months	0	+
	(0.17)	(<0.001)
Observations	47,373	40,668

*Based on the Standard Normal and Chi Squared c.d.f.

Table A1. Rotation Plan of the HLFS 2000–2002 and the Observation Plan of the Present Study[*]

	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
Expected no. of total 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.

Visit counter shows possible visit number values in a given round.

Expected no. of total visits shows number of times address should be in the sample.

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	26,244	25,669	24,770	12,587	13,397	14,201	14,009	14,376	13,670	13,391	12,025	0	**184,339**
p(A=1)	0.0831	0.1151	0.0910	0.0725	0.0789	0.0917	0. 0944	0.0949	0.0797	0.1328	0.0990		**0.0944**
A(9) : at risk	13,474	0	0	0	0	0	0	0	0	0	0	0	**13,474**
p(A=1)	0.1092												**0.1092**
A(12) : at risk	12,158	11,482	11,534	11,674	12,340	12,899	12,686	13,012	0	0	0	0	**97,785**
p(A=1)	0.1286	0.1150	0.1244	0.1266	0.1334	0.1332	0.1484	0.1235					**0.1294**
A(15) : at risk	10,594	10,161	10,099	10,196	10,694	11,181	10,803	0	0	0	0	0	**73,728**
p(A=1)	0.0853	0.0881	0.0957	0.0882	0.0806	0.1204	0.0838						**0.0919**
All Households													
A(3) : at risk	9,439	9,244	8,914	4,585	4,759	5,201	5,028	5,252	4,887	4,816	4,342	0	**66,467**
p(A=1)	0.0749	0.1049	0.0720	0.0624	0.0727	0.0808	0.0817	0.0864	0.0722	0.1296	0.0799		**0.0837**
A(9) : at risk	4,857	0	0	0	0	0	0	0	0	0	0	0	**4,857**
p(A=1)	0.0681												**0.0681**
A(12) : at risk	4,408	4,181	4,204	4,299	4,413	4,781	4,617	4,798	0	0	0	0	**35,701**
p(A=1)	0.0980	0.0897	0.1035	0.1042	0.1072	0.1094	0.1282	0.0965					**0.1048**
A(15) : at risk	3,976	3,806	3,769	3,851	3,940	4,258	4,025	0	0	0	0	0	**27,625**
p(A=1)	0.0709	0.0686	0.0738	0.0704	0.0607	0.1080	0.0561						**0.0730**
Households Headed by 20–54 Year–old Individuals													
A(3) : at risk	6,833	6,505	6,365	3,276	3,385	3,665	3,553	3,788	3,488	3,405	3,110	0	**47,373**
p(A=1)	0.0771	0.1101	0.0743	0.0687	0.0750	0.0849	0.0850	0.0913	0.0771	0.1389	0.0846		**0.0878**
A(9) : at risk	3,414	0	0	0	0	0	0	0	0	0	0	0	**3,414**
p(A=1)	0.0723												**0.0723**
A(12) : at risk	3,169	2,928	2,998	3,051	3,131	3,354	3,251	3,442	0	0	0	0	**25,324**
p(A=1)	0.0985	0.0943	0.1081	0.1105	0.1159	0.1139	0.1387	0.1020					**0.1104**
A(15) : at risk	2,857	2,652	2,674	2,714	2,768	2,972	2,800	0	0	0	0	0	**19,437**
p(A=1)	0.0763	0.0758	0.0767	0.0711	0.0647	0.1178	0.0629						**0.0783**

Table A3. Probit Estimates of Attrition at 3 Months, Households with 20–54 Year–old Heads

Dependent Variable: *A*(3) = 1 if attritor, = 0 else.

Variable ***denotes the characteristics of the HH head**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
yr2001	0.002	0.005	0.001	–0.003	–0.003
	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)
yr2002	0.090**	0.094**	0.088**	0.080**	0.079**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
q2	0.204**	0.205**	0.208**	0.207**	0.207**
	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)
q3	0.022	0.023	0.027	0.027	0.026
	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)
q4	0.052	0.049	0.051	0.047	0.047
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
*age		–0.162**	–0.161**	–0.077	–0.027**
		(0.055)	(0.055)	(0.057)	(0.010)
*age2		0.338*	0.337*	0.157	0.023
		(0.147)	(0.148)	(0.153)	(0.013)
*age3		–0.239	–0.24	–0.116	
		(0.129)	(0.129)	(0.133)	
*female		0.073*	0.066*	–0.086*	–0.088*
		(0.030)	(0.033)	(0.044)	(0.044)
urban		0.302**	0.298**	0.247**	0.246**
		(0.023)	(0.023)	(0.024)	(0.024)
*emp1			–0.02	–0.063*	–0.060*
			(0.029)	(0.029)	(0.029)
*unemp1			0.133**	0.130**	0.132**
			(0.039)	(0.039)	(0.039)
*sch				0.012	–0.026**
				(0.019)	(0.010)
*sch2				–0.284	0.276**
				(0.241)	(0.052)
*sch3				2.298*	
				(0.956)	
*single				0.185**	0.189**
				(0.046)	(0.045)
*divorced				0.218**	0.218**
				(0.064)	(0.064)
*widow				–0.028	–0.039
				(0.062)	(0.062)
hhsize				–0.174**	–0.174**
				(0.033)	(0.033)
hhsize2				2.029**	2.020**
				(0.535)	(0.535)
hhsize3				–6.759**	–6.734**
				(2.495)	(2.496)
constant	–1.452**	0.834	0.827	–0.039	–0.572**
	(0.018)	(0.655)	(0.656)	(0.691)	(0.182)
Observations	47373	47373	47373	47373	47373
Log–likelihood w/o covariates	–14089	–14089	–14089	–14089	–14089
Log–likelihood	–14019	–13826	–13814	–13649	–13652
LR test: Incremental Chi–sq. (d.f)	139 (5)	386 (5)	24 (2)	330 (9)	6 (2)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**).
All quadratic terms are scaled by 1/100, all cubic terms by 1/10000.

Table A4. Probit Estimates of Attrition at 12 Months, Households with 20–54 Year–old Heads

Dependent Variable: *A*(12) = 1 if attritor, = 0 else.

Variable *denotes the characteristics of the HH head**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
yr2001	0.077**	0.083**	0.078**	0.074**	0.075**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
q2	–0.016	–0.017	–0.015	–0.014	–0.014
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
q3	0.085**	0.084**	0.088**	0.092**	0.092**
	(0.029)	(0.030)	(0.030)	(0.030)	(0.030)
q4	–0.008	–0.01	–0.009	–0.011	–0.011
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
*age		–0.175*	–0.187**	–0.11	–0.011**
		(0.072)	(0.072)	(0.075)	(0.001)
*age2		0.400*	0.427*	0.258	
		(0.195)	(0.195)	(0.200)	
*age3		–0.32	–0.337*	–0.217	
		(0.170)	(0.170)	(0.175)	
*female		0.113**	0.149**	0.055	0.057
		(0.039)	(0.043)	(0.059)	(0.059)
urban		0.271**	0.270**	0.237**	0.234**
		(0.029)	(0.029)	(0.029)	(0.029)
*emp1			0.062	0.031	0.026
			(0.038)	(0.038)	(0.037)
*unemp1			0.265**	0.264**	0.260**
			(0.051)	(0.052)	(0.051)
*sch				–0.002	–0.027*
				(0.023)	(0.012)
*sch2				–0.085	0.267**
				(0.300)	(0.067)
*sch3				1.437	
				(1.214)	
*single				0.165**	0.219**
				(0.064)	(0.059)
*divorced				0.038	0.055
				(0.094)	(0.093)
*widow				–0.01	–0.012
				(0.079)	(0.078)
hhsize				–0.134**	–0.059**
				(0.044)	(0.018)
hhsize2				1.650*	0.381**
				(0.696)	(0.146)
hhsize3				–5.711	
				(3.217)	
constant	–1.281**	1.087	1.176	0.38	–0.922**
	(0.024)	(0.868)	(0.870)	(0.908)	(0.104)
Observations	25324	25324	25324	25324	25324
Log–likelihood w/o covariates	–8797	–8797	–8797	–8797	–8797
Log–likelihood	–8782	–8660	–8645	–8584	–8587
LR test: Incremental Chi–sq. (d.f)	29 (4)	244 (5)	31 (2)	123 (9)	8 (4)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**).
All quadratic terms are scaled by 1/100, all cubic terms by 1/10000.

Table A5. Probit Estimates of Attrition at 15 Months, Households with 20–54 Year–old Heads

Dependent Variable: *A*(15) = 1 if attritor, = 0 else.

Variable ***denotes the characteristics of the HH head**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
yr2001	0.034	0.041	0.04	0.036	0.036
	(0.028)	(0.029)	(0.029)	(0.029)	(0.029)
q2	0.176**	0.177**	0.177**	0.174**	0.174**
	(0.034)	(0.034)	(0.034)	(0.035)	(0.035)
q3	–0.008	–0.004	–0.004	–0.001	0
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
q4	0.021	0.018	0.019	0.015	0.017
	(0.046)	(0.047)	(0.047)	(0.047)	(0.047)
*age		–0.128	–0.13	–0.078	–0.005**
		(0.092)	(0.092)	(0.095)	(0.002)
*age2		0.287	0.29	0.175	
		(0.245)	(0.246)	(0.252)	
*age3		–0.216	–0.217	–0.136	
		(0.214)	(0.214)	(0.219)	
*female		0.157**	0.167**	0.136	0.141
		(0.049)	(0.055)	(0.076)	(0.075)
Urban		0.279**	0.279**	0.238**	0.236**
		(0.036)	(0.036)	(0.037)	(0.037)
*emp1			0.017	–0.01	–0.016
			(0.047)	(0.047)	(0.046)
*unemp1			0.043	0.046	0.039
			(0.069)	(0.069)	(0.068)
*sch				0.027	0.021**
				(0.029)	(0.004)
*sch2				–0.281	
				(0.387)	
*sch3				1.668	
				(1.568)	
*single				0.021	0.044
				(0.087)	(0.082)
*divorced				0.12	0.116
				(0.114)	(0.113)
*widow				–0.066	–0.067
				(0.099)	(0.098)
hhsize				–0.086	–0.085**
				(0.046)	(0.021)
hhsize2				0.633	0.576**
				(0.646)	(0.158)
hhsize3				–0.353	
				(2.435)	
constant	–1.488**	0.173	0.181	–0.421	–1.413**
	(0.029)	(1.108)	(1.108)	(1.152)	(0.115)
Observations	19437	19437	19437	19437	19437
Log–likelihood w/o covariates	–5338	–5338	–5338	–5338	–5338
Log–likelihood	–5316	–5266	–5266	–5233	–5235
LR test: Incremental Chi–sq. (d.f)	42 (4)	101 (5)	0 (2)	67 (9)	4 (5)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**).
All quadratic terms are scaled by 1/100, all cubic terms by 1/10000.

Table A6. Probit Estimates of the 1st Visit Participation Outcome, 20-54 Year-old Household Heads

Dependent Variable: LFP = 1 if participant, = 0 if not.

	Expected no. of visits					
Variables	**4**		**3,4**		**2,3,4**	
*** Characteristic of the HH head**	**I**	**II**	**I**	**II**	**I**	**II**
Attritor at 3 months	–0.002	–0.056	0.005	–0.037	–0.014	–0.042
$A(3) = 1, A(12) = A(15) = 0$	(0.038)	(0.046)	(0.034)	(0.042)	(0.025)	(0.030)
Attritor at 12 months	0.069*	0.036	0.074*	0.049		
$A(3) = 0, A(12) = 1, A(15) = 0$	(0.035)	(0.042)	(0.032)	(0.039)		
Attritor at 15 months	–0.05	–0.056				
$A(3) = A(12) = 0, A(15) = 1$	(0.039)	(0.047)				
yr2001	0.008	-0.017	–0.01	–0.033	–0.011	–0.045*
	(0.022)	(0.027)	(0.019)	(0.023)	(0.017)	(0.020)
yr2002					–0.076**	–0.061**
					(0.019)	(0.022)
q2	0.076**	0.166**	0.075**	0.161**	0.088**	0.169**
	(0.027)	(0.033)	(0.027)	(0.032)	(0.019)	(0.023)
q3	0.016	0.072*	0.016	0.070*	0.066**	0.135**
	(0.027)	(0.032)	(0.027)	(0.032)	(0.019)	(0.023)
q4	0.016	0.081	-0.019	0.044	0	0.059*
	(0.035)	(0.042)	(0.026)	(0.031)	(0.023)	(0.028)
*age		0.278**		0.278**		0.284**
		(0.014)		(0.013)		(0.010)
*age2		–0.436**		–0.434**		–0.437**
		(0.017)		(0.016)		(0.012)
*female		–1.793**		–1.765**		–1.711**
		(0.052)		(0.047)		(0.035)
urban		–0.315**		–0.334**		–0.292**
		(0.033)		(0.030)		(0.023)
*sch		–0.086**		–0.079**		–0.069**
		(0.013)		(0.012)		(0.009)
*sch2		0.707**		0.659**		0.638**
		(0.078)		(0.071)		(0.054)
*single		–0.458**		–0.507**		–0.481**
		(0.067)		(0.060)		(0.045)
*divorced		0.211*		0.209*		0.174**
		(0.089)		(0.081)		(0.057)
*widow		–0.287**		–0.250**		–0.268**
		(0.070)		(0.064)		(0.048)
hhsize		0.029**		0.022**		0.022**
		(0.008)		(0.007)		(0.005)

Table A6. Probit Estimates of the 1st Visit Participation Outcome, 20-54 Year-old Household Heads – Contd.

constant	1.057**	–2.154**	1.062**	–2.168**	1.052**	–2.440**
	(0.022)	(0.271)	(0.021)	(0.249)	(0.015)	(0.187)
Observations	23790	23790	27578	27578	47373	47373
Log-likelihood w/o covariates	–9518	–9518	–11152	–11152	–19316	–19316
Log-likelihood	–9510	–6416	–11142	–7612	–19294	–13300
LR test: Incremental Chi-sq (d.f.)	15 (7)	6188 (10)	20 (6)	7059 (10)	44 (6)	11988 (10)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**).
All quadratic terms are scaled by 1/100.

.

Table A7. Probit Estimates of the 1st Visit Unemployment Outcome, 20-54 Year-old Household Heads

Dependent Variable: Unemp = 1 if participant, = 0 if not (conditional on LFP =1).

	Expected no. of visits					
Variables	**4**		**3,4**		**2,3,4**	
*** Characteristic of the HH head**	**I**	**II**	**I**	**II**	**I**	**II**
Attritor at 3 months	0.213**	0.219**	0.198**	0.208**	0.168**	0.183**
A(3) = 1, *A*(12) = *A*(15) = 0	(0.045)	(0.046)	(0.041)	(0.042)	(0.030)	(0.031)
Attritor at 12 months	0.210**	0.210**	0.209**	0.214**		
A(3) = 0, *A*(12) = 1, *A*(15) = 0	(0.040)	(0.041)	(0.037)	(0.038)		
Attritor at 15 months	0.047	0.053				
A(3) = *A*(12) = 0, *A*(15) = 1	(0.051)	(0.052)				
yr2001	0.149**	0.166**	0.179**	0.198**	0.187**	0.205**
	(0.028)	(0.029)	(0.024)	(0.025)	(0.021)	(0.022)
yr2002					0.298**	0.316**
					(0.023)	(0.024)
q2	–0.127**	–0.125**	–0.127**	–0.123**	–0.167**	–0.169**
	(0.034)	(0.035)	(0.034)	(0.035)	(0.024)	(0.024)
q3	–0.140**	–0.144**	–0.142**	–0.146**	–0.179**	–0.185**
	(0.034)	(0.035)	(0.034)	(0.035)	(0.024)	(0.025)
q4	–0.109*	–0.108*	–0.041	–0.036	–0.065*	–0.064*
	(0.047)	(0.047)	(0.033)	(0.034)	(0.029)	(0.030)
*age		–0.029		–0.021		–0.038**
		(0.016)		(0.015)		(0.011)
*age2		0.036		0.026		0.047**
		(0.020)		(0.019)		(0.014)
*female		0.211*		0.195*		0.211**
		(0.093)		(0.085)		(0.063)
Urban		0.308**		0.305**		0.336**
		(0.037)		(0.033)		(0.025)
*sch		–0.055**		–0.058**		–0.064**
		(0.004)		(0.004)		(0.003)
*single		0.368**		0.353**		0.325**
		(0.078)		(0.073)		(0.056)
*divorced		0.261*		0.282*		0.464**
		(0.130)		(0.117)		(0.080)
*widow		0.144		0.144		0.039
		(0.138)		(0.125)		(0.095)
Hhsize		0.015		0.019*		0.020**
		(0.008)		(0.008)		(0.006)
Constant	–1.479**	–0.909**	–1.491**	–1.029**	–1.445**	–0.645**

Table A7. Probit Estimates of the 1st Visit Unemployment Outcome, 20-54 Year-old Household Heads – Contd.

	(0.029)	(0.294)	(0.027)	(0.272)	(0.019)	(0.205)
Observations	20523	20523	23726	23726	40668	40668
Log-likelihood w/o covariates	–5386	–5386	–6449	–6449	–11368	–11368
Log-likelihood	–5337	–5190	–6386	–6201	–11227	–10834
LR test: Incremental Chi-sq (d.f.)	97 (7)	293 (9)	126 (6)	370 (9)	281 (6)	786 (9)

Standard errors are in parentheses. Asterisks denote statistical significance at the 5 (*), and 1 percent (**). All quadratic terms are scaled by 1/100.