ORIGINAL ARTICLE   **Open Access**

The Egypt labor market panel survey: introducing the 2012 round

Ragui Assaad[1*] and Caroline Krafft[2]

* Correspondence: assaad@umn.edu
[1]Humphrey School of Public Affairs, University of Minnesota, 301 19th Ave. S, Minneapolis, MN 55455, USA
Full list of author information is available at the end of the article

Abstract

This paper introduces the 2012 round of the Egypt Labor Market Panel Survey (ELMPS), a publicly-available nationally representative longitudinal household survey. We describe the key characteristics of the ELMPS, including the samples and questionnaires for each round. Additionally, we examine the attrition processes observed in the panel and discuss the creation of weights to correct for attrition. We compare our data to other statistical sources for Egypt to evaluate the sample's representativeness. To demonstrate how the ELMPS allows for an improved understanding of the labor market, we compare unemployment trends using the ELMPS and other data on unemployment in Egypt.

JEL codes: J00, C81, C83, J64

Keywords: Survey panel data; Public use data; Sample weights; Unemployment; Egypt

1. Introduction

The Egypt Labor Market Panel Survey, carried out by the Economic Research Forum (ERF) in cooperation with Egypt's Central Agency for Public Mobilization and Statistics (CAPMAS) since 1998, has become the mainstay of labor market and human resource development research in Egypt, being the first and most comprehensive source of publicly available micro data on the subject. The 2012 round of the survey provides a unique opportunity to ascertain the impact of the momentous events accompanying the January 25th revolution on the Egyptian economy and labor market and on the lives of Egyptian workers and their families. The Egypt Labor Market Panel Survey of 2012 (ELMPS 2012) is the third round of this longitudinal survey, which was also carried out in 1998 and 2006[1].

Like other labor markets in the MENA region, the Egyptian labor market suffers from high youth unemployment, low and stagnant rates of female labor force participation, rapid increases in educational attainment, but poor quality education, anemic private sector growth and continued informalization, and strong preference for public sector employment among new entrants (see Assaad 2013). In addition, the fact that Egypt is in the midst of a deep economic downturn, which started with the world financial crisis of 2008/9 and deepened considerably after the January 25th 2011 revolution, affords an opportunity to study how labor markets react to such economic downturns. Anecdotal evidence suggests that although hiring slows in such crises, there are appear to be few outright layoffs and relatively little job destruction. Instead, real earnings appear to adjust downward and employment in some sectors becomes more precarious and irregular. The

economic crisis has also interacted in complex ways with the evolution of the youth bulge, whose peak seems to have shifted beyond the age of labor market entry, thus placing reduced pressure on labor supply, just at the time when labor demand slowed. The 2012 round of the ELPMS will allow for a careful examination of all these issues and many others. It contains a great deal of information about current employment characteristics for a nationally representative sample of individuals, but also allows for the study of labor market dynamics and employment trajectories over time, both through the panel design as well as through a large number of retrospective questions about the individual's employment history.

The ELMPS is also notable for the fact that it allows for clear distinctions between market and subsistence work, allowing flexibility for researchers to explore various notions of economic activity, such as the market and extended labor force definitions. It uses a battery of questions to identify participation in market and subsistence work, an approach that is particularly critical to correctly identifying the work status of women and children. Additionally, the survey allows for various definitions of unemployment to be explored by asking separate questions about the desire to work, availability for work, and search behavior. Having detected employment, the ELMPS collects a great deal of information about the characteristics of employment, including its formality, regularity, the form it takes, the benefits that are associated with it, the nature of the workplace, and the conditions of work. The 2012 round includes some subjective assessments about satisfaction with one's work and attitudes about changes that occurred since the January 25th revolution. It also asks respondents to report about some of the characteristics of their employers, where relevant, such as size, legal form, formality status, export orientation, and degree of reliance on tourism.

In addition to standard labor market topics such as labor force participation, employment, unemployment, and earnings, the ELMPS also covers topics such as parental background, education, housing, access to services, residential mobility, migration and remittances, time use, marriage patterns and costs, fertility, women's decision making and empowerment, savings and borrowing behavior, and the operation of household enterprises and farms. Besides the survey's panel design, which permits the study of various phenomena over time, the survey also contains a large number of retrospective questions about the timing of major life events such as education, residential mobility, jobs, marriage and fertility. The survey provides detailed information about place of birth and subsequent residence, as well information about schools and colleges attended at various stages of an individual's trajectory, which permit the individual records to be linked to information from other data sources about the geographic context in which the individual lived and the educational institutions s/he attended. These are just a few of the features of the ELMPS that make it a particularly valuable data source for researchers, and also an excellent model for researchers interested in undertaking similar surveys.

This paper introduces the 2012 round of the ELMPS and coincides with the release of the public use micro data from this round as well as the integrated data from the previous rounds in both repeated cross section and panel formats. We proceed as follows. First, we describe the samples and questionnaires for each survey round, as well as information on access to the micro data. Second, we describe the data collection process in the 2012 round and present information on the attrition processes observed in the panel and the creation of weights to correct for this attrition. Third, we present a comparison of our data to other statistical sources for Egypt to evaluate the representativeness of the sample. We also

present a comparison of unemployment trends based on other sources and the ELMPS to demonstrate how the collection of the high quality data provided in the ELMPS allows for a substantially improved understanding of labor market trends and conditions.

1.1. The samples

The first round of the panel, the Egypt Labor Market Survey of 1998 (ELMS 1998) was carried out on a nationally representative sample of 4,816 households containing 23,997 individuals[2]. The original sample of the ELMS 1998 was a two-stage stratified random sample selected from a CAPMAS master sample[3]. Two hundred primary sampling units (PSUs) were systematically selected from among the enumeration areas of the CAPMAS master sample from geographic strata representing the 22 governorates covered by the survey divided into urban and rural subcomponents, where relevant. Urban areas were over-represented in that sample (140 out of 200 PSUs at a time when the rural population made up 56.8 percent of the population), an issue that was taken into account in the creation of sampling weights.

The ELMPS 2006 followed the initial ELMS 1998 sample, locating 3,685 households from the original 4,816 surveyed in 1998, and adding 2,168 new households that emerged from these households as a result of splits, as well as a refresher sample of 2,498 households, all totaling 8,351 households containing 37,140 individuals. In 2006, the refresher sample of households was selected from an additional 100 PSUs randomly selected from a new master sample prepared by CAPMAS. The additional PSUs were selected to represent the actual distribution of rural and urban areas across governorates at that time.

The ELMPS 2012 is therefore the third round of a periodic longitudinal survey that tracks the labor market and the demographic characteristics of households and individuals interviewed in 2006, some of whom were also interviewed in 1998. The field work for the ELMPS 2012 was carried out from March to June of 2012. The final sample for 2012 round of the survey was 12,060 households, consisting of 6,752 households from the 2006 sample, 3,308 new households that emerged from these households as a result of splits, and a refresher sample of 2,000 households. This refresher sample was selected from an additional 200 PSUs selected from an updated CAPMAS master sample. It was designed to over-sample areas with high migration rates to increase the representation of current migrants in the sample[4]. Of the 37,140 individuals interviewed in the 2006 survey, 28,770 (77 percent) were successfully re-interviewed in 2012. These individuals, 13,218 of whom were also tracked in 1998, form a panel that can be used for longitudinal analysis. The 2012 sample also includes 20,416 new individuals adding up to a total of 49,186 individuals.

The attrition that occurred from the original 1998 sample to 2006 was mostly random in nature, due to the loss of records containing identifying information for 1998 households (Assaad and Roushdy 2009). The attrition that occurred from 2006 to 2012 was due to a variety of processes. We discuss the factors related to these attrition processes and how they are corrected for using non-response weights below.

1.2. The questionnaires

The 1998, 2006, and 2012 questionnaires all have three major sections: (i) a household questionnaire administered to the head of household or his or her spouse, requesting

information about basic demographic characteristics of all household members, owner-ship of durable goods and assets, and housing conditions, services, and facilities; (ii) an individual questionnaire administered to every individual in the household age six and above; and (iii) an enterprise and income module that collects information on enter-prises operated by the household, as well as all non-labor income sources. In 2006 and 2012 this section also includes information on current migrants, remittances, and transfers.

The questionnaires for the different survey rounds are intentionally similar to ensure data comparability over time. However, additional modules and questions have been added to the surveys over time. Table 1 lists the different modules in the individual questionnaire, and indicates which modules were included in all three rounds, as well as modules added in 2006 and 2012. The initial 1998 ELMS and subsequent surveys collected information on parents' characteristics, respondents' education, employment, and unemployment, job characteristics, job mobility and geographic mobility, and earn-ings. The 2006 ELMPS added modules on siblings' characteristics, fertility, women's sta-tus, the cost of marriage, and questions on the characteristics of the first job. The 2006 round also added additional questions to detect employment and a more detailed edu-cation history. The ELMPS 2012 added a life events calendar, tracking education, mar-riage, work, and migration over time, as well as modules on health, return migration, information technology, and savings and borrowing. Additionally, in 2012, the module on female employment was extended to include previous as well as current wage workers. Questions were added on non-wage earnings in 2012. The mobility module was also expanded to better capture information on first jobs and questions were in-cluded about any job changes since the revolution.

As well as including detailed labor market data, the ELMPS surveys include detailed information on education, geographic residence, occupation, and economic activity, which can be linked to other sources of data on Egypt. The data includes the codes for all schools attended, as well as for the educational certificate an individual obtained. Geographic codes identifying up to four different places of residence are collected. Additionally, detailed economic activity and occupation codes are collected. This level of detail allows researchers to connect the ELMPS data to other datasets on Egypt, and further expands the opportunities for research and analysis.

Table 1 Modules in the individual questionnaire, 1998, 2006, 2012

In 1998, 2006, & 2012	Added in 2006 (and included in 2012)	Added in 2012 only
Parents' characteristics	Siblings' characteristics	Life events calendar
Education	Characteristics of first job	Health
Employment	Fertility	Return migration
Unemployment	Women's status	Information technology
Job characteristics	Cost of marriage	Savings & borrowing
Secondary jobs		
Female employment		
Job mobility		
Geographic mobility		
Earnings		

1.3. Public use micro data access

The public use micro data for the 1998 and 2006 surveys, as well as a harmonized dataset including the 1988 special LFS, are available from ERF at www.erfdataportal. com. The public use micro data and documentation for the 2012 round will become available on November 1, 2013 from ERF, also at www.erfdataportal.com. The public use micro data will include the full 2012 cross-section, as well as a harmonized pooled cross-section dataset covering 1988–2012, and a panel for all the individuals included in the 1998–2012 rounds. Documentation on the creation and definition of variables will also be made available. Additionally, with the release of the 2012 data, we will release the STATA do files used to generate and harmonize the 2012 datasets; interested users will be able to recreate all variables and weights from the raw data in 2012.

ERF will provide the data and documentation free of charge to researchers as long as they meet certain basic criteria. ERF requires users to fill out a data usage agreement, and collects data on users, restricts data use to educational and scholarly activities, and requires the user to cite ERF as the source of data. Dataset inquiries can be made to erfdataportal@erf.org.eg. Data is made available as STATA .dta files and includes English labels. Accompanying documentation includes codebooks for the data and technical reports, as well as English and Arabic versions of the actual questionnaires and enumerators' manuals used in the field.

2. Data collection, sample attrition and the calculation of sample Weights[5]

2.1. Data collection

The 2012 data collection process proceeded in two phases. First, in late 2011, an enumeration phase was undertaken. This phase focused on locating households and individuals from the 2006 sample. If households or individuals had moved, every effort was made to collect current contact information. Additionally, the refresher sample was designed to over-sample high-migration areas, and refresher sample PSUs and households were randomly selected based on this sampling approach (see below for more information on the selection of the 2012 refresher sample).

The 2012 ELMPS was implemented by 39 teams in the field, each consisting of one supervisor, one reviewer, and four enumerators. Additionally, there were two teams undertaking quality control. All interviewers were trained for 10 days by the technical director and CAPMAS prior to fielding the survey. The fielding of the full 2012 survey took place from March 1, 2012 to June 10, 2012, with more than 90% of households and individuals surveyed during March and April. Desk review, coding, and data entry and validation at CAPMAS occurred after fielding.

2.2. The pattern of attrition from 2006 to 2012

Sample attrition is an issue that arises in all longitudinal surveys due to the possibility that a household interviewed in a previous round of the survey cannot be located again in a subsequent round or that an individual who split from existing households could not be found. Attrition is rarely entirely random as the probability that households move in their entirety or that individuals move to unknown locations can often depend on both the observable and unobservable characteristics of the households and the individuals. If such attrition is not addressed, it calls into question the representativeness

of the sample and could introduce significant bias in any analysis. The objective of this section is to analyze the nature of the attrition that occurred from 2006 to 2012 and use that analysis to estimate appropriate sampling weights to at least partially adjust for the occurrence of attrition[6]. By necessity, the adjustments we make will only adjust for attrition that is associated with observable characteristics of individuals and households and will not be able to address attrition based on unobservables. Adjustment factors for attrition are combined with weights from previous rounds to create the appropriate weights for panels that follow individuals from 1998 to 2012, from 1998 to 2006, and from 2006 to 2012. In addition to panel weights, the paper describes how we calculated weights to account for the sampling design of the refresher sample that was added in the 2012 round of the survey and how those weights were combined with those of the attrition-corrected original sample to create a nationally-representative cross-section in 2012[7]. The refresher sample over-sampled areas with high rates of migration in order to improve statistical power for studying issues related to migration. This section includes a discussion of our construction of sampling weights to account for this sampling strategy.

As noted earlier, the final sample interviewed in 2012 consists of 12,060 households, which includes 6,752 original households from the 2006 sample, 3,308 new households that emerged from those households as a result of splits (i.e. split households), and a refresher sample of 2,000 households. In terms of individuals, the 2012 sample includes 49,186 individuals, who are made up of 28,770 individuals initially interviewed in 2006 and successfully re-interviewed in 2012 and 20,416 'new' individuals. Of those new individuals, 5,009 joined the original 2006 households, 6,900 joined split households, and 8,507 were members of the refresher sample of households.

In analyzing attrition in the ELMPS 2012 sample, we distinguish between two types of attrition processes, namely (i) the inability to locate an entire household interviewed in 2006, and (ii) the inability to track a new household that resulted from one or more individuals who split together from an original 2006 household. In the second case, we are considering attrition among households that split off from original households, with split households consisting of either the single individual who had left the original household, or multiple individuals who had left together and any additional individuals who may have joined them later. The first type of attrition (Type I) was mainly due to a failure to track households due to either inaccurate address information or because they moved to an unknown location. The second type of attrition (Type II) occurred when the original household had been located, but one or more individuals in that household could not be accounted for in 2012 nor tracked into a split household. We use observable characteristics of households and individuals in 2006 to predict the probability that each of these two attrition processes will occur. We then use these predicted probabilities to create adjustment weights to apply to households found in 2012 to correct both the panel and cross sectional samples for attrition.

2.2.1. Type I attrition: attrition of original households from 2006 to 2012

In November and December of 2011, three months prior to the commencement of the main fieldwork for the 2012 round of the survey, an enumeration phase of the fieldwork was conducted where the field teams attempted to locate all of the households interviewed in 2006 and enumerate their members. If a household was found in its original location, its address information was checked for accuracy and information was

collected about its existing members and about any who may have left or died[8]. If a household was not found in its original location, there was an attempt to determine the reason for the absence, including the death of all household members, emigration of all household members to another country, or a move to another location[9]. In case of the latter, an attempt was made to obtain the new address from the neighbors. Table 2 shows the status of all 8,351 households interviewed in 2006 at the enumeration stage at the end of 2011.

Equipped with the information obtained in the enumeration phase, the fieldwork team visited the households that were found or those that moved to a known address again in the period from March-June 2012. In addition to the households that were determined to have died out (135), left the country in their entirety (41) and that refused to be enumerated (16), there were 1,108 households that could not be located at all in the enumeration phase or that had moved to an unknown address. These were necessarily excluded from the start of the 2012 fieldwork (see Table 2). Of the 7,051 original 2006 households that remained in place or moved to a known address, 287 were not located in 2012 due to inaccurate address information. An additional two households left the country in the interim period, one refused to be interviewed, and an additional nine died out entirely. Ultimately 6,752 households of the original 8,351 were found and interviewed in 2012. As shown in the last column of Table 2, of the remaining 1,599 households, 43 left the country in their entirety, 144 died out completely, 17 refused to respond and a total of 1,395 could not be located. We refer to the loss of the latter two categories (1,412 households) as Type I attrition. This attrition is relative to the 8,164 households that would have been around in 2012 had no attrition occurred, i.e. the original 8,351 minus the 43 that left the country and the 144 that died out completely. This implies a Type I attrition rate of 17.3% at the household level.

Of the 37,140 individuals who were interviewed in 2006, 31,589 were members of households that were successfully tracked in 2012, while 5,212 were members of households that were not successfully tracked, and 339 were members of households that left the country or died out entirely. The Type I attrition rate at the individual level is therefore 14.2%.

2.2.2. Type II attrition: attrition among individuals who split whose original households were found

The second type of attrition we examine is that involving individuals who are no longer with their original 2006 households, conditional on these households being found in 2012. Natural attrition in this case involves individuals who died or left the country. All

Table 2 Status of household at the end of the enumeration state in December 2011 & at the main fieldwork stage in March-June 2012

	2011 status	2012 updates	Final 2012 status
Initial households	8,351	7,051	8,351
Household died out	135	9	144
Household left the country	41	2	43
Household refused	16	1	17
Household not found or at unknown address	1,108	287	1,395
Household located or at a known address	7,051	6,752	6,752

other individuals who could not be located are considered as having experienced Type II attrition.

When a household was found at the enumeration stage, the enumerators inquired about the status of each of its 2006 individuals. If these individuals were no longer present, they marked the reason for their absence (death, emigration, move to another location within Egypt). If the move was to a location within Egypt, the enumerators noted the individual's new location as best they could and attempted to gather any available contact information.

The 6,752 original households that were found at the fieldwork stage in 2012 would have had 31,589 individuals in 2006 (Table 3). Of those individuals, 24,656 were found still in their original households and 6,933 were not found with their original households. Of those, 1,139 had experienced natural attrition through death and migration, implying that the remaining 5,794 had split off to form independent households within Egypt. After accounting for multiple individuals who moved to the same location, these 5,794 individuals were found to have formed 4,931 potential 'new households', some of which contained multiple individuals from a 2006 household.

An additional complication emerged for the 1,680 individuals who did not show up in the 2012 data (either in the original households or as splits) although they had been included in lists obtained from the enumeration phase. For these individuals we have no information about whether they died, left the country, or moved to another location within the geographical scope of the survey. We cannot therefore distinguish for this group between natural attrition and Type II attrition. In general we treat these 1,680 individuals as having experienced Type II attrition with the exception of three categories that we exclude from the attrition model because of the presumption that their attrition is probably due to death. These are (i) "heads" of split households above the age of 65

Table 3 Status of individuals and split households in 2011 & 2012, conditional on 2006 household being found

	Final 2012 status
Individuals present in 2006 in original households found in 2012	31,589
Individuals still in original households in 2012	24,656
Individuals no longer in original households in 2012	6,933
Natural attrition through death and migration	1,139
Individual known to have died	912
Individual who emigrated or left for a governorate outside scope of survey	227
Individual splits to form households within Egypt	5,794
Potential split households (accounting for individuals who split together)	4,931
Households excluded from type II attrition model	
-Potential split households made up of individuals 65 or older, permanently handicapped or under six who were not found in 2012	203
-Households whose split individual is in the above three categories who were found 2012	46
Potential split households minus exclusions	4,682
Split households found	3,262
Split households not found (attrited)	1,420

in 2006 who were not found in 2012 (70 cases), (ii) "heads" of split households who were children under the age six who did not split with other adult members of the household (122 cases), (iii) "heads" who were permanently handicapped individuals (11 cases). This leaves 5,574 individuals who split from their 2006 households as of 2012, forming 4,728 potential split households. We also exclude from the Type II attrition model 46 split households that were found whose split head was in one of the above three categories[10]. This leaves 4,682 'split households' to examine for Type II attrition, of which 3,262 were successfully located and 1,420 were not, for a Type II attrition rate of 30.3%.

2.3. Predicting attrition in the ELMPS 2012

In this section we perform bivariate and multivariate analyses to investigate the potential observable correlates of attrition for each of the two possible attrition processes. We use the multivariate analysis of each type of attrition to predict attrition probabilities as a function of observables and use these probabilities to formulate appropriate weights to adjust for attrition.

2.3.1. Correlates of type I attrition

Table 4 shows the 2006 characteristics of the households and the household heads by the household attrition status in 2012. The sample in this analysis consists of the 8,164 original households that were interviewed in 2006 and did not completely die out or leave the country. The table shows a number of statistically significant differences between the characteristics of found households and those that attrited.

The households that attrited had a smaller number of children 6 to 14 in 2006, as well as fewer working age males and females, and slightly fewer elderly males. The households that attrited were much more likely to be from Greater Cairo, Alexandria & Suez Canal, and somewhat more likely to be from urban Lower Egypt, but less likely to be from either rural Lower Egypt or Rural Upper Egypt than households that did not attrite. The households that attrited were also much less likely to own their dwelling, and, if renting, were much more likely to have rented under the "new," more flexible rental laws.

In terms of household head characteristics, the heads of households that attrited were more likely to be divorced, less likely to be illiterate or read & write, and more likely to be secondary or university educated. The heads of households that attrited were more likely than the heads of those that did not attrite to be engaged in formal or informal private regular wage work, but less likely to be engaged in informal temporary work or to be a household enterprise worker.

The multivariate analysis for Type I attrition is presented as marginal effects from a probit model for the probability that a household was not found (attrited), based on the same set of characteristics presented in Table 4. Table 5 presents the results of the probit regression as marginal effects. The reference case, in terms of 2006 variables, is at the mean of the household composition variables, lived in Greater Cairo, owned his or her dwelling, was 35 or older, male, single, illiterate or read & write, and was employed in Government work. The base probability of Type I attrition, for this reference, omitted case, is 0.149. Consistent with Table 4, the probit shows a 2.3 percentage point lower probability that a household will attrite for each additional child age 6 to 14, a 3.1 percentage point lower probability of attrition for each additional working age male, a 2.0 percentage point lower

Table 4 2006 household characteristics by household attrition status

Household characteristics	Attrited	Found	Significant difference	Total
Household composition (Mean)				
No. of children 0 to 5 in household	0.68	0.68		0.68
No. of children 6 to 14 in household	0.45	0.77	***	0.72
No. of working age males in household	1.16	1.51	***	1.45
No. of working age females in household	1.24	1.51	***	1.46
No. of elderly males in household	0.08	0.10	**	0.10
No. of elderly females in household	0.09	0.11		0.11
Region (Percentage)				
Greater Cairo	30.8	11.5	***	14.9
Alex & Suez Canal	19.3	10.2	***	11.8
Urban Lower	16.1	13.8	*	14.2
Urban Upper	17.5	17.6		17.6
Rural Lower	8.7	26.9	***	23.7
Rural Upper	7.5	20.1	***	17.9
Ownership status (Percentage)				
Own	58.5	81.5	***	77.5
Old rental law	35.3	17.5	***	20.6
New rental law	6.2	1.0	***	1.9
HH head age group (Percentage)				
15-24	2.9	2.6		2.6
25-34	31.2	20.0	***	21.9
35+	65.9	77.4	***	75.4
Sex of HH head (Percentage)				
Male	83.1	83.9		83.7
Female	16.9	16.1		16.3
HH head marital status (Percentage)				
Single	2.1	1.8		1.8
Married	83.0	83.1		83.1
Divorced	2.1	1.1	**	1.3
Widowed	12.9	14.0		13.8
HH head education (Percentage)				
Illiterate or read & write	26.8	44.1	***	41.1
Primary/preparatory	15.3	16.0		15.9
Secondary	32.3	26.8	***	27.8
University	25.6	13.1	***	15.2
HH head status/sector of employment (Percentage)				
Government	26.0	26.6		26.5
Formal private regular wage	13.7	6.5	***	7.7
Informal private regular wage	12.7	8.4	***	9.1
Informal temporary	2.3	4.7	***	4.3
HH enterprise worker	20.0	28.5	***	27.0
Out of labor force or manpower basis	24.1	24.5		24.4
Unemployed	1.3	0.8		0.9
No. of households	1,412	6,752		8,164

Asterisks indicate significance of a bivariate difference in proportions test: *p < 0.05; **p < 0.01; ***p < 0.001.

Table 5 Marginal effects from probit regression of household and household head characteristics accounting for type I attrition

Base probability	**0.149**
Household composition in 2006	
No. of children 0 to 5 in household	−0.003
No. of children 6 to 14 in household	−0.023***
No. of working age males in household	−0.031***
No. of working age females in household	−0.020**
No. of elderly males in household	−0.040*
No. of elderly females in household	−0.011
Region of residence in 2006 (Greater Cairo omitted)	
Alex & Suez Canal	−0.037**
Urban Lower	−0.077***
Urban Upper	−0.086***
Rural Lower	−0.127***
Rural Upper	−0.117***
Home ownership status in 2006 (Own omitted)	
Old rental law	0.065***
New rental law	0.292***
HH head age group in 2006 (35+ omitted)	
15-24	0.017
25-34	0.034*
HH head gender (Male omitted)	
Female	0.019
HH head marital status in 2006 (Single omitted)	
Married	0.069*
Divorced	0.105
Widowed	0.038
HH head education in 2006 (Illiterate or read & write omitted)	
Primary/Preparatory	0.002
Secondary	0.026*
University	0.084***
HH head status/sector of employment in 2006 (Government omitted)	
Formal private regular wage	0.067***
Informal private regular wage	0.069***
Informal temporary	−0.036
HH enterprise worker	0.008
Out of labor force/manpower	0.009
Unemployed	0.043
No. of households	8,164

Dependent variable: probability that the original household attrited (was not found).
Asterisks indicate significance of marginal effects: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
Marginal effects are estimated at the omitted category for binary and categorical variables, and at the mean for continuous variables.

probability for each additional working age female, and a 4.0 percentage point lower probability for each additional elderly male.

The probability of attrition from 2006 to 2012 is lower in all other regions than Cairo. Compared to the omitted category of Cairo, the Alexandria and Suez Canal region has

a 3.7 percentage point lower probability of attrition. Urban Lower Egypt has a 7.7 percentage point lower probability of attrition, while urban Upper Egypt has an 8.6 percentage point lower probability of attrition. In rural Lower Egypt the probability of attrition is 12.7 percentage points lower and rural Upper Egypt is 11.7 percentage points lower than Greater Cairo. Households under the old rental laws are 6.5 percentage points more likely to attrite than those that owned their homes, and households under the new rental laws are 29.2 percentage points more likely to attrite than owners.

Compared to households with heads age 35 or older, households with heads age 25–34 are 3.4 percentage points more likely to attrite. Compared to households with single heads, households with married heads are 6.9 percentage points more likely to attrite. Households with secondary educated heads are 2.6 percentage points more likely to attrite, and with university educated heads are 8.4 percentage points more likely to attrite, compared to illiterate or read & write heads. Households headed by an individual in formal regular wage employment are 6.7 percentage points more likely to attrite than government workers, and households headed by informal private regular wage workers are 6.9 percentage points more likely to attrite than government workers.

2.3.2. Correlates of type II attrition

Among 2006 households that were found in 2012, if an individual or group of individuals split together, we modeled whether or not their "potential split household" was found in 2012. Type II attrition is when such a potential split household is not found. After accounting for households that were presumed dead, and excluding potential split household heads who were either over 65, under six, or permanently handicapped, there are 4,682 potential split households. Most of these were formed by single individuals who split, but some are made up of multiple individuals from the same 2006 household who split together. In such cases, the characteristics of the first individual listed in the group were used as the characteristics of the "split household head"[11].

As shown in Table 3, of the potential split households, 1,420 were not found and 3,262 were found. Table 6 presents the bivariate relationship between the 2006 characteristics of the "split household head" and whether the split household was successfully tracked. All variables except one are based on characteristics in 2006. The variable "Female: Separated" identifies women who were not found with their 2006 husband, but whose (2006) husband was found in 2012. In the field, in households where the woman had separated from their husbands, the husbands were disinclined to share the woman's current location.

As can be seen in Table 6, there are significant differences between the characteristics of the head of split households that were found, and those that attrited. Households that attrited were more likely to have male heads. Attrited split households were more likely to have a new head age 35–64, and less likely to have a new head 15–24 or 25–34. Split households that were not found were more likely to have heads who were married, but less likely to have heads that were single or divorced. Split households with female heads that were not found were more likely to be separated.

Table 6 Correlates of type II attrition

	Attrited	Found	Significant difference	Total
Region of residence in 2006 (Percentage)				
Greater Cairo	17.4	6.9	***	10.1
Alex & Suez Canal	6.4	7.0		6.8
Urban Lower	12.1	12.0		12.1
Urban Upper	14.4	15.3		15.0
Rural Lower	27.2	33.8	***	31.8
Rural Upper	22.5	25.0		24.3
Sex of "head" (Percentage)				
Male	51.5	47.5	*	48.7
Female	48.5	52.5	*	51.3
Age of "head" in 2006 (Percentage)				
6-14	8.9	8.0		8.3
15-24	54.2	62.2	***	59.8
25-34	21.3	24.8	**	23.7
35-64	15.6	5.0	***	8.2
Marital status of "head" in 2006 (Percentage)				
Single	72.6	85.6	***	81.7
Married	25.6	12.4	***	16.4
Divorced	0.5	1.1	*	0.9
Widowed	1.3	0.8		1
Female: separated	5.8	1.3	***	2.7
Education of "head" in 2006 (Percentage)				
Illiterate or read & write	5.8	1.3	***	2.7
Primary/preparatory	17.2	12.8	***	14.2
Secondary	22.5	28.2	***	26.5
University	39.1	40.4		40.0
Status/sector of employment of "Head" in 2006 (Percentage)				
Government	13.2	11.1	*	11.7
Formal private regular wage	7.9	7.1		7.3
Informal private regular wage	4.6	5.3		5.1
Informal temporary	11.5	12.4		12.1
HH enterprise worker	3.9	4.1		4.0
Out of labor force & not student	12.4	13.8		13.4
Out of labor force & student	24.9	21.9	*	22.8
Unemployed	26.6	27.7		27.4
No. split households	1,420	3,262		4,682

Asterisks indicate significance of a bivariate difference in proportions test: *p < 0.05; **p < 0.01; ***p < 0.001.

Looking at the education of the split household head, attrited households were more likely to be headed by an illiterate or read & write head, or a head with a primary/preparatory education, but less likely to be headed by an individual with a secondary education. Turning to the new head's status/sector of employment, attrited households are only slightly more likely to be out of the labor force and not a student, or in government work. While Type II attrition does not appear to be a random process, differences are primarily based on demographic characteristics, and not on economic ones.

Table 7 presents the multivariate analysis of Type II attrition. The dependent variable is the probability of a split household attriting (not being found), given that the original 2006 household was found. In addition to the variables used in Table 6, we include

Table 7 Marginal effects from probit regression for type II attrition among split households, conditional on original (2006) household found

Base probability	0.708
	Marginal effect
Region of residence in 2006 (Greater Cairo omitted)	
Alex & Suez Canal	−0.245***
Urban Lower	−0.227***
Urban Upper	−0.231***
Rural Lower	−0.288***
Rural Upper	−0.261***
Sex of "head" (Male omitted)	
Female	−0.251***
Age of "head" in 2006 (35–64 omitted)	
6-14	0.093
15-24	−0.260***
25-34	−0.314***
Female: 6-14	−0.183
Female: 15-24	0.139***
Female: 25-34	0.159***
Marital status of "head" in 2006 (Single omitted)	
Married	0.087**
Divorced	−0.439*
Widowed	0.119
Female: married	0.117**
Female: divorced	0.237***
Female: widowed	−0.048
Female: separated	0.121**
Education of "head" in 2006 (Illiterate or read & write omitted)	
Primary/preparatory	−0.051*
Secondary	0.008
University	0.037
Status/sector of employment of "head" in 2006 (Government omitted)	
Formal private regular wage	−0.020
Informal private regular wage	0.039
Informal temporary	0.060
HH enterprise worker	0.032
Out of labor force & not student	0.075*
Out of labor force & student	0.110***
Unemployed	0.079*
No. split households	4,682

Dependent variable: probability of a split household attriting (not being found) given that original household was found, and the new household split from the original household.
Asterisks indicate significance of marginal effects: *p < 0.05; **p < 0.01; ***p < 0.001.
Marginal effects are estimated at the omitted category for all variables.

interactions between the age of the head of the split household and his/her sex, as well as between his/her marital status and his/her sex. The reference case is constructed from the omitted categories of the 2006 variables. The "head" of the reference case split household was in Greater Cairo in 2006, was male, was age 35–64, was single, was illiterate or read & write, and was in Government work. The base probability of type II attrition for this reference case was 0.708.

The multivariate analysis shows that households that split from a Greater Cairo household are the most likely to attrite. Compared to Greater Cairo, split households originally from Alexandria & Suez Canal are 24.5 percentage points less likely to attrite, households originally from urban Lower Egypt 22.7 percentage points less likely to attrite, households originally from urban Upper Egypt 23.1 percentage points less likely to attrite, households originally from rural Lower Egypt 28.8 percentage points less likely to attrite, and households originally from rural Upper Egypt 26.1 percentage points less likely to attrite.

If the split household is headed by a female who is otherwise observationally identical to the reference case male, it is 25.1 percentage points less likely to attrite. Compared to individuals who were 35 or older in 2006, individuals 15–24 year-old are 26.0 percentage points less likely to attrite, and individuals 25–34 year olds 31.4 percentage points less likely to attrite. Looking at the interaction term between age and gender, compared to male heads of the same age and adding together the female, age, and female/age interaction effects, females ages 15–24 are 37.2 percentage points less likely to attrite than the reference males 35–64, and female 25–34 year olds are 40.6 percentage points less likely to attrite.

Compared to single heads, married heads are 8.7 percentage points more likely to attrite, but divorced heads 43.9 percentage points less likely to attrite. Looking at the interactions between gender and marital status, being female and married has a net effect of decreasing the probability of attrition by 4.7 percentage points compared to the reference single male. Females who were divorced in 2006 have a 45.3 percentage point lower probability of attrition than single males. Females who are separated are 12.1 percentage points more likely to attrite than single females, but the net effect of being female and separated as compared to male and single is a 13 percentage point decrease in the probability of attrition.

Split household heads that had a primary/preparatory education are 5.1 percentage points less likely to attrite than illiterate household heads. Compared to a reference case of a government employee, split household heads that were out of the labor force and not a student in 2006 are 7.5 percentage points more likely to attrite, while those out of the labor force who were students in 2006 are 11.0 percentage points more likely to attrite, and those who were unemployed are 7.9 percentage points more likely to attrite.

2.4. The computation of panel and cross-sectional weights for the ELMPS 2012

To correct for possible biases that could result from the two attrition processes, we constructed weights based on the inverse of the probability of non-response to correct for attrition in both the panel and cross sectional 2012 data. Since we have two new panel data sets, one that follows individuals from 1998 to 2006 to 2012 and one that follows individuals from 2006 to 2012, we constructed two sets of panel weights[12]. The weights for

the cross-sectional 2012 sample were calculated separately for two subsamples and then combined. These subsamples are the panel sample, which includes original and split households, and the refresher sample. These weights were designed to make the sample nationally representative in 2012. This section describes in detail the construction of the weights for both panel samples and the ELMPS 2012 cross-sectional sample.

Weights were constructed at the level of the 2012 household, and applied to all 2012 household members. Constructing the household weights for the ELMPS 2012 panel and cross sectional samples involved the following steps.

1. We determined the probability of Type I attrition for the original (2006) household h and, if applicable, the probability of Type II attrition, the loss of a split household s. The combined probability of attrition for a split household s that split from an original household h is $\Pr(A_{hs})$, where:

$$\begin{aligned} \Pr(A_{hs}) &= \Pr(h\,found\,\&\,s\,not\,found) + \Pr(h\,not\,found) \\ &= \Pr(s\,not\,found|h\,found)[1{-}\Pr(h\,not\,found)] + \Pr(h\,not\,found) \end{aligned}$$

   where $\Pr(h\,not\,found)$ is the probability predicted from the Type I attrition probit model above and $Pr(s\,not\,found|h\,found)$ is the probability predicted from the Type II attrition probit model above. For an original household h, Type II attrition is not relevant and the probability of attrition reduces to $\Pr(A_h) = \Pr(h\,not\,found)$.

2. We computed the response adjustment factors for the households in the original or split samples as, $r_{hs}(2012) = {}^1/_{(1-Pr(A_h))}$ and $r_{hs}(2012) = {}^1/_{(1-Pr(A_{hs}))}$, respectively. The weights of the cross-sectional 2012 sample need to be calculated separately for the original subsample, the split subsample and the refresher sub-sample. The weight of a household h in stratum n in the original sub-sample are calculated as follows: $wo_{hn} = r_{hs}(2012)/\bar{r}_n$, where $\bar{r}_n$ is the mean adjustment factor in each region n in the original/split sub-sample in 2012. There are 6 regions making up the six strata of the sample[13]. Weights for split households are similar: $ws_{hsn} = r_{hs}(2012)/\bar{r}_n/c_s$, where c_s is the number of component households in a split household s. In order to remain representative of the population and avoid double counting while following split households, it is necessary to account for the number of different component households that went into the new household[14].

3. The refresher sample of 2,000 households was designed to increase the representation of households with international migrants. As such, PSUs from the CAPMAS master sample were first stratified into the six regions mentioned above, with the PSUs for each region further stratified into high and low migration areas. This high/low migration classification was based on a ratio of migrants abroad as a percent of the adult population, calculated from the 2006 population census at the neighborhood/village level[15]. We used a different cutoff rate to define high and low migration villages/neighborhoods in urban and rural areas. The cutoff rate for rural areas was 4% whereas it was 1% in urban areas. We thus have 12 strata in total in the refresher sample.

4. The 2,000 households in the refresher sample were drawn from 200 PSUs, 150 of which were in high migration areas and 50 in low migration areas. Twelve households per PSU were randomly selected from the 1,500 households in each PSU in the Master Sample, with ten serving as the refresher sample and 2 as spare

households in case of non-response. The distribution of the 200 PSUs included in the refresher sample by stratum m is available in Assaad and Krafft (2013a) together with the implied sampling rates F_m from among the total number of PSUs in each stratum included in the CAPMAS Master sample. It is worth noting that the CAPMAS master sample is designed to be self-weighted.

5. The weights for a household k in a PSU l in stratum m in the refresher sample were computed as follows: $w_{klm} = 1/F_m * 1500/HH_l$ where F_m is the sampling rate in each stratum m, HH_l is the number of households sampled in PSU l, and 1500 is the total number of households in each PSU in the master sample.

6. The refresher household weight is then given by: $wr_i = w_{klm}/\bar{w}$, $\bar{w}$ is the average of w_{klm} in the refresher sample. Thus the refresher weights, wr_i, have mean 1 in the refresher sample.

7. The final cross-sectional weight for a household j in stratum n is given $w_{jn} = wo_{hn}$ if a household is an original household, $w_{jn} = ws_{hsn}$ if a household is a split household, or $w_{jn} = wr_i$ if the household is a refresher household.

8. To convert the weights into expansion factors, we normalized the weights to one in each region at the individual level, and multiply the weights in each stratum by the ratio of individuals in the sample to individuals in the population as follows:[16] $e_{jn} = w_{jn}{}^{I_n}/{}_{i_n}$, where I_n is the number of individuals in the population for region n, and i_n is the number of individuals in the sample for region n[17].

9. To calculate the weights for the panel data sets, we used revised versions[18] of the appropriate ELMPS 2006 weights to calculate the new panel weights after adjusting them by the response adjustment factors $r_{hs}(2012)$ as follows:

   a. The panel weights for individuals followed from 1998 to 2006 to 2012[19] are calculated as: $pw_{98\_06\_12} = pw_{98\_06} \, r_{hs}(2012) = [w_{98}r_{hs}(2006)]r_{hs}(2012)$, where w_{98} is the cross-sectional weight in 1998 and $r_{hs}(2006)$ is the response adjustment factor due to attrition from 1998 to 2006.

   b. The panel weights for individuals followed from 2006 to 2012[20] are calculated as: $pw_{06\_12} = w_{06}r_{hs}(2012)$, where w_{06} is the cross-sectional weight in 2006.

3. A comparison of selected results from the ELMPS, population censuses, and labor force Surveys[21]

To illustrate the representativeness of the ELMPS, we compare the 1998, 2006, and 2012 ELMPS rounds with ten percent samples of the 1996 and 2006 population censuses[22] and the 2010 & 2011 rounds of the Labor Force Survey (LFS),[23] focusing on demographic and labor market characteristics. The 1996 census 10 percent sample consists of 5,902,243 observations, a random subsample of the 59 million people surveyed by the 1996 census. The 2006 census 10 percent sample consists of 7,282,434 observations, a random subsample of the 73 million people surveyed by the 2006 census. The 2010 LFS data contains 343,913 observations, the 2011 LFS data 379,733, and both are nationally representative after applying the appropriate sample weights.

3.1. Basic demographic and socio-economic characteristics

Figure 1 compares the age distribution of the 1996 census and the 1998 ELMS, using the age of 1998 ELMS individuals in 1996, and only individuals who were at least two

Age, 1996 Census and 1998 ELMPS

Figure 1 Age distribution of 1996 Census and 1998 ELMS (1996 Ages).
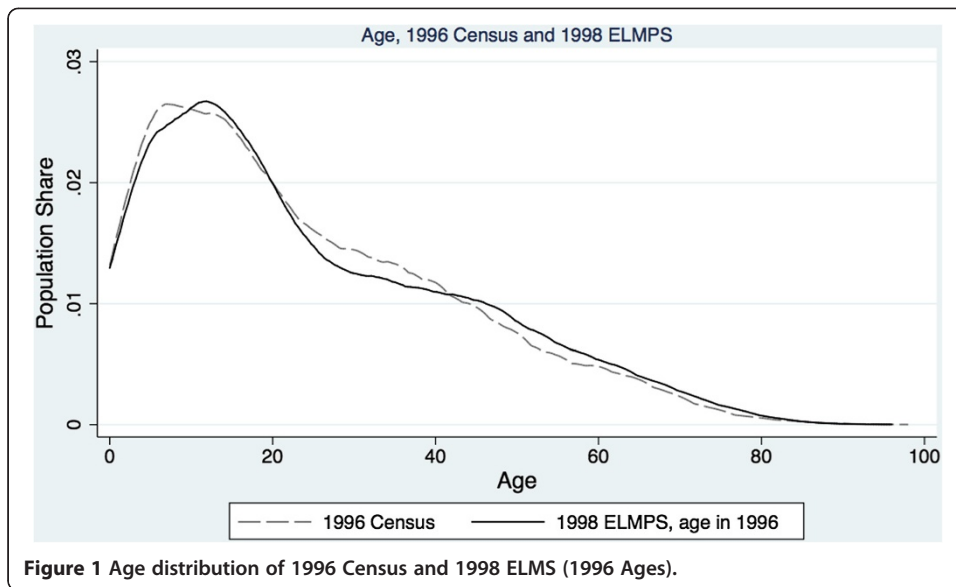
years of age and therefore would have been alive in 1996. The age distributions are generally quite similar, with a few small differences. The 1996 census finds a slightly higher proportion of young children than the 1998 ELMS, while the 1998 ELMS samples more children who would have been in their early teens in 1996 than the 1996 census. The 1996 census samples more individuals in the 20–39 age range, while the 1998 ELMS sampled more individuals who were 40 or older in 1996.

Figure 2 compares the age distribution of the 2006 census and the 2006 ELMPS. The age distributions are generally quite similar, with a few small differences. Both surveys show the 'youth bulge' around age 22 in 2006, the 'trough' of fewer youth in their teens, and the start of the 'echo', the young children of the youth bulge. The 'youth bulge' created substantial pressures on the labor market in 2006. As of 2012, the peak of the youth bulge was 28, and largely integrated into the labor force. The leading edge of the

Age, 2006 Census and 2006 ELMPS

Figure 2 Age distribution of 2006 Census and 2006 ELMPS.
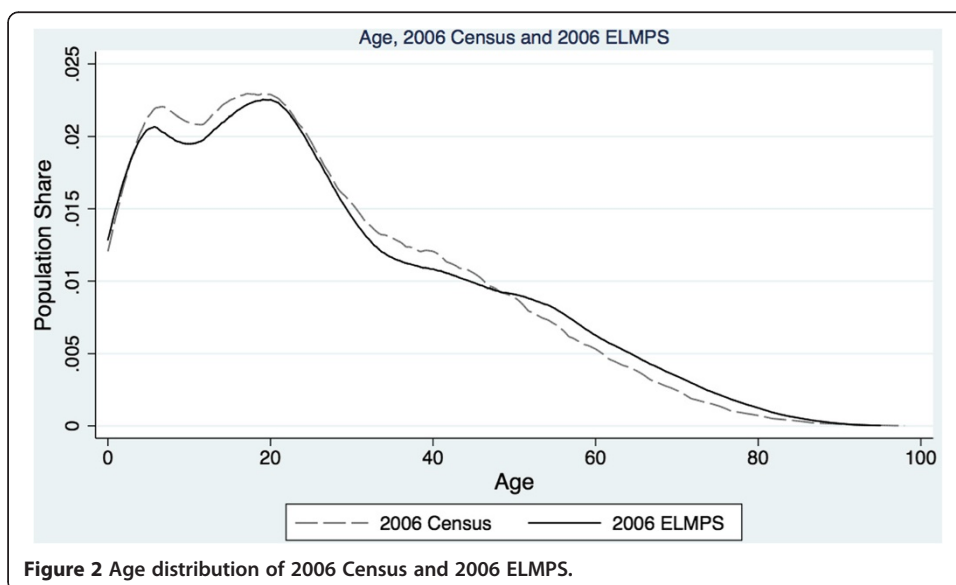
echo was around age 11 in 2012, and as this group ages, it will once again create substantial pressures on the labor market. The 2006 census finds a slightly higher proportion of young children up through age 20 than the 2006 ELMPS. The 2006 census also finds a higher share of individuals 25–45 while the 2006 ELMPS samples more older adults than the 2006 census.

Figure 3 compares the age distribution of the 2006 census and the 2012 ELMPS, using the age of 2012 ELMPS individuals in 2006, and only individuals who were at least six years of age and therefore would have been alive in 2006. The age distributions are generally quite similar, and in fact closer than earlier ELMPS surveys, which may be a consequence of increasing sample size. There are a few very small differences. The 2006 census finds a slightly lower proportion of young children than the 2012 ELMPS, while the 2012 ELMPS samples fewer youth who would have been 5–25 in the 2006 census. The 2006 census samples more individuals in the 35–45 age range, while the 2012 ELMPS sampled slightly more individuals who were 45–60 in 2006.

Figure 4 compares the age distribution of the 2010 LFS sample (all quarterly rounds) and the 2012 ELMPS, using the age of 2012 ELMPS individuals in 2010, and only individuals who were at least two years of age and therefore would have been alive in 2010. The age distributions are quite different for the youth population. The 2012 ELMPS samples slightly more children under age 5 and has both the 'youth bulge', peaking in their late 20s in terms of 2010 ages, and the growing 'echo' of young children. The LFS sample does not capture the youth bulge and echo phenomena, which can be seen starting to occur in the 2006 census (Figure 3) as well as the ELMPS sample. The LFS represents many more 10–20 year olds than the 2012 ELMPS, fewer individuals in their mid-20s to mid-30s, and a larger population in their late 30s to mid 50s. This suggests possible inaccuracies in age reporting or in household composition in the LFS.

Table 8 presents the distribution of education by survey and round for individuals age 10 and older. The different surveys use somewhat different questions to identify education, which may account for some of the differences in education distributions. The 1998 ELMS finds a smaller proportion of illiterates (30%) than the 1996 census (39%) and more
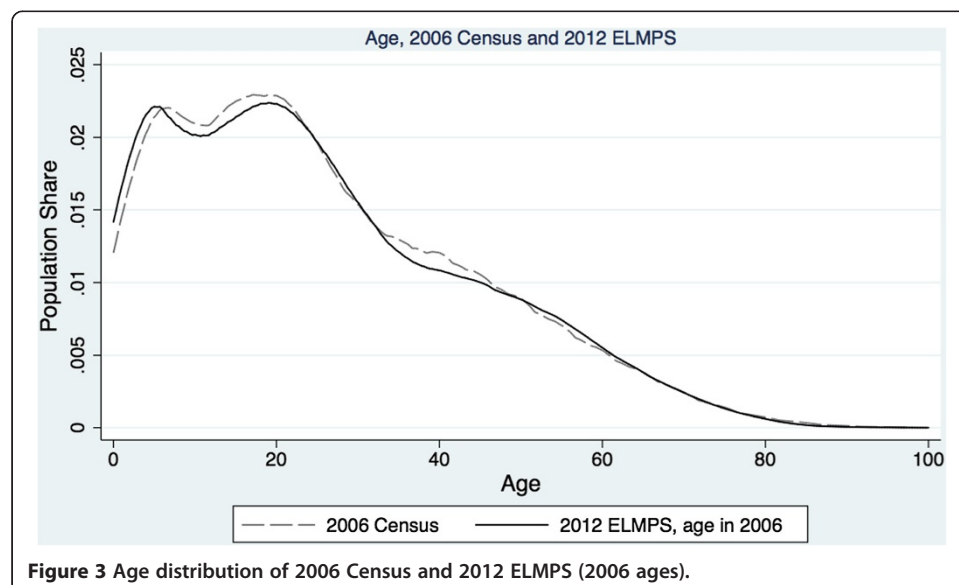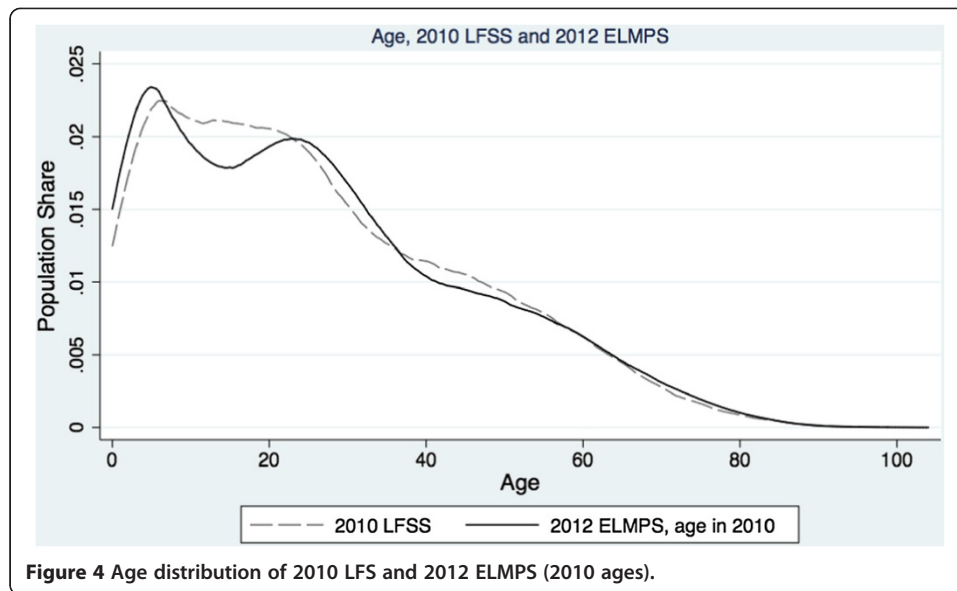

Figure 3 Age distribution of 2006 Census and 2012 ELMPS (2006 ages).

Figure 4 Age distribution of 2010 LFS and 2012 ELMPS (2010 ages).

individuals with less than intermediate education (26%) as compared to the 1996 census (18%), which is consistent with improving educational attainment. The proportion of the population with intermediate (equivalent to upper secondary), above intermediate (equivalent to 2-year colleges), and university and higher education is very close in the 1996 census and 1998 ELMS. The 2006 census and 2006 ELMPS show very similar distributions of education. Both find 30% of the population is illiterate[24]. The 2006 census classifies 13% of the population as "reads and writes" and 19% as having less than intermediate education (equivalent to primary and lower secondary), while the 2006 ELMPS classifies 10% of the population as "reads and writes" and 23% as having less than intermediate education. Very similar proportions are observed in intermediate through university education when comparing the 2006 census and 2006 ELMPS. Comparing the 2010 LFS and 2012 ELMPS, they represent individuals with relatively similar education. While 26% of the 2010 LFS sample was illiterate and 15% could read and write, only 23% of the 2012 ELMPS sample was illiterate and 19% could read and write. About 20% of individuals were educated at less than the intermediate level in both samples. In the 2010 LFS 25% of individuals had an intermediate education, compared to 23% in the 2012 ELMPS, while the percentage with an above intermediate or university and above education is very similar.

Table 8 Education distribution by survey and round, age 10 and older

| Education	1996 census	1998 ELMS	2006 census	2006 ELMPS	2010 LFS	2012 ELMPS
Illiterate	39	30	30	30	26	23
Reads & writes	19	18	13	10	15	19
Less than intermediate	18	26	19	23	20	21
Intermediate	16	16	26	24	25	23
Above intermediate	2	3	3	3	3	3
University & above	6	6	10	10	10	11
Total	100	100	100	100	100	100

Table 9 examines individual's marital status by survey and round. Between 2006 and 2012 the law changed such that the minimum age of marriage for women changed from 16 to 18. The minimum age of marriage for men was already 18. For comparability, the table classifies individuals as 'under age' if they were under age 18, regardless of the law at the time of the particular survey. Almost half (48%) of individuals in the 1996 census were under age, and 11% were never married, which is similar to the 45% under age and 14% never married sampled two years later in the 1998 ELMS. The 2006 census and 2006 ELMPS find very similar distributions of marital status, around 40% under age, 15% never married, 39% married, 1% divorced, and 5% widowed. The 2010 LFS and 2012 ELMPS also produce relatively similar distributions in terms of marital status. For both samples, 41% of individuals are under age. In the 2010 LFS, 14% of the sample was never married. In the 2012 ELMPS 11% of the individuals were never married. The 2010 LFS and 2012 ELMPS have similar percentages of individuals married, divorced, and widowed, within a percentage point.

All the surveys define a household as individuals who live under the same roof and eat from the same pot, but may operationalize this definition differently in the field. This is a common definition for identifying households. However, a common definition does not preclude challenges in operationalizing a definition in the field, and difficulties identifying household members (Beaman and Dillon 2012; Glewwe 2000). In Egypt, households are increasingly difficult to identify because of changing rural shelter patterns. Especially in rural areas and poorer urban areas, it can be difficult to clearly distinguish households when extended families live in the same building. Increasingly, these households live in multiple story buildings containing separate apartments with their own kitchens. These families may be sharing some but not all meals and may be pooling some of their consumption budgets but not all. Identifying whether they are a single household or multiple households in the field is therefore a question of judgment that can vary from one survey to another and even from one enumerator to the other.

As a result, there are some small differences in household size by survey and round (Figure 5)[25]. Different definitions of households used in the different surveys and rounds yield different distributions of household sizes. Table 10, which presents average household sizes in each survey and round, also shows variation in the distribution of household sizes. Both use the percent of individuals living in a given household size, not the percent of households of a given size. The 1998 ELMS sampled the largest average household sizes, 6.4, and the greatest percent of households with 8 or more people. However, the 1996 census actually sampled the greatest percent of 6 or 7 person households. The 2006 ELMPS represented individuals living in an average household size of 5.8, while the 2006 census found that individuals lived in households with an average of 5.1 people. The 2010

Table 9 Marital status by survey and round

Marital status	1996 census	1998 ELMS	2006 census	2006 ELMPS	2010 LFS	2012 ELMPS
Under age	48	45	41	40	41	41
Never married	11	14	14	15	14	11
Married	37	36	40	39	41	42
Divorced	0	1	1	1	1	1
Widowed	4	5	4	5	4	5
Total	100	100	100	100	100	100

Figure 5 Household size by survey and round.
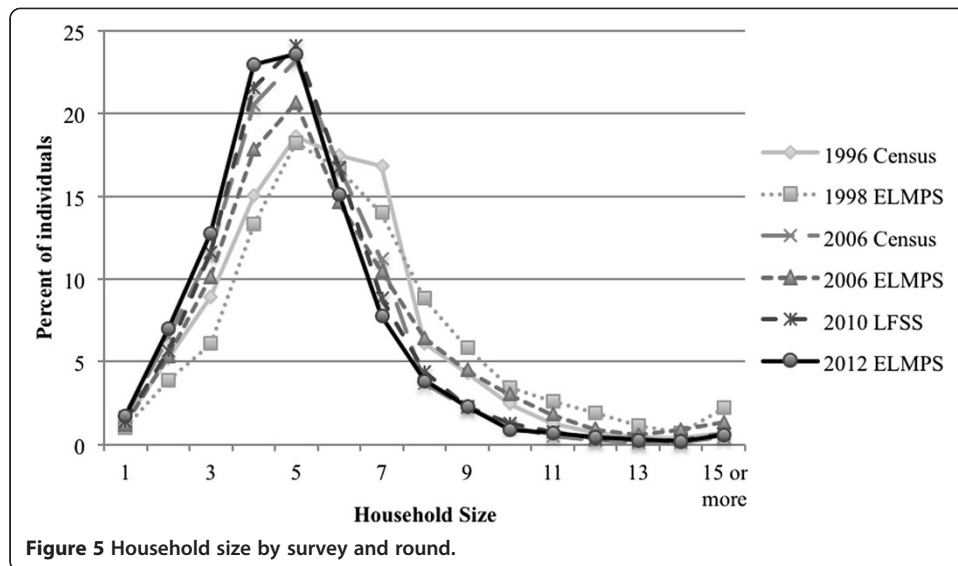
LFS sampled individuals living with 5.2 individuals on average, while the 2012 ELMPS sampled individuals living with 5.0 individuals on average. The 2012 ELMPS and 2010 LFS generally found quite similar distributions of household size (Figure 5) as well as having similar averages.

3.2. Trends in unemployment

The LFS over the 2010–2011 period shows a sharp spike in unemployment at the time of the revolution, in the first quarter of 2011 (Table 11). Unemployment, which had been around 8.8% for 2010, increased to 11.8% in the first quarter of 2011, and had risen to 12.4% by the last quarter of 2011. Yet the ELMPS finds an 8.7% unemployment rate overall. The surveys are actually quite consistent in terms of the female unemployment rate, which is 23.2% in LFS 2011 Q4 and 23.7% in the ELMPS 2012. The picture of male unemployment is, however, quite different. Male unemployment is the component that increased in LFS 2011 Q1, almost doubling from 4.8% in LFS 2010 Q4 to 8.9% in LFS 2011 Q1, while female unemployment actually decreased slightly. In contrast, the ELMPS finds a male unemployment rate of 4.2% in 2012.

3.2.1. Why are male unemployment rates different in LFS and ELMPS?

We investigate why the LFS and ELMPS generate very different (male) unemployment rates for subsequent years; while the standard market unemployment rate in the LFS in

Table 10 Average household size by survey and round

Survey	Average household size
1996 census	5.7
1998 ELMS	6.4
2006 census	5.1
2006 ELMPS	5.8
2010 LFS	5.2
2012 ELMPS	5.0

Table 11 Unemployment rates by sex and survey, standard market labor force, ages 15-64

	Male	Female	Total
LFS			
2010 Q1	4.9	20.8	8.7
2010 Q2	4.9	22.0	8.9
2010 Q3	4.6	23.0	8.8
2010 Q4	4.8	22.4	8.8
2011 Q1	8.9	21.5	11.8
2011 Q2	8.7	22.3	11.8
2011 Q3	8.7	22.7	11.9
2011 Q4	9.2	23.2	12.4
ELMPS			
2012	4.2	23.7	8.7

the fourth quarter of 2011 is 12.4%, the ELMPS, in the first quarter of 2012, finds an unemployment rate of 8.7%. We show these differences are due primarily to the more detailed information collected in the ELMPS and the fact that the ELMPS collects information from the individual him or herself, while the LFS gathers the information from any adult members who happens to be present in the household during the first visit by the enumerator[26]. The LFS is likely classifying many visibly under-employed individuals (those with precarious employment and who may not be able to find employment every day of the week) as unemployed, while the ELMPS correctly identifies them as employed, but visibly under-employed. While the ELMPS unemployment rate may seem low, it is not unusual for a low-income country with no unemployment benefits to maintain a low unemployment rate. People cannot afford to be unemployed, but do tend to be under-employed.

Differences in unemployment rates are not due to differences in the labor force; labor force participation rates are relatively comparable between LFS 2011 Q4 and ELMPS 2012 as Table 12 demonstrates. In LFS 2011 Q4, the total participation rate was 51.3%. Male

Table 12 Labor force participation by sex and survey, standard market labor force, ages 15-64

	Male	Female	Total
LFS			
2010 Q1	77.6	24.6	51.5
2010 Q2	78.3	24.7	51.9
2010 Q3	78.8	23.8	51.8
2010 Q4	78.8	23.8	51.6
2011 Q1	77.4	23.6	51.0
2011 Q2	77.9	23.7	51.2
2011 Q3	78.5	23.6	51.5
2011 Q4	78.2	23.7	51.3
ELMPS			
2012	80.2	23.1	51.1

labor force participation was 78.2% and female participation was 23.7%. In the ELMPS the total participation rate was 51.1%. The female rate was slightly lower, 23.1%, and the male rate slightly higher, 80.2%. Overall the participation patterns are quite consistent. This suggests that some of the males detected as unemployed in the LFS would have actually been detected as employed in the ELMPS rather than out of the labor force.

Examining visible under-employment trends shows a likely explanation for the discrepancy in male unemployment rates between the LFS and the ELMPS 2012. An individual is considered to be visibly ("time-related") under-employed if he or she worked less than 35 hours in the reference week, and the reason for working less than 35 hours is the unavailability of work opportunities the rest of the time. Underemployment is a common problem in developing countries, but accurately identifying underemployment is complex (Hussmanns 2007; Wilkins and Wooden 2011). While the ELMPS 2012 in fact asks about the reasons for working less than 40 hours, the LFS only asks about working less than 35 hours, so for the sake of comparability we apply the lower threshold to the ELMPS for this paper; however, it should be noted that visible under-employment would be even higher if a 40 hour standard were used.

Table 13 presents visible under-employment rates as a share of the standard market labor force. This data is only available for LFS in 2011. If individuals are identified as unemployed rather than employed but visibly under-employed, the labor force will be the same size but the number of unemployed will be higher. Thus differences in the visible under-employment rate as a share of the labor force will map onto differences in the (un)employment rate. The total visible under-employment rate in the labor force in the LFS 2011 Q4 is 1.8%, while the ELMPS finds a 6.2% visible under-employment rate. This 4.4 percentage point higher visible under-employment rate in the ELMPS is quite similar to the 3.7 percentage point lower unemployment rate. The gender differences are a little larger; the male visible under-employment rate is 2.2% in the LFS 2011 Q4 and 6.6% in the ELMPS 2012, a 4.4 percentage point difference that accounts for most of the 5.0 percentage point difference in the male unemployment rate. The female visible under-employment rate was 0.4% in the LFS 2011 Q4 and 5.1% in the ELMPS, a 4.7 percentage point difference. The total combined unemployment and underemployment rate is 14.9% for the ELMPS 2012 and 14.2% for LFS 2011 Q4, a 0.7 percentage point higher combined rate in the ELMPS than LFS 2011 Q4.

How the LFS is fielded drives these differences. While the ELMPS collects labor market data directly from individuals, and will follow up with a household up to three

Table 13 Visible under-employment rates as share of standard market labor force

	Male	Female	Total
LFS			
2011 Q1	2.5	1.0	2.1
2011 Q2	1.7	0.5	1.4
2011 Q3	1.6	0.5	1.4
2011 Q4	2.2	0.4	1.8
ELMPS			
2012	6.6	5.1	6.2

times to find an individual, the LFS collects data on the labor market characteristics of all individuals in the household from a single respondent, whoever is home at the time of the survey. A single respondent reporting labor information for the entire household is not as likely to accurately identify whether or not an individual is working, or if that individual has intermittent work and is visibly under-employed.

These differences can also be observed in terms of employment (Table 14). The 2011 Q4 LFS finds a male employment rate of 71.0% and female rate of 18.2%. The ELMPS 2012 finds a slightly lower female employment rate, 17.6%, and a much higher male employment rate, 76.8%. This difference in employment rates is driven by the same share of the population that is classified as unemployed by the LFS and underemployed by the ELMPS.

Despite these differences in classifying individuals into employment, under-employment, and un-employment, the ELMPS and LFS find very similar results for other labor market phenomena, such as wages. As Table 15 shows, the nominal monthly wages for regular workers in the LFS 2011 Q4 and ELMPS 2012 are nearly identical, 1225 LE per month. Wages actually rose substantially over 2011 for regular workers, especially between 2011 Q1 and 2011 Q2, in large part due to an increase in wages for public employees post-revolution. For irregular workers, who receive wages on a daily basis, the ELMPS 2012 finds slightly lower nominal daily wages (40.6 LE) than the LFS, which found daily wages of 43.3 LE in 2011 Q4 and daily wages of 41.9 LE in 2011 Q3 for irregular workers. The slightly lower daily wage rate for irregular workers in the ELMPS 2012 may be due to the detection of additional irregular and under-employed workers in the ELMPS. Furthermore, irregular wages tend to fluctuate a lot with economic conditions and, since the LFS does not usually collect the wage from the worker him or herself, it may do a poorer job of capturing these fluctuations.

We also examined differences in wages disaggregated by sex and type of employment across LFS 2011 Q4 and ELMPS 2012. These comparisons are shown in Table 16. For both sexes taken together, differences in wages across the two surveys are less than 10% except for the public enterprise sector, where ELMPS reports much higher wages than the LFS. It is worth noting that because this is the smallest of the sectors presented here, its average wage rates have fluctuated considerably even within the LFS, where they were LE 1,529/month in Q2 as compared to LE 1,366/month in Q4. Discrepancies in average female wages across the two surveys are larger (up to 25%),

Table 14 Employment rates, market definition, LFS and ELMPS, ages 15-64

	Male	Female	Total
LFS			
2010 Q1	73.7	19.5	47.0
2010 Q2	74.5	19.3	47.3
2010 Q3	75.2	18.3	47.2
2010 Q4	75.0	18.5	47.0
2011 Q1	70.5	18.5	45.0
2011 Q2	71.1	18.4	45.1
2011 Q3	71.7	18.3	45.4
2011 Q4	71.0	18.2	44.9
ELMPS			
2012	76.8	17.6	46.7

Table 15 Nominal monthly wages (Regular workers) and daily wages (Irregular workers), employed (Market definition) wage workers, LFS and ELMPS, ages 15-64

	Monthly wage, regular workers	Daily wage, irregular workers
LFS		
2011 Q1	930	37.9
2011 Q2	1,178	42.4
2011 Q3	1,236	41.9
2011 Q4	1,225	43.3
ELMPS		
2012	1,225	40.6

but this is probably due to the smaller sample sizes in these categories and the fact that LFS collects data from any member of the household rather than the worker herself.

4. Conclusions

High-quality, publicly available and up-to-date data is key to studying labor markets. The need for high quality micro data on the Egyptian labor market cannot be exaggerated. Since the implementation of the ELMS 1998 and ELMPS 2006, there has been no reliable detailed information on labor market conditions in Egypt. The regular Labor Force Survey carried out quarterly by CAPMAS produces very limited information on current labor market conditions, but suffers from the fact that it is not collected from the worker him or herself and from lack of detailed information about conditions of employment and employment dynamics.

Over the past years, the publicly available data of the ELMS 1998 and ELMPS 2006 have proved to be the workhorse of labor market research in Egypt (see, for instance: Angel-

Table 16 Nominal monthly wages (Regular workers) and daily wages (Irregular workers) by sex and type of employment across LFS 2011 Q4 and ELMPS 2012, ages 15-64

	LFS 2011 Q4	ELMPS 2012
Male		
Government (monthly)	1,306	1,288
Public enterprise (monthly)	1,380	1,681
Private regular (monthly)	1,124	1,238
Private irregular (daily)	43.5	40.8
Female		
Government (monthly)	1,302	1,048
Public enterprise (monthly)	1,236	1,441
Private regular (monthly)	1,041	778
Private irregular (daily)	35.1	32.8
Both		
Government (monthly)	1,305	1,201
Public enterprise (monthly)	1,366	1,650
Private regular (monthly)	1,114	1,189
Private irregular (daily)	43.3	40.6

Urdinola and Semlali 2010; Belhaj Hassine 2011; Biltagy 2012; Langsten and Salem 2008; Namoro and Roushdy 2009; Said 2012; Wahba and Zenou 2005). However, economic and political conditions have changed substantially since 2006. Fortunately, the ELMPS 2012 is now available, with rich data covering a wide variety of different labor market topics.

The breadth and depth of the survey, as well as its longitudinal nature, will allow the data to be used in a variety of different ways and to answer a multitude of research questions. The addition of the 2012 round will be particularly important for understanding the impact of the recent global financial crisis, as well as the economic impacts of the January 25th revolution. The longitudinal and retrospective data is ideal for examining long-term labor market trajectories and labor market dynamics. While the focus of the survey is on the labor market, the life history added in 2012 and modules such as health, education, geographic mobility, and women's status can provide a rich source of data for a wide variety of other research questions. The data from the ELMPS can also be linked to other sources, such as school-level data, to further enrich research. Although the ELMPS 2012, with its rich longitudinal data, is unique to the MENA region, the Jordan Labor Market Panel Survey (JLMPS), which fielded its first round in early 2010, is based on the same design, allowing for comparative work. The data from JLMPS 2010 can also be obtained from the ERF data Portal (www.erfdataportal.com). The first wave of the Tunisia Labor Market Panel Survey is planned for the spring of 2014 and data from it will be made available through the ERF Data Portal. We hope to see additional comparable surveys in the region and globally in the future.

As this paper has demonstrated in examining unemployment rates, the ELMPS's more accurate and detailed information generates a substantially different picture of the challenges facing the labor market than the rough snapshot provide by the Labor Force Survey. Accurately identifying and understanding these challenges requires high-quality data that only a survey such as the ELMPS can provide. Our case study is just one example of the many ways the ELMPS data will be a vital part of accurately understanding Egypt's labor market in this time of transition.

Endnotes

[1] See Assaad and Barsoum (2000), Assaad (2002), Barsoum (2009), and Assaad (2009) for additional information on the 1998 and 2006 rounds of the survey.

[2] Although the sample was designed to be nationally representative, this round and all subsequent rounds do not cover the five Border Governorates of Matruh, New Valley, Red Sea, North and South Sinai. These remote governorates are sparsely populated, containing no more that 2% of the population in 2006 (Minnesota Population Center 2011).

[3] The CAPMAS master sample used to draw the ELMS 1998 sample was originally prepared in 1995 and subsequently updated. It is subdivided into 500 Enumeration Areas (EAs) each containing about 1,500 households. The sample is stratified into urban and rural strata, with each of these strata further stratified into substrata by governorate. The first stage consisted of selecting the village and *shyakhas* (urban neighborhoods) to be represented in the sample, which was done by the principle of probability proportional to size, meaning that some shyakhas or villages could appear more than once if their size warrants it. The selected villages and *shyakhas* were then subdivided into EAs of approximately 1500 dwelling units each and one or more EA was selected from each. The final master sample contained 306 urban EAs and 194 rural EAs. See Assaad and Barsoum (2000).

[4] Areas with high migration rates were identified by classifying all PSUs in the CAPMAS master sample into high or low migration areas based on a migration variable from the 2006 population census. Rural areas with at least a 4% migration rate among adults were classified as high migration, and urban areas with at least 1% migration rate among adults were classified as high migration. Out of 200 PSUs in the refresher sample, 50 were from high migration urban areas, 100 from high migration rural areas, 22 from low migration urban areas and 28 from low migration rural areas.

[5] This section is based on Assaad et al. (2013).

[6] For an analysis of sample attrition from 1998 to 2006, please refer to Assaad and Roushdy (2009).

[7] See Assaad and Barsoum (2000), Barsoum (2009) and Assaad and Roushdy (2009) for descriptions of the sample designs in 1998 and 2006.

[8] See section on Type II attrition for further details about the information collected about the status of individuals at the enumeration stage.

[9] We treat those who have relocated to parts of Egypt that are not covered by the survey (i.e. the border governorates of Matrouh, Red Sea, New Valley, North and South Sinai) similarly to those who left the country altogether. Hereafter, they will be referred to as "having gone abroad" or having "left the country" even though this is not strictly true.

[10] These split households will be assigned the average attrition probability estimated in the model.

[11] This individual may not in fact be the actual head of the split household, but we are considering him or her to be the head of the group that split off together from the original household.

[12] There are also panel weights for individuals followed from 1998 to 2006. See Assaad and Roushdy (2009) for information on these weights.

[13] See Assaad and Krafft (2013a) for how regions were defined.

[14] To identify component households empirically, we counted the number of split household members who were alive during the previous wave (were not individuals born subsequently), and the number of those who were not in the 2006 sample. If there were no individuals in the split household who were alive in the previous wave and not in the 2006 sample, then there was only one component household. If there was one or more than one individual in the split household who was alive in the previous wave and not in the 2006 sample, then we considered the split household to consist of two component households. Although it is theoretically possible that there are more than one or two component households forming a new household, we did not have data on whether additional members had previously lived together, and an examination of the relationships among additional members indicated that they almost always had a prior relationship, for instance a wife and older daughter who had been absent the previous wave. We therefore considered the maximum possible number of component households to be two.

[15] Every household was asked in the 2006 census whether they had members abroad and asked to provide the number and distribution by country of these members. Although this method clearly understates the number of Egyptian migrants abroad because it does not account for households that left in their entirety, it can be used to create an acceptable indicator of the geographic distribution of emigration.

[16] The source of the number of individuals in the population in each region was an official population projection by region (as defined in this survey) as of 1/1/2012.

[17] These expansion factors are named expan_12 in the 2012 cross-sectional data.

[18] We revised the 2006 weights in two ways. First, although in 2006 the Type II attrition process was random, and therefore there was no need to model the attrition process to remain representative of the population, in 2006 average Type II attrition (15.4%) was not incorporated into the weights of split household. Average 2006 Type II attrition was incorporated into the revised weights. The second revision was accounting for the occurrence of multiple component households in split households, to prevent double-counting; see the discussion in the cross-sectional weights and footnote there for further information.

[19] These panel weights are named panel_wt_98_06_12 in the panel data.

[20] These panel weights are named panel_wt_06_12 in the panel data. Additionally, panel weights for individuals observed in 1998 and 2006 but not 2012 are calculated in a similar manner (Assaad and Roushdy 2009) and included in the dataset named panel_wt_98_06.

[21] This section is based on Assaad and Krafft (2013b).

[22] We use 10 percent samples for the 1996 and 2006 Censuses from the IPUMS-International database of the Minnesota Population Center (2011).

[23] The LFS is conducted quarterly and is a product of CAPMAS (2011; CAPMAS 2010).

[24] Whether an individual is reported as illiterate or "reads and writes" depends on their own assessment of their literacy and can therefore vary from one survey to another depending on how the question is asked.

[25] As a result of the operational differences in identifying households, there are also differences in the distribution of individuals' relationships with the head of household, see Assaad and Krafft (2013a).

[26] ELMPS enumerators are instructed to make at least three visits at different times of day to make sure that they can interview the individual in person and only after failing to do so after three visits can they collect the information from a proxy respondent. In the LFS, any adult respondent present (usually during the day) can respond on behalf of all individuals in the household.

Author details
[1]Humphrey School of Public Affairs, University of Minnesota, 301 19th Ave. S, Minneapolis, MN 55455, USA.
[2]Department of Applied Economics, University of Minnesota, 1994 Buford Avenue - Room 217H, St. Paul, MN 55108-6040, USA.

References

Angel-Urdinola D, Semlali A (2010) Labor markets and school-to-work transition in Egypt: diagnostics, constraints, and policy framework. MPRA paper 27674. University Library of Munich, Munich, Germany

Assaad R (2002) The Egyptian Labor Market in an Era of Reform. American University in Cairo Press, Cairo, Egypt

Assaad R (2009) The Egyptian Labor Market Revisited. American University in Cairo Press, Cairo, Egypt

Assaad R (2013) Making sense of Arab labor markets: the enduring legacy of dualism. IZA discussion paper No. 7573. Institute for the Study of Labor, Bonn, Germany

Assaad R, Barsoum G (2000) Egypt Labor Market Survey, 1998: Report on the Data Collection and Preparation. http://www.erf.org.eg/CMS/uploads/pdf/1194970697_ELMS_98_Data_Report.pdf. Accessed 29 Apr 2013

Assaad R, Krafft C (2013a) The Egypt labor market panel survey: introducing the 2012 round. Economic research forum working paper series No. 758. The Economic Research Forum, Cairo, Egypt

Assaad R, Krafft C (2013b) Comparisons between the Egyptian Labor Market Panel Survey, Egyptian Population Census, and Egyptian Labor Force Surveys. University of Minnesota, Minneapolis, Minnesota

Assaad R, Krafft C, Yassine C (2013) An analysis of sample attrition and the calculation of sample weights in the Egypt labor market panel survey 2012. Mimeo

Assaad R, Roushdy R (2009) Methodological appendix 3: an analysis of sample attrition in the Egypt labor market panel survey 2006. In: Assaad R (ed) The Egyptian Labor Market Revisited. American University in Cairo Press, Cairo, Egypt, pp 303–316

Barsoum G (2009) Methodological appendix 1: the Egypt labor market panel survey 2006: documentation of the data collection process. In: Assaad R (ed) The Egyptian Labor Market Revisited. American University in Cairo Press, Cairo, Egypt, pp 259–284

Beaman L, Dillon A (2012) Do household definitions matter in survey design? Results from a randomized survey experiment in Mali. J Dev Econ 98:124–135

Belhaj Hassine N (2011) Inequality of Opportunity in Egypt. World Bank Econ Rev 26:265–295

Biltagy M (2012) Quality of education, earnings and demand function for schooling in Egypt: an economic analysis. Procedia - Social and Behavioral Sciences 69:1741–1750

CAPMAS (2010) Labor Force Survey 2010. Central Agency for Public Mobilization and Statistics, Cairo, Egypt

CAPMAS (2011) Labor Force Survey 2011. Central Agency for Public Mobilization and Statistics, Cairo, Egypt

Glewwe P (2000) Household Roster. In: Grosh M, Glewwe P (eds) Designing household survey questionnaires for developing countries: Lessons from 15 years of the Living Standards Measurement Survey. The World Bank, Washington, DC, pp 135–142

Hussmanns R (2007) Measurement of employment, unemployment and underemployment – Current international standards and issues in their application. In: Bulletin of Labor Statistics 2007–1., http://www.ilo.org/public/english/bureau/stat/download/articles/2007-1.pdf. Accessed 7 May 2013

Langsten R, Salem R (2008) Two Approaches to Measuring Women's Work in Developing Countries : A Comparison of Survey Data from Egypt. Popul Dev Rev 34:283–305

Minnesota Population Center (2011) Integrated Public Use Microdata Series, International: Version 6.1 [Machine Readable Database]. University of Minnesota, Minneapolis

Namoro S, Roushdy R (2009) Intrahousehold resource allocation in Egypt: women empowerment and investment in children. Middle East Dev J 01:105–121

Said M (2012) Trade reform, Job quality and wages of the working poor in Egypt: evidence from manufacturing panel data. J Dev Areas 46:159–183

Wahba J, Zenou Y (2005) Density, social networks and job search methods: Theory and application to Egypt. J Dev Econ 78:443–473

Wilkins R, Wooden M (2011) Economic approaches to studying underemployment. In: Maynard DC, Feldman DC (eds) Underemployment: psychological, economic, and social challenges. Springer New York, New York, NY, pp 13–34