

ECONOMIC  
RESEARCH  
F O R U M



منتدى  
البحوث  
الاقتصادية

2016

# working paper series

**COMPARING RETROSPECTIVE  
AND PANEL DATA COLLECTION METHODS  
TO ASSESS LABOR MARKET DYNAMICS**

**Ragui Assaad, Caroline Krafft and Shaimaa Yassin**

**Working Paper No. 994**

**COMPARING RETROSPECTIVE  
AND PANEL DATA COLLECTION METHODS  
TO ASSESS LABOR MARKET DYNAMICS**

Ragui Assaad, Caroline Krafft and Shaimaa Yassin

**Working Paper 994**

**April 2016**

**Send correspondence to:**  
Shaimaa Yassin  
University of Paris I, Pantheon-Sorbonne  
cyassine.eco@gmail.com

First published in 2016 by  
The Economic Research Forum (ERF)  
21 Al-Sad Al-Aaly Street  
Dokki, Giza  
Egypt  
[www.erf.org.eg](http://www.erf.org.eg)

Copyright © The Economic Research Forum, 2016

All rights reserved. No part of this publication may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without permission in writing from the publisher.

The findings, interpretations and conclusions expressed in this publication are entirely those of the author(s) and should not be attributed to the Economic Research Forum, members of its Board of Trustees, or its donors.

## Abstract

There is potential for measurement problems in both retrospective and panel microdata. In this paper we compare results on basic indicators related to labor markets and their dynamics from retrospective and panel survey data on the same individuals in Egypt, in order to determine the conditions under which results are similar or different. Specifically, we (1) assess the consistency of reporting of time-invariant characteristics in different waves of the panel, (2) compare the retrospective and panel data results on past labor market statuses, (3) assess the consistency of estimates of labor market transition rates across two specific dates by comparing panel and retrospective data, (4) assess the consistency of estimates of the level and trends of annual labor market transition rates across retrospective data from different waves of the survey, and (5) assess whether retrospective data can provide accurate trends of labor market aggregates, such as unemployment rates. We find that it is possible to garner useful information on labor market dynamics from retrospective data, but one must be cautious about which information to trust and at what level of detail. We conclude with a discussion of implications for future research as well as future survey design.

**JEL Classification :** C83, C81, J01, J62, J64

**Keywords:** Panel Data, Retrospective Data, Survey Data, Measurement Error, Labor Markets, Egypt

## ملخص

هناك إمكانية لقياس المشكلات في كل من البيانات الجزئية بأثر رجعي والمسوح. في هذه الورقة نقارن النتائج على المؤشرات الأساسية المتعلقة بأسواق العمل وديناميتها من بيانات المسح بأثر رجعي على نفس الأفراد في مصر، من أجل تحديد الشروط التي تكون بموجبها النتائج متشابهة أو مختلفة. وعلى وجه التحديد، نقوم (1) بتقييم اتساق الإبلاغ عن خصائص الوقت الثابتة في موجات مختلفة من المسح، (2) بمقارنة نتائج البيانات بأثر رجعي وخلق نقاش حول الأوضاع في سوق العمل الماضية، (3) بتقييم اتساق تقديرات سوق العمل ومعدلات الانتقال عبر تواريخ محددة بمقارنة البيانات بأثر رجعي والمسوح، (4) بتقييم اتساق تقديرات مستوى واتجاهات معدلات الانتقال السنوية لسوق العمل عبر البيانات بأثر رجعي من موجات مختلفة من الدراسة، و (5) نقوم بتقييم ما إذا كانت البيانات بأثر رجعي يمكن أن توفر اتجاهات دقيقة من مجاميع سوق العمل، مثل معدلات البطالة. نجد أنه من الممكن جمع معلومات مفيدة حول ديناميات سوق العمل من البيانات بأثر رجعي، ولكن يجب على المرء أن يكون حذرا حول الثقة بأي معلومات وعلى أي مستوى من التفاصيل. ونختتم بمناقشة الآثار المترتبة للأبحاث في المستقبل، فضلا عن تصميم المسوح في المستقبل.

## **1. Introduction**

The analysis of labor market dynamics requires the availability of data about the same individuals at multiple points in time. This kind of data allows for the examination of flows between different labor market states rather than simply assessing labor market stocks over time, which is what is usually possible with cross-sectional data. Data about the same individuals over time can either be in the form of panel data, where individuals are visited and interviewed multiple times over the course of several months or years, or retrospective data, where individuals are asked about their past labor market trajectories at one point in time. Although both methods of data collection suffer from different kinds of measurement errors, panel data are often deemed superior because they minimize recall error, which could be substantial in retrospective data. Panel data, however, are expensive and difficult to collect and are, therefore, rarely available to researchers in developing countries. If available, they are generally not collected frequently enough to observe complete labor market trajectories and transitions. It is therefore useful to examine how well retrospective data perform in assessing labor market dynamics and the extent to which analyses that depend on them conform to results obtained from panel data.

It is well known that retrospective data suffer from problems such as difficulties in recalling dates or even that certain events occurred at all (Artola & Bell, 2001; Bound, Brown, & Mathiowetz, 2001; Magnac & Visser, 1999). Panel data, that is data that are collected contemporaneously at different points in time for the same individual, are unlikely to suffer from recall errors but may have other problems. Because they are collected at discrete points in time, they only provide information at those points in time and not on the course of events between those points (Blossfeld, Golsch, & Rohwer, 2007). Moreover, panel data can suffer from sample attrition and misclassification errors (Artola & Bell, 2001). They can also suffer from the fact that individuals may be unwilling to accurately report their current status due to fear of taxation or other government interference.

Due to potential problems with both retrospective and panel data, it is worthwhile to compare results on basic indicators related to labor market dynamics from retrospective and panel data on the same sample of individuals, in order to determine the conditions under which they provide similar or substantially different results. To date, no study has undertaken such a comparison in the Middle East and North Africa (MENA) region. This paper takes advantage of a unique opportunity to undertake such a comparison, where both panel and retrospective data are available for the same individuals in the Egypt Labor Market Panel Survey (ELMPS). Three waves of the ELMPS were carried out by the Economic Research Forum (ERF) in 1998, 2006 and 2012. All three waves of the ELMPS contain both contemporaneous and retrospective data, including detailed labor market histories for all individuals 15 and older who have ever worked, as well as other life course variables. Not only do the reference periods of the retrospective data overlap with the dates of the previous waves of the survey, allowing for comparisons of retrospective and panel data at the same point in time, but the retrospective periods from different waves of the survey overlap with each other as well, allowing for comparison of past events in one wave with the same events as captured in another wave.

In this paper we assess the soundness of both the distribution of past statuses and transitions among them obtained from the two sources of data. Specifically, we (i) assess the consistency of reporting of time-invariant characteristics in different waves of the panel, (ii) compare the retrospective and panel data results on past labor market statuses, including the estimation of multivariate models of the determinants of alignment between the two data sources, (iv) assess the consistency of

estimates of labor market transition rates across two specific dates by comparing panel and retrospective data, (v) assess the consistency of estimates of the level and trends of annual labor market transition rates across retrospective data from two different waves of the survey, and (vi) assess whether retrospective data can provide accurate trends of labor market aggregates, such as employment-to-population ratios and unemployment rates.

The rest of the paper is organized as follows. Section 2 reviews the theory and past evidence on measurement error in contemporaneous and recalled data. Section 3 includes a discussion of our data source and methods of analysis. Section 4 lays out all our findings on the various comparisons we make and section 5 concludes with recommendations as to what kinds of information can be reliably collected using retrospective questions, how to improve retrospective data collection strategies to obtain more reliable information, and potential methods for correcting measurement errors.

## **2. Theories and Past Evidence on Measurement Problems in Current and Recalled Data**

Whether labor market states and transitions can be accurately represented by current (contemporaneous panel) or recalled (retrospective) data is essentially an issue of measurement error. Surveys are attempting to measure a certain phenomenon—such as the current labor market status, or the date of the first job—but the data reported may be erroneous. The literature on measurement error suggests a wide variety of issues that might contribute to measurement errors in both current and recalled information. This section begins with a summary of the threats to research created by measurement error, then discusses some of the key issues that contribute to measurement error in current and recalled data. The section concludes with the evidence to date characterizing these problems, with a particular focus on findings relating to labor markets in developing countries.

### ***2.1 The implications of measurement errors***

The implications of measurement error depend substantially on the nature of the problems. Truly random errors in continuous variables sometimes do not present a substantial problem to research, as they will not affect estimates of key statistics such as means. In estimating linear regression models with a mis-measured continuous dependent variable,  $y$ , so long as the measurement error in  $y$  is random, results will not be biased (although standard errors will be increased) (Bound, Brown, & Mathiowetz, 2001). Random errors in an explanatory variable,  $x$ , will downward-bias or attenuate the estimated coefficient on  $x$  in a linear regression model (Bound, Brown, & Mathiowetz, 2001).

Random errors in categorical or binary variables are more problematic. For example, if a variable is binary, such as whether or not an individual is employed, an error must always be the opposite of the true value. That is, if the true value of some  $y^*=1$ , then the measured value,  $y$ , may be 0 or 1, and it is always the case that  $y-y^*\leq 0$ . Likewise if  $y^*=0$ , then the measured value,  $y$ , may be 0 or 1, and it is always the case that  $y-y^*\geq 0$ . Thus, the correlation between the true value and the error is always negative (Bound, Brown, & Mathiowetz, 2001). In the case of even random errors in a limited (categorical, binary) dependent variable, unlike in the continuous case, regression results will be biased downwards (attenuated). As in the continuous variable case, when the mis-measured variable is an independent, explanatory variable, this will lead to downward-biased (attenuated) estimates (Bound, Brown, & Mathiowetz, 2001).

Often, measurement errors are not random, but are instead systematic—that is, related to characteristics or covariates. In this case, measurement errors will bias both basic statistics and

regression coefficients in complex ways (Bound, Brown, & Mathiowetz, 2001). For instance, when studying the incomes of the self-employed, individuals with more education may keep accounting books and be able to more accurately report their incomes. If less educated individuals systematically under-report their incomes, this will systematically bias a regression estimating the relationship between years of education and income.

## ***2.2 Why do measurement errors occur?***

A variety of different processes can generate errors in data. Measurement problems can be either unintentional or intentional misreporting, and we discuss here a number of the processes that contribute to these different types of misreporting. The process of respondents providing information to survey data collectors can suffer from a number of errors. The question answering process for a respondent requires, first, comprehension of the question, the recollection of the information from one's memory, comparing the retrieved information with the original question, and communicating this information to an enumerator (Bound, Brown, & Mathiowetz, 2001). A large body of literature focuses on the recall or retrieval process and the nature of errors in recall. These are particularly likely to be affected by the recall period. That is, the longer the recall period (the further back in time the event in question is), the more likely that respondents will report with error, although the extent to which this is a problem varies substantially over studies of different outcomes (Bound, Brown, & Mathiowetz, 2001).

In reporting the dates of various events, the misreporting of dates may be a function of how far back in time the event occurred. Respondents are more likely to move forward the date ("forward telescoping") of an event that has a short reference period (a few weeks) while respondents are more likely to move back in time an event that occurred a year or more in the past (Bound, Brown, & Mathiowetz, 2001). Studies of panel data on dates have identified what is commonly referred to as a "seam effect," i.e. excessive numbers of changes at the "seam" between one study period and the next (Bound, Brown, & Mathiowetz, 2001).

The "salience" or importance of events may affect the accuracy with which they are reported (Bound, Brown, & Mathiowetz, 2001; Judge & Schechter, 2009). For instance, unemployment spells of only a few weeks may be of lower salience than unemployment spells that last a year and therefore be more likely to be forgotten. Individuals may forget when, or even whether, certain events occurred at all. If individuals do remember events, they may not readily remember the exact timing of events. This leads to measurement errors such as "heaping," where individuals tend to report certain numbers as responses (Roberts & Brewer, 2001). For instance, respondents often report adult ages in years in multiples of 5 or child ages in months rounded to the nearest year or half year (Heitjan & Rubin, 1990; Roberts & Brewer, 2001). Question and questionnaire design can play an important role in whether respondent errors occur. Identifying the best respondent within a household, deciding on the level of aggregation for data, and asking for information in the most appropriate units and for the most appropriate reference period are important elements of design that will affect the accuracy of measurement (Grosch & Glewwe, 2000; Puetz, 1993).

As well as unintentional errors, primarily due to difficulties accurately recollecting information about states and events, responses in surveys may suffer from intentional respondent misreporting. Particularly for topics that relate to behaviors or states that have strong connotations of social (un)desirability, such as the intention to send children to school or the receipt of charity, respondents may misreport. Under-reporting will occur for socially undesirable phenomena, and over-reporting for desirable phenomena, generating "social desirability bias" (Bound, Brown, & Mathiowetz, 2001).

As well as respondents providing inaccurate information, interviewer practices and data processing may generate inaccurate information. Differences arise in the quality of government-collected and academic researcher-collected data (Judge & Schechter, 2009), which may be due to differences in the qualities and characteristics of interviewers and data processing. As well as subtler issues such as interviewers with poor training and weak incentives (Puetz, 1993), outright fieldworker fraud may occur. Such fraud is particularly likely to bias panel data estimates (Finn & Ranchood, 2013). Quality control during data collection can help address such issues (Puetz, 1993).

### ***2.3 Evidence on the extent of errors in survey data***

The Malaysian Family Life Survey (MFLS), with panel data 12 years apart and substantial retrospective elements, suggests some of the issues that may occur in developing country data. The survey focused on issues of fertility and health and therefore targeted ever-married women. The MFLS also collected life histories on issues such as employment, migration, and marriage. The findings demonstrate that substantial errors can occur, but also that reporting of retrospective events can be quite accurate. For instance, while 95% of the ever-married sample reported being currently married in the first wave of the survey in 1976, twelve years later, only 84% of panel respondents reported in 1988 that they had been married in 1976. This difference is substantial and statistically significant. However, the same rate of mortality for children born prior to 1976 results from both the 1976 and 1988 interviews (Beckett, Da Vanzo, Sastry, Panis, & Peterson, 2001). The level of detail in the question affected the accuracy of reporting as well; for instance, agreement was much higher in reporting whether a child was ever breastfed than the duration of breastfeeding. The salience of events also mattered; women reported inter-district moves (a more substantial move) more consistently in 1976 compared to 1988 than intra-district moves (only 80% of the rate of moves prior to 1976 was reported in 1988). Substantial “blurring” of dates also occurred, with mothers not reporting exact months or years for children’s births when they were relatively further back in time across the two waves. Quantities were more likely to be rounded (akin to heaping), with rounding increasing with the length of recall. A study found different reporting errors with the MFLS to be related to respondent characteristics (Beckett, Da Vanzo, Sastry, Panis, & Peterson, 2001).

A number of studies have also been conducted on measurement of income, assets, and consumption. A study of boat-based fisherman in India looked at self-reported income over 34 months and compared it to administrative data from the fishermen society (de Nicola & Giné, 2014). The study found that the mean of income is maintained but variance reduced when going back 24 months. Findings suggested that boat owners reverted to inference, i.e. reporting mean income, as recall periods lengthened. Asking about the date of boat purchase directly elicited responses of similar quality to asking in relation to time cues (anchoring) important to the respondent; using unimportant time cues generated substantially worse results. The timing of the question within the survey did not, however, affect results (de Nicola & Giné, 2014).

Using data from Africa, Beegle, Carletto, and Himelein (2012) look for recall bias in agricultural data from three household surveys in African countries. Agricultural data usually refer to an agricultural season or year, and may be subject to recall bias of varying degrees depending on, for example, the time since harvest. The authors regress information on harvest sales and input use on the time elapsed between harvest and interview. They find little recall bias, although more salient events may be reported more accurately.

As well as examining income data, a number of studies have examined recall errors in expenditure and consumption data. Using household survey from Vietnam and resurveying respondents,



Nakata, Sawada and Tanaka (2010) find that questions on total rather than categorical expenditure suffer less recall bias. They also find that errors are systematically related to household size, and that errors are more serious for goods produced for own consumption than purchased goods. Errors tend to be mean-reverting, which will bias coefficients downward. Beegle, De Weerd, and Friedman (2012) compare eight different methods for measuring household consumption, comparing to a benchmark of personal diary use other diary and retrospective approaches. Recall is lower for other approaches than diaries, with particularly acute problems for poorer, larger, and less educated households.

Using United States data including a longitudinal survey, long-term retrospective recall data, and company records, comparisons demonstrated that the means of earnings were very similar in retrospective responses and company records. However, transitory variations were under-reported, generating another case of mean-reverting errors (Gibson & Kim, 2010). In a similar vein to our work on labor market dynamics, Dercon and Shapiro (2007) examine the role of measurement problems in poverty dynamics in panel data. They review past work on poverty mobility and discuss several key errors that are also relevant in our work: (1) inaccurate measures of income or consumption (2) price deflation and (3) mismatching of households over survey waves.

### **3. Data and Methods**

#### ***3.1 Data sources***

To compare results from panel and retrospective data, it is necessary to have a survey that collects contemporaneous data at different points in time as well as retrospective data for the same individuals. The only survey that currently meets these criteria in the MENA region is the Egypt Labor Market Panel Survey (ELMPS). With waves in 1998, 2006, and 2012, it is possible to use the ELMPS to compare retrospective and panel data over multiple periods. The ELMPS is a nationally representative household survey with detailed modules on current and past labor market statuses. Of the original 23,997 individuals interviewed in 1998, 13,218 (55.1%) were re-interviewed in both 2006 and 2012. Of the 37,140 individuals interviewed in 2006, 18,770 (77.5%) were re-interviewed in 2012.<sup>1</sup> A retrospective panel of annual statuses is constructed from retrospective data in each wave and compared to panel and retrospective data from previous waves. Reporting of time invariant information, such as parent's education, is also compared based on reports in different waves of the survey.

A particularly important element of our analyses relies on the labor market history section of the ELMPS surveys, which is administered to all individuals 15 and older who ever worked. In 2012, this section asks for the start dates (year, month) and characteristics of labor market statuses lasting six months or more from the time the individual exited school.<sup>2</sup> A status is defined as any labor market state lasting six months or more, be it employment, unemployment or out of the labor force. If the individual is employed in that status, she or he is asked about the details of such employment, including employment status (wage work, self-employment, etc.), sector of employment, occupation, economic activity, incidence of a formal contract and/or social insurance coverage, location of work, and reason for changing the status. The questionnaire inquires about the first four statuses lasting six months or more. Statuses of less than six months are dropped and if four statuses are not enough to reach the current status, the fifth and later statuses are also dropped. However, the total number of employment spells and their start and end dates can be obtained from the life

---

<sup>1</sup> See Assaad and Roushdy (2009) and Assaad and Krafft (2013) for a discussion of attrition from the various waves of the ELMPS.

<sup>2</sup> For individuals who never went to school, the retrospective period starts at age 6.

events calendar section of the questionnaire. In addition to the first four statuses and the current status, which is obtained elsewhere in the questionnaire, the questionnaire inquires whether the individual's current status (in early 2012) was different from their status in the month prior to the January 25<sup>th</sup> 2011 revolution. If it was, the questionnaire elicits information about the individual's status during that month.

It is important to note that in the preceding waves of the ELMPS survey (1998 and 2006), the labor market history questions were sequenced differently. Specifically, these waves of the survey used a reverse chronological order in eliciting labor market trajectories as compared to the chronological method used in 2012. In 1998 and 2006, the questionnaire first inquired about the current labor market status, then the previous status and the status previous to that, collecting information about the date of start of each of these statuses. In addition, information was collected in a separate part of the questionnaire about the first job in which the individual was engaged for a period of more than six months. Unlike the 2012 wave, the 1998 and 2006 waves did not contain a life events calendar and therefore no information on the total number of primary jobs the individual engaged over his/her lifetime. This questionnaire design implies that initial unemployment and out of labor force states could be missed, as well as employment states between the first job and the pre-previous status.

### **3.2 Methods**

To compare retrospective and panel data, the retrospective data were mapped on to panel data from previous waves in such a way that retrospective and current information is available for the same individual at the same point in time. We then draw on the econometric literature on measurement error to assess and compare the data sources and suggest possible corrections to account for measurement error (Black, Berger, & Scott, 2000; Bound, Brown, & Mathiowetz, 2001; Carroll, Ruppert, Stefanski, & Crainiceanu, 2012; Fuller, 1987; Magnac & Visser, 1999).

As a first check on the accuracy of the panel data, we begin by comparing the consistency of time invariant information across different waves of the panel. We do this for own education for adults 30-54 in 1998, father's sector of work when the individual was 15 years of age, and recalled costs of marriage. We then compare labor market statuses at a given point in time (1998 and 2006) across retrospective and panel data to assess the accuracy of recall and identify statuses that are particularly prone to erroneous recall. We subsequently estimate a multivariate model of the probability of alignment in labor market status between the two kinds of data as a function of individual characteristics, whether the information was elicited from the individual him/herself or a proxy respondent, the nature of the past employment status itself, and the contemporaneous employment status in 2012.

The next step is to assess the consistency of reporting of labor market transitions in retrospective and panel data. To do this, we convert the retrospective data into an annual retrospective panel, which contains information about the main labor market variables every year since the individual exited school for the first time. Using this retrospective panel we calculate the rate of change in employment status from 1998 to 2006 using the respective waves of the panel for those dates to the rates of change over the same period as reported by the 2012 retrospective data. We then move to comparing annual transition rates derived from the retrospective data in different waves of the survey.

In examining labor market transitions, we examine two types of transitions of particular interest to the study of labor market dynamics: transitions among employment, unemployment and out-of-

labor-force states, and job-to-job transitions among the employed. Within the first type, we include job-finding rates for the unemployed and those out of the labor force, and separation rates from employment to either unemployment or out of the labor force. The second type includes two-way transitions across different types of jobs, such as public and private employment and wage and non-wage work. We examine how different waves of the retrospective data generate transition rates, by type of transition. Finally, we revisit the question of whether the levels and trends in important labor market measures, such as the employment-to-population ratio and the unemployment rate, can be accurately assessed using the retrospective data, by comparing different waves of retrospective data and the retrospective and contemporaneous sources of these data, such as the official labor force survey.

## 4. Findings

### *4.1 Consistency of reporting of time-invariant information across different waves of a panel survey*

#### *4.1.1 Own education for adults*

The accuracy of the characteristics individuals report in any survey, such as their age, education, or labor market characteristics, plays an important role in researchers' ability to accurately describe economies and labor markets. Often researchers are focusing on cross-sectional, contemporaneous labor market characteristics. Comparing how individuals report static characteristics over time can help researchers understand how accurately contemporaneous characteristics are reported in cross-sectional data. Because the ELMPS is a panel, we can compare characteristics that should remain unchanged, such as education (for adults) as reported in different waves of the survey in order to assess their accuracy.

Figure 1 compares the reporting of education in 1998 with that in 2006<sup>3</sup> for individuals aged 30-54 in 1998. It is important to keep in mind that either the 1998 or the 2006 response could be inaccurate when they disagree, or both could be consistent (but still inaccurate) over time. Education is examined categorically, in terms of eight different categories. Overall, 79% of responses are the same over time, but there is substantial variation in terms of which education categories are reported consistently. For instance, 90% of those who reported being illiterate in 1998 report being illiterate again in 2006. The remainder primarily reported being able to read and write but having no education certificate (7%), which could be a genuine change in literacy, or having primary education (3%). The ability to read and write appears to be the most poorly reported, with only 34% of those reporting they could read and write but having no education certificate in 1998 reporting the same status in 2006. The most common response in 2006 for this group is being illiterate, which may represent a genuine decay in reading and writing ability. In general, when reporting is different, the reported alternative is usually a proximate level of education. For instance, 20% of those who reported general secondary in 1998 then report they attended vocational secondary in 2006.<sup>4</sup> Likewise 22% of those who reported post-secondary education in 1998 reported being vocational secondary graduates in 2006. Since the distinction is not always clear between special five-year vocational secondary programs and three-year

---

<sup>3</sup> Comparisons are not made with 2012 because, to reduce the burden of responding to the survey, respondents who answered educational questions in 2006 and had no change in education status were not re-interviewed about education.

<sup>4</sup> Smaller categories, such as general secondary, may suffer from more mis-reporting due to their tiny size. For instance, if random typos are uniformly distributed across the categories, more responses in the smallest categories are likely to be errors.

vocational secondary plus two-year post-secondary programs, this may have contributed to the lower consistency.

Although every effort is made to collect data from the individual him or herself, one dimension of data collection that is likely to affect measurement problems is whether or not the respondent is, in fact, the individual in question. Data consistency can also be assessed along this dimension of reporting. We compare the education status in 1998 and 2006 of individuals who were consistently responding for themselves and those who were not consistently responding for themselves.<sup>5</sup> Among individuals in the panel who were 30-54 in 1998, 71.3% reported for themselves in both waves (94.9% in 1998 and 75.7% in 2006). When the individual in question is consistently the respondent, there is a slight (but not dramatic) improvement in the consistency of data. There is only a one percentage point increase in consistency of illiteracy, primary, and university reporting, but larger improvements in read and write (5 percentage points), preparatory (6 percentage points), vocational secondary (6 percentage points) and post-secondary (4 percentage points) reporting.

Creating less finely aggregated categories, such as only four education levels, can lead to more consistency in reporting (Figure 3). Combining illiterate and read & write leads to a category that is consistently reported 93% of the time. Combining primary and preparatory education into a single category leads to 76% of respondents reporting consistently over time. A category that combines general, vocational and post-secondary into the definition “secondary” is 91% accurately reported over time. University and above was consistently reported as 94% over the two waves. Overall, it is clear that using more aggregated categories leads to more consistent reporting; more detailed categories should be viewed with some caution (particularly read and write and general secondary).

#### *4.1.2 Father’s Sector of Work*

Particularly for analyzing patterns across generations, it is necessary to gather information on parents’ characteristics, even if they are not present in the respondent’s household. The ELMPS collects information on father’s characteristics when the respondent was age 15 from respondents when the father is not in the household. Figure 4 shows the consistency of responses over time (2006 versus 2012) among respondents aged 30-54 whose father was not present in 2006 or 2012. Approximately 71% of respondents who reported their father worked in government in 2006 then report that their father worked in government in 2012, while 7% said their father worked in public enterprise and 22% worked in the private sector. Private sector work is relatively consistently reported (91% the same from 2006 to 2012). Consistency is most problematic in terms of identifying work in public enterprises. Only 35% of individuals who reported that their father worked in public enterprise in 2006 reported the same status in 2012. Instead, 40% reported their father worked in government and 25% reported their father worked in the private sector. As with education, certain categories are less clearly defined. The results suggest that respondents are sometimes inferring or reconstructing their father’s sector of work; the results suggest that, for instance, individuals may know that their father works in a utility but not know the sector of employment and reconstruct it.

One possible factor contributing to inconsistencies and misreporting of father’s status in particular is whether or not the respondent is actually answering the questionnaire. If, for instance, an individual’s spouse is responding to the questionnaire, they will often be in a very poor and

---

<sup>5</sup> When individuals were not consistently responding for themselves, it is possible that the same individual was responding in their place in both waves (e.g. a spouse) but the data does not allow us to determine whether this is so.

uniformed position to answer questions about their spouse's parents' characteristics when their spouse was age 15. For the same sample as in Figure 4, we examine whether father's sector is consistently reported depending on whether the respondent is (consistently) the individual in question. Responses when the individual in question answers in both 2006 and 2012 are compared to those where the individual in question does not answer in one or both of the waves. Overall, 69.4% of individuals in the sample were consistently reporting for themselves.

The patterns of reporting by respondent are presented in Figure 5. Having the respondent consistently be the individual in question improves the results at most a little. Reporting of private sector work and public sector work is nearly identical. Only in regards to public enterprise work does having the respondent consistently report appreciably improve consistency, 39% reporting public enterprise in both waves when consistently reported by the respondent and 33% otherwise.

#### *4.1.3 Recalled costs of marriage*

Understanding the investments individuals have made over time often requires asking about past outlays of expenditure. Individuals are expected to provide the costs of expenditures and investments when they occurred. However, especially when individuals are inferring or reconstructing the value of an investment, for instance inferring the cost of their housing at the time of marriage based on their current cost or value of housing, this can cause problems in assessing trends over time. Essentially, individuals may (fully or partially) update past expenditures from nominal to real terms. Figure 6 shows the trends in the total costs of marriage over time for individuals who were married in 2012 and present in both the 2006 and 2012 waves (they may not have yet been married). Marriage is an enormous investment for young people and their families, and the cost of marriage and its trends in Egypt and the MENA region are the subject of substantial concern and discussion (Assaad & Krafft, 2015a, 2015b; Assaad & Ramadan, 2008; Dhillon & Yousef, 2009; Salem, n.d., 2014, 2015; Singerman & Ibrahim, 2003; Singerman, 2007). The figure shows both the nominal (reported) costs and the real (inflated to 2012 LE) costs by year, as reported in both 2006 and 2012. Nominal marriage costs are clearly rising over time using both the 2006 and 2012 data, and the difference between those reports (when they overlap in time) may be due to inflation. Using the 2006 data and 2012 prices, it appears that from 2000 to 2006, marriage costs were flat or slightly declining, and averaged around 60,000 to 70,000 LE. Using the 2012 data and 2012 prices, looking at the same respondents' reported marriage costs from 2000 to 2006, it appears marriage costs were flat or slightly declining, but averaged around 90,000 LE. This is clearly inconsistent with what was reported in 2006, even when updated to 2012 prices. It appears individuals are partially (but not fully) updating nominal costs into real terms, as nominal costs for 2000-2006 as reported in 2012 are too high compared to 2006 nominal costs, but real costs for 2000-2006 as reported and then inflated using 2012 data are too high compared to 2006 reports updated into 2012 prices. Additionally, further investigation suggests different elements of marriage costs are updated differentially, likely related to how easy they are to recall or reconstruct.

Continuing to examine the 2012 data out to 2012 in real terms, it appears that marriage costs have fallen substantially over time, from around 90,000 in 2000-2006 to around 60,000-70,000 by 2012. This implies the cost of marriage over the 2000-2012 period decreased almost a third. By this comparison, marriage costs are falling over time. However, looking back at marriage costs as reported in 2001-2006, and updated to 2012 terms, marriage costs have remained essentially constant, in the 60,000-70,000 range (in 2012 LE). This is evidence that, particularly when asked about events that are now a number of years in the past, individuals may be inferring their value or inflating into current terms. This suggests that retrospective data should not be used to assess

time trends for investments and financial outlays; repeated cross sections or panel data are required. Comparing investments in the few years preceding a survey wave to investments in the few years preceding different survey waves will be more accurate for such cost data.

## ***4.2 Comparing labor market statuses across retrospective versus panel data***

### *4.2.1 Alignment of labor market statuses in general*

Individuals' labor market statuses, namely whether they are out of the labor force, unemployed, or employed, and if employed, their employment status, are at the heart of labor market surveys. Both the accuracy of individuals' contemporaneous statuses and the accuracy of the labor market histories they report are of great interest. The latter are particularly important for assessing labor market dynamics. This section first assesses whether aggregate statistics vary by data source and then whether individuals report consistently across contemporaneous and retrospective data.

Figure 7 presents aggregate labor market statistics by gender for 1998 and 2006 using both contemporaneous statistics from the waves of the panel and retrospective reports from 2012 for those years. Notably, for males, the aggregates from both the retrospective and contemporaneous data are quite similar, with a few exceptions. Reporting of public sector work, private formal and informal regular wage work, and self-employment are fairly similar. Irregular wage work is differentially reported in the retrospective data, which is likely because hours of work fluctuate over time; individuals may remain at the same job over time, but report that it is irregular in 2012 and map that back onto their status in previous years. Changes in regularity within the same job are also not captured within the labor market history and are difficult to detect. Regularity of employment should therefore not be assessed from retrospective data. Being an employer was more likely to be reported in the contemporaneous than retrospective data. This is again because whether a self-employed worker hires other workers or not is a varying, time-bound status that is not easily recalled.

For individuals 30-54 in 2012, their ages would have been approximately 24-48 in 2006 and approximately 16-40 in 1998. Thus, it is only in 1998 that many would have been out of or transitioning into the labor force (unemployed). These statuses appear to be under-reported in the retrospective data, when comparing 1998 contemporaneous data to the retrospective data for 1998 from 2012. For instance, while 7% of males were unemployed in 1998 contemporaneously, in the retrospective data only 2% of males report being unemployed at that date. Likewise in the 1998 contemporaneous data, more individuals report being out of the labor force.

There are so few females in a number of labor market statuses that our assessment for women focuses primarily on the public sector, unpaid family work, unemployment, and being out of the labor force. Public sector work is quite consistently reported in the aggregates, which may be due in part to the stability of this employment status. Unpaid family work, which includes subsistence work, is much more frequently reported in the contemporaneous data (6-10% across years) than in retrospective data (3-4%). This may be in part because individuals are only asked the labor market history in 2012 if they report having ever worked in market work, and unpaid family workers may frequently switch into and out of market work, sometimes producing agricultural goods for their own subsistence and sometimes selling them on the market. Unemployment is also more frequently reported in the contemporaneous data than in the retrospective data. This is likely due to the fact that many women who search for work never end up working (Assaad & Krafft, 2014) and thus are not asked the questions in the labor market history. As a result of these patterns in

reporting employment, being out of the labor force is higher in the retrospective than contemporaneous data for women.

A number of labor market statuses are particularly prone to misreporting over time, comparing retrospective and contemporaneous data. Figure 8 presents the distribution of retrospective statuses going back from 2012 to various years by the status reported contemporaneously from 2006 or 1998 for individuals 30-54 in 2012, by sex. There also is somewhat greater inconsistency comparing 1998 statuses than 2006 statuses, which is likely due to recall deteriorating over time. While public sector employment tends to be reported quite consistently, other labor market statuses are frequently not reported consistently. Formal private sector work tends to be more consistently reported than informal work. Most reporting of wage work is consistent, but the type of wage work is not consistently reported. Distinctions between self-employment and being employers are likewise blurred. A higher degree of inconsistency is also apparent for those who were unemployed or out of the labor force. Some of this may be because the duration of these statuses is shorter, so the contemporaneous status may be off relative to the status that is measured as the predominant status for the year in the retrospective data. Females have a much higher probability than males of reporting that they are out of the labor force (which is their predominant status). Less formal forms of employment in the contemporaneous data for women, such as being an employer, self-employment, and unpaid family work are particularly likely to be reported as being out of the labor force. Being unemployed suffers from a similar problem, likely, as mentioned earlier, due to the large share of women who are unemployed but ultimately never find work.

These patterns, as with education, suggest a number of issues for analyzing labor market statuses and dynamics. For instance, a category of private wage work, incorporating regular formal and informal and irregular workers would be more consistently reported than the disaggregated categories, and transitions between regular/irregular and, to a lesser extent, formal/informal may be poorly reported over time. Self-employment and being an employer also often are mixed up, and might be better combined into a single category. For females, retrospective data should be treated with particular caution, as women may not report ever working when they have done so, or report being out of the labor force when they were in fact working in the private sector. Although it is not certain from the comparison of the contemporaneous and retrospective data which is correct, contemporaneous information on women's status shows greater differences from retrospective data than the same comparison for men.

Using more aggregated categories of employment status leads to somewhat greater consistency in responses. Figure 9 compares the distribution of retrospective statuses going back from 2012 to various years of by the status reported contemporaneously from 2006 or 1998, using only four categories: public, private wage, non-wage, and not working. Public sector wage work continues to be the most consistently reported category, as before. For males, private wage work is fairly consistently reported comparing 2012 retrospective statuses to 2006 (72%) and 1998 (66%). More males tended to retrospectively report that they were not working when they reported they were working in private wage work in 1998 (10%) than 2006 (3%). Only around half of those who reported being non-wage workers in 1998 or 2006 reported the same status in the retrospective data (55% for 2006 and 47% for 1998). Most of the remainder reported private sector wage work. Half (57%) of those who reported not working in the 1998 wave also reported not working in the 2012 retrospective data for 1998. Despite the recall time being shorter, there was less consistency between those who reported not working in 2006 and the 2012 recall data for 2006 (27% consistency). This may be because those males 30-54 in 2012 were 16-40 in 1998, and therefore

were more likely to have a long-term status of not working (preceding entry into work for the first time). Those males 30-54 in 2012 were 24-48 in 2006, and so their time spent not working would more likely have been short in duration and easier to forget in the retrospective data. Not working, for instance short spells of unemployment, may not have even met the definition for a status lasting six months or more and therefore would not have been included by design in the labor history.

Although aggregating labor market statuses causes some important improvements in consistency across males' labor market histories and contemporaneous statuses, there is less improvement in consistency in females' statuses, primarily because they fail to report employment at all. The problems associated with detecting employment even contemporaneously among marginally employed women in agriculture and animal husbandry in Egypt are well known (Anker & Anker, 1995; Assaad, 1997; Langsten & Salem, 2008). These problems are compounded when the question refers to a reference period well in the past. Women who were in public sector work according to their 1998 or 2006 reporting do consistently report that in the retrospective data and those not working according to the 1998 and 2006 waves report not working in the retrospective status. However, less than half of those in private wage work in one wave reported this in their retrospective data for 2012 (43% for 2006 and 24% for 1998). Consistent reporting of non-wage work is even lower (27% for 2006 and 13% for 1998). The inconsistencies are primarily due to individuals saying they were not working at the time. Further examination of the data demonstrated that a key problem is detection of whether women ever worked at all. Among the women examined, just two-thirds (67%) of those who were identified in 2006 as engaging in market work reported that they ever worked in 2012. Likewise just 73% of females who were identified as engaged in market work in 1998 reported ever working in 2012. This was not a problem for men (<1%). The problem is primarily driven by women who were no longer working in 2012; all of those working in 2012 were, of course, identified as having ever worked. However, among women who were not working in 2012 but were working in 2006, only 16% reported ever working in 2012. Among women who were not working in 2012 but were working in 1998, only 18% reported ever working in 2012. Only those who report ever working are asked the labor market history, and thus these women are considered to never have worked and no labor market history data is collected.

We had initially expected substantially more consistent reporting of labor market statuses when the individual was responding for him/herself. However, that does not appear to be the case. Figure 10 shows the collapsed labor market statuses by sex and reporter status comparing the 2006 contemporaneous data to the 2012 retrospective data for 2006. Some statuses are more consistently reported but others are not. For men, consistency in reporting public sector work is slightly higher but for women rates are similar. For both men and women private wage work is lower when the individual is consistently the respondent. When the individual reported being a non-wage worker in the 2006 wave, for both males and females this is much more consistently reported when the respondent is reporting (60% consistency for males when the individual is consistently the respondent, 48% if not and 34% consistency for females when the individual is consistently the respondent, 12% otherwise). Reporting not working tends to be more consistent for males when the individual is *not* consistently the respondent, but is similar for females. The lack of higher consistency when the individual is reporting for him or herself could be due to a variety of reasons. It may be that individuals are more accurate reporters, but not necessarily more consistent reporters, in that when others are reporting on behalf of an individual they provide consistent but potentially inaccurate responses, or increase consistency by simplifying the labor market history.



Further analysis of the data demonstrated that reporting whether women ever worked at all varied substantially by the respondent. Among the women examined where the respondent was not providing data, just 55% of those who were identified in 2006 as engaging in market work reported that they ever worked in 2012, compared to 71% when the respondent was consistently the individual herself. Among women who were not working in 2012 but were working in 2006, only 11% reported ever working in 2012 when it was not consistently the respondent reporting, and only 19% when it was the respondent reporting. While both illustrate extremely low rates of reporting work, having the respondent as the reporter did lead to increased accuracy in regards to ever working.

#### *4.2.2 Recalling past unemployment spells*

While in the aggregate labor market statistics are not substantially different, the inconsistency of individuals' responses over time is troubling. This section attempts to analyze some of the patterns and sources of disagreement in the data sources, focusing on the case of unemployment, the occurrence and duration of which is of particular interest within the Egyptian and MENA labor markets (Assaad & Krafft, 2014; Kherfi, 2015). The inconsistencies between contemporaneous unemployment and retrospective unemployment reporting could be occurring for a variety of reasons. Because only individuals who ever worked are asked the retrospective questions, excluding women who sought but never began work, this section focuses solely on the unemployment dynamics of individuals who ever worked and examines several different questions essentially revolving around the issue of why there are inconsistencies across the data sources. Do individuals report unemployment in their retrospective histories, but just during a different year? Are shorter spells of unemployment more likely to be forgotten over time?

Since the primary concern is that unemployment is under-reported in the retrospective data, in Table 1, for those who reported unemployment in the 2006 or 1998 waves,<sup>6</sup> we examine the reports of unemployment in the retrospective data and a number of characteristics, including the mean current unemployment duration at the time of survey, and the percentage of individuals who experienced short (less than six month) current unemployment durations as of the time when they were surveyed. Notably, for those unemployed in the contemporaneous data for 1998, just 9% of unemployment statuses in the 2012 retrospective data for 1998 included a report of unemployment. The alignment was slightly better in 2006, when 13% were aligned. Individuals who were unemployed in 2006 were more likely to report unemployment within one year (5%) or two to five years (12%) than those unemployed in 1998 (1% reported unemployment within one year and 7% within two to five years). More individuals reported being unemployed at some point more than five years out in 1998 (11%) than in 2006 (7%). Reporting of unemployment is less accurate, both in terms of reporting at all and the timing of unemployment, going further back in time.

Notably, 71% of individuals who were contemporaneously unemployed in 1998 did not ever report being unemployed in the labor market histories. Because the labor market histories in 2012 go forward in time, it is possible that unemployment occurred after the fourth status (the last status asked in the labor market history). Therefore, those with a fourth status are separated out, and comprise a small share of the distributions (6% for those unemployed contemporaneously in 1998 and 4% of those in 2006).

The characteristics of unemployment, specifically its duration to date as of the contemporaneous status reported in 1998 or 2006, are related to the probability of accurately reporting. Those whose

---

<sup>6</sup> Data is not separated by gender or restricted by age so as to ensure an adequate sample size.

reporting aligned had, on average, long durations of unemployment to date, 23 months in 1998 and 16 months in 2006. Those who reported their unemployment, but with imprecise timing, tended to have shorter durations of unemployment than the average, a year or less. Those who never reported being unemployed in the retrospective data had slightly longer than average unemployment durations. Overall, it appears that gathering data on historical patterns of unemployment, even among those who ever worked, is likely to produce substantially different results than using contemporaneous data. It seems likely that retrospective data will both under-report past unemployment and distort its characteristics.

Having the respondent reporting for his or her self does not substantially improve the reporting of unemployment. Table 2 presents the patterns of unemployment reporting by gender and whether or not an individual was consistently the respondent. Males are less likely to report their unemployment as aligned overall (8% for males, 26% for females) with similar rates for those reporting consistently for themselves and otherwise. Overall, males are slightly less likely to report never being unemployed and have no fourth status if responding for themselves consistently (63%) than if otherwise (67%), but the opposite is true for women, among whom 48% of those responding for themselves report never being unemployed and have no fourth status, compared to 40% of those not consistently responding for themselves.

#### *4.2.3 Multivariate models of alignment between retrospective and panel data*

Particularly concerning in assessing measurement error is whether errors are systematic (related to covariates). Such relationships will bias any attempts to examine the relationship between covariates and mis-measured outcomes. To assess whether there are systematic patterns of misreporting, in Table 3 we run probit models for whether individuals' responses about their contemporaneous (panel) data in 1998 and 2006 were consistent with their (2012) retrospective data for those years. Models are restricted to those 30-54 in 2012 and run separately for males and females and therefore allow for a comparison of how characteristics are related to reporting both by gender and over varying retrospective spans from 2012.

The probability of alignment in reporting is high for the reference case, a 30-34 year-old university educated individual living in Greater Cairo, who did not consistently respond for him or her self, was a public wage worker in the 1998/2006 contemporaneous (panel) data, was employed in 2012, was a regular worker in 2012, and was a formal worker in 2012. For retrospective data referring to 1998, the reference case has a probability of alignment between retrospective and panel data of more than 0.9. This is actually lower, around 0.8, for retrospective data referring to 2006. Those 30-54 in 2012 would have been 24-48 in 2006 and 16-40 in 1998. Individuals may have an easier probability with retrospective recall about first statuses than subsequent non-current statuses that causes the 1998 data, with more first statuses, to be more consistent. Compared to the university educated, for males recalling distant (1998) statuses, all other education levels perform significantly worse, but this pattern does not hold for females or males recalling less distant (2006) statuses.

For males, compared to those 30-34 in 2012, those 35-44 in 2012, but not those 45-50 had significantly less alignment. For women there was at most a small increase in consistency in reporting among older females (45-50) for more recent (2006) statuses. Few regional differences occurred, with only slightly better alignment in Upper Egypt for male's more recent (2006) statuses. After controlling for other characteristics, there were not significant differences in consistency dependent on whether or not the respondent was consistently the individual in question.

Where large differences did occur was by both the retrospective status and 2012 employment characteristics. Compared to public wage workers in the contemporaneous data (1998/2006), private wage workers were significantly less likely to have consistent reports, non-wage workers even more so. For men, those not working in the contemporaneous data were also significantly less likely to report consistently, but there were no such differences for women. The magnitude of the differences is substantial; non-wage males had around a 30 percentage point higher probability of disagreement, and non-wage women 66-72 percentage points. For males, there are mixed differences comparing the effect sizes back to 2006 versus 2012. For females, more recent reporting is consistently more aligned, although not by large margins. In terms of 2012 employment characteristics, females not employed in 2012 are significantly more likely to consistently report their 1998 status, but not their 2006 status, while for males those not employed in 2012 are significantly less likely to report their 2006 status but not their 1998 status. Both males and females who were irregular in 2012 were significantly less likely to report their 2006 statuses consistently but not their 1998 ones, possibly due to the rising volatility of their employment being relatively recent (Assaad & Krafft, 2015c). Those males who were informal in 2012 were significantly less likely to report their 1998 status consistently. Overall, there are mixed relationships between 2012 status and recall of past statuses, but definite disagreements related to the contemporaneous (panel) employment type in the preceding 1998 or 2006 wave.

#### ***4.3 Comparing labor market transition rates across retrospective versus panel data***

An important application of retrospective and panel data on labor market statuses is measuring transition rates between different labor market statuses in order to assess labor market dynamics. We have demonstrated that there could be substantial misalignment between contemporaneously measured statuses and ones measured by means of retrospective questions, but also that the overall distribution of statuses is fairly similar (Figure 7). If the measurement errors are primarily an issue of random errors in the reporting of the timing of statuses, measures of labor market transition rates could still be fairly sound. However, if entire statuses are lost (as appears to be the case for unemployment), then measures of labor market dynamics will be understated and will point to a more rigid labor market than is actually the case. Because the ELMPS contains three panel waves, it is actually possible to assess labor market transition rates by using either purely retrospective or purely panel data. This section specifically compares transition rates, by status, from 1998 to 2006, based at first on the 1998 and 2006 panel data, and second, on the retrospective data collected in 2012 for 1998 and 2006. This analysis is performed only for individuals who appear in all three waves and who were 30-54 in 2012. The status used for classification purposes comes from either the retrospective or the panel data, depending on which data are being used to calculate the transition rates.

There are some key points to keep in mind when considering this comparison. The contemporaneous status is (as is the case throughout this paper) the “usual” status in the 3-month period preceding the survey, if an individual is employed. In the retrospective data coming from the labor market history module of the survey, statuses have to be at least six months long to be reported. It is therefore likely that in the panel data some of the transitions that are detected relate to statuses that lasted less than six months and that would not be observed by definition in the retrospective data. This would tend to inflate panel data transition rates upward, but probably not by much. We know from the 2012 contemporaneous data that only 1.6% of employed individuals have a different primary job in the reference week than in the reference three months, suggesting that short-term transitions are rare. Transition rates in the panel data are therefore only likely to be

inflated by a few percentage points at most. Although the probability of reporting statuses across panel and retrospective data is fairly similar (Figure 7), the differences that do exist are going to affect the measurement of transition rates as well.

In Figure 11, the rates of change in the various labor market statuses are assessed using panel and retrospective data sources, according to the 1998 status. Notably, transition rates for males are understated by about half in the retrospective data relative to the panel data (35% versus 59%) and by about two-thirds for females (9% versus 33%). Looking across statuses, every employment status in 1998 suffers from under-reporting problems in the retrospective data, but to varying degrees. For males, transitions out of unemployment and OLF statuses are fairly comparable, but this is not the case for females.

As well as differential rates of change, there are differential patterns of change in terms of which transitions are detected or not detected (not shown). More subtle transitions, such as transitions from informal to formal private wage work or from employer to self-employed and vice versa, are more likely to be missed in the retrospective data. More distinctive transitions—such as those between public and private sector jobs and between wage and non-wage work—are also somewhat under-reported in the retrospective data, but to a lesser extent. Particularly for women, the retrospective data is less able to detect transitions into and out of the out of the labor force, a problem related to the issue we discussed earlier about the difficulty in detecting women’s self-employment and unpaid family labor in the Egyptian context. Women in the public sector are much more likely to report being employed in the past. Since they typically have low transition rates, this tends to understate overall transition rates for women.

#### ***4.4 Comparing the levels and trends of annualized labor market transition rates across two sets of retrospective data from different waves of the survey***

##### *4.4.1 Measuring annualized transition rates from retrospective data*

To further investigate the extent to which the ELMPS retrospective data suffer from measurement problems, we compare the transition probabilities obtained from the retrospective data for the same time period as assessed by different waves of the survey. This analysis could serve as a guide for researchers wishing to use similar surveys to generate annualized data from retrospective questions. Our overall conclusion is that retrospective data tends to understate the degree of dynamism of the labor market and the longer the recall period, the greater the information loss.

The retrospective data from the ELMPS suffers from two major problems, the first being the typical recall bias that attenuates the number of past transitions and the second being the tendency of respondents to only recall past employment spells and overlook non-employment spells. The latter problem may be an artifact of the confusion on the part of interviewers and respondents between past labor market statuses and past jobs. Accordingly, this section aims to deliver two key messages about the retrospective data obtained from the ELMPS surveys. First, transition rates tend to be underestimated when calculated using the annualized retrospective data. Second, the time trends of these transitions are relatively reliable when analysing job-to-job flows, especially if the time-series curves are smoothed. However, the trends for job/non-employment flows are not well identified; these trends are distorted not only because of recall errors but also due to the way the labor market history questions are interpreted in the field. Throughout the retrospective accounts, non-employment spells are often skipped, whether they are initial spells preceding first employment or interim spells between two employment statuses.

To illustrate some of the challenges in working with retrospective data, we demonstrate three different approaches to constructing retrospective data on the labor market. Firstly, a naïve basic panel of annualized labor market statuses is constructed using only the retrospective chapter of the questionnaire. The analysis is therefore limited to individuals who have ever worked and who report their past labor market statuses. Second we augment this labor market history data with information about recent unemployment spells from the current unemployment section of the questionnaire. We also incorporate information from the life events calendar on the timing of non-employment spells of those who are currently out of the labor force. We refer to this type of retrospective panel data as the “augmented panel.” Finally, we incorporate information from those who have never worked, which should be considered in all analyses of dynamics. By means of this analysis, we show that it would be quite misleading to rely only on the labor market history information for those who have ever worked to assess labor market dynamics. For these retrospective data to be useful, they must be combined with the information on current unemployment for new entrants as well as those who ever worked and with the longer employment history obtained from the life events calendar.

The dynamics we focus on are primarily the job finding ( $f$ ) and separation rates ( $s$ ), which can be defined as the share of employed,  $E$ , and non-employed,  $NE$ , changing states over time,  $t$ ;

$$f = \frac{NE_{t-1} \rightarrow E_t}{NE_{t-1}}$$

$$s = \frac{E_{t-1} \rightarrow NE_t}{E_{t-1}}$$

#### 4.4.2 Separation rates

Turning first to the dynamics of separation rates, Figure 12 compares the differences across the three different annualized retrospective data construction approaches using the 2012 retrospective data. The ELMPS 2012 questionnaire was redesigned to ask about the past labor market statuses in a chronological order. Consequently, the fourth status does not necessarily coincide with the current status for certain individuals and moving from the labor market history data to the current state information detects additional separations. Since separation rates are calculated relative to the employed, adding in the never worked is irrelevant. The addition of the life events calendar in the ELMPS 2012 might have also played a role, given that now individuals who have gone recently out of the labor force have been captured, using the start date of their inactivity status.

Although some dynamics are still missing in the new ELMPS 2012 design, as we show below, the new structure of the retrospective questionnaire has generated smaller bias in the dynamics obtained using the ELMPS 2012 than those obtained using the ELMPS 2006 and the ELMS 1998, especially when going back further in time. Yet a substantial amount of information about states individuals have occupied between the fourth status of the labor market history and the current status is lost. According to calculations from ELMPS 2012, about 34% of individuals who ever worked and had a fourth status have exited that status. These two points indicate that while the new design represents an improvement in the retrospective chapter in the questionnaire, the addition of a fifth and sixth status to the labor market history section of the survey might be valuable.

Using the augmented panels of the three waves of the survey, in Figure 13 we overlap the separation rates calculated over the years. This is done for both male and female workers between

15 and 54 years of age in year  $t$ . A remarkable jump in the separation rates in the most recent year of each survey is observed. This tends to be true for both men and women but is much more pronounced in the male workers' trends.

It may be the case that individuals lose accuracy as we ask them to recall labor market states further in the past. At this stage we can not confirm if this bias is due to recall error or to a questionnaire design issue. One additional possibility is that the jump in these separation rates is because people declare having lost their jobs right before the year of the survey out of fear that enumerators are tax collectors. The latter argument is unlikely given all the evidence about the underestimation of unemployment as retrospective data are overlapped with previous waves of the panel in the previous section. Overall, non-employment spells and therefore separation rates of workers are underestimated through the individuals' retrospective accounts.

Given that the survey was designed to capture only retrospective labor market statuses that last for more than 6 months, we modify the way our retrospective panels are augmented. By including all currently unemployed individuals, we might have captured current unemployment spells that would have lasted less than 6 months. Although this might actually be giving us a more accurate picture of dynamics in the labor market, we reclassify, for comparability, those short unemployment spells as employment lags. In the ELMPS 2012 and ELMPS 2006, those would be the unemployment spells that started during the second half of the years 2005 (for the ELMPS 2006) and 2011 (for the ELMPS 2012). This might actually bring up how crucial it is to analyze short-term unemployment spells in a separate chapter in future surveys.

In Figure 14, we replot the employment to non-employment separation rates for males using the reclassified augmented data. The separation rate increase in the year preceding the survey is of a smaller magnitude. Yet, we still note the increasing trend in the separation time series. We again note that the ELMPS 2012, most likely due to the chronological design of the retrospective accounts, is doing a better job than the ELMPS 2006 in capturing the transitions of individuals that are further back in time. Still, if we take more recent rates as true, both are underestimating the employment to non-employment transitions.

As a further investigation into how employment to non-employment transitions are underestimated using retrospective panels, we repeat the above exercise making a distinction between unemployment and inactivity states. We replot the three separation rate time series, first using employment to unemployment transitions then second using employment to inactivity transitions. Figure 15 and Figure 16 show that the jump in the separation rates continues to appear for both types of separations, although this may be due, in part, to including statuses lasting less than six months. The increase in the employment-to-inactivity separations seem to increase gradually over the most recent two years while the jump occurs suddenly for the unemployment separations.

If we examine the age distributions of these separations by year, throughout our observation period separations are concentrated around the age of 20 for employment to inactivity transitions and late twenties for the employment-to-unemployment transitions. This age distribution helps explain the under-reporting of these transitions in the retrospective section. These are interim non-employment spells that occur at early stages of an individual's labor market trajectory. Only the most recent non-employment spells are reported, while past job statuses are merged or over-reported.

Apart from measurement error issues, the observed dynamics confirm the static nature of the Egyptian labor market once an individual gets stabilized in a job. A lot of churning occurs among the young, either between employment and unemployment until they find what is probably a

“suitable” job from their perspective, or between employment and inactivity, most likely for their military service. This explanation does not mean however that if we exclude the young age group from our sample, we avoid this separation uptick.

#### *4.4.3 Job-finding rates*

Moving now to the dynamics of the job finding process, recall that these rates are the flow of workers from non-employment to employment relative to the stock of non-employed. In this case, augmenting our retrospective panels, first with the currently unemployed and then with the never worked does make a difference, unlike the case of separation rates. Figure 17 shows that adding information about current unemployment and inactivity spells is not trivial; it gives a higher estimate of the stock of non-employed especially during the last two to three years before the survey. These are most likely initial unemployment spells, i.e people who have never had a job and consequently are not captured in the retrospective part of the survey. Including everybody (ever and never work) in the analysis gives the conceptually correct estimate of the stock of non-employed. Yet, it is only through the current unemployment information that we can differentiate between unemployed and inactive individuals with more than four statuses.

As was true for the job finding rates time series, we again suspect an underestimation of the non-employed as we go back in time. This time, the job-finding rates calculated from our retrospective panels are over-estimated. The most reliable point, in terms of the level of the job-finding rate, is likely to be the most recent point. Figure 18 shows the overlap of the the three finding rates time series calculated from the three retrospective panels using augmented data and incorporating those who never worked. We note likely over-estimation in the job finding rates mirroring the likely under-estimation we noted above in the separation rates. If we set aside the levels issue, and focus of the trend of both finding and separation rates over time, we note that there has been a very slight increase in the job finding rates over time for males and almost no substantial change for the females. A relatively higher increase in the separation rates over time is observed for both male and female workers.

#### *4.4.4 Job-to-job transitions*

Having examined states and transitions between employment and non-employment, we now examine job-to-job transitions among the employed. The comparisons of retrospective and panel data showed that more aggregated employment statuses are likely to be more consistently reported. For instance, it appears that respondents have difficulty distinguishing between informal and formal employment states as well as regular and irregular work. Therefore, we limit the analysis of the retrospective transitions rates to three broad employment sectors, namely public wage work, private wage work and non-wage work.

In Figure 19, we overlap the job-to-job transition rates calculated using the ELMPS 2006 and ELMPS 2012 retrospective panels. These rates are obtained by dividing the number of workers transitioning from one sector to another between years  $t$  and  $t+1$  by the number of workers employed in the origin sector in year  $t$ . Generally, we observe a close overlap of the job-to-job transition rates obtained using the two different retrospective panel data sets. This finding suggests that using retrospective accounts give consistent conclusions about the trends of job-to-job transition rates over time, especially when these trends are smoothed. However, it’s crucial to note that the levels of these transition rates are under-estimated given what we saw earlier in Figure 11.

#### ***4.5 Do retrospective data provide accurate trends of past labor market aggregates?***

The problems we observe in assessing labor market dynamics using retrospective data also present challenges to assessing stocks over time. This section examines the stocks derived from the retrospective data for two specific statistics: the unemployment rate and the employment to population ratio (employment rate). Figure 20 illustrates the evolution of employment to population ratio from the augmented retrospective panels, including the never worked over our observation period. The pattern suggests that as we go back in time, we only retain the employment states of our sample and lose track of their non-employment history. We obtain as a result a decreasing employment-to-population ratio, which is not consistent with patterns observed contemporaneously in the panel (Assaad & Krafft, 2015c). The magnitude of the decrease differs from one survey to the other. The ELMPS 2012 seems to have less of a decreasing trend than the ELMPS 2006. The most likely explanation for this observation is the different structure of the ELMPS 2012 questionnaire. Asking individuals about their past statuses in a chronological order, starting with the first status, rather than backward in time as was the case in 1998 and 2006, may have increased consistency of employment trends. However, none of these approaches recovers the pattern of employment observed in the panel contemporaneous statuses, which is an increase in employment rates over time for men and a rise and fall in employment rates for women.

Superimposing the retrospective data and the unemployment rates from the official Labor Force Sample Survey (LFSS) further illustrates how the proportions of different labor market states and consequently labor market transitions get distorted if one uses retrospective data. Figure 21 shows that the retrospective data does not align with the evolution of Egypt's unemployment rate over the past two decades as reported in official statistics.

### **5. Conclusions**

The primary objective of this paper is to assess whether it is possible to collect information about labor market dynamics using retrospective data or if recall error is so great as to make panel data the only viable option. As expected, we conclude that it is possible to garner useful information on labor market dynamics from retrospective data, but one must be cautious about which information to trust and at what level of detail. One of our most basic conclusions is that information on past employment collected using retrospective data can be fairly reliable, so long as fine distinctions between employment states are not made. For instance, the distinctions between employer and self-employed, between formal and informal wage work, or regular and irregular wage work are not easily made using retrospective data. The regularity of work is something that can change frequently depending on the state of labor demand in the economy and should therefore not be a subject of retrospective questions. Even the distinction between self-employment and irregular wage work is sometimes difficult to make especially for men engaged in small-scale agriculture. Smallholders often do not have enough land to keep them fully occupied on their small farms and must often engage in multiple livelihood strategies that may either involve non-agricultural self-employment or irregular wage work.

In the case of women engaged in self-employment, whether in agriculture or outside agriculture, the distinction between being employed and not employed is hard enough to make in contemporaneous data, let alone in retrospective data. In Egypt, women in this kind of employment typically do not consider themselves to be employed and may move frequently between employment and non-employment states, as defined by international labor statisticians. To assess their current status accurately, researchers must use complex keyword-based questions that inquire about a large number of activities, and even this detailed approach often fails to elicit reliable



estimates of female participation in home-based self-employment and unpaid family labor (Anker & Anker, 1995; Assaad & El-Hamidi, 2009; Assaad, 1997; Langsten & Salem, 2008). It is impossible to ask questions at this level of detail about a retrospective period, casting doubt on the employment transitions obtained from retrospective data for women in self-employment. Conversely, transitions across well-defined employment states, such as between public and private wage work, or between public wage work and non-wage work can be captured fairly reliably using retrospective data.

Spells of non-employment interspersed between employment spells are usually hard to recall, whether they are unemployment spells or spells outside the labor force altogether. For instance, 71% of those observed as unemployed in the 1998 wave and 64% of the unemployed in 2006 wave of the survey never reported any unemployment at any time in the past in the retrospective data obtained from them in the 2012 wave. Thus transitions from non-employment to employment and vice-versa will be understated in retrospective data, with important implications for the accurate reporting of separation rates for the employed and job-finding rates for either the unemployed or those outside the labor force, and the stock of unemployed in past dates. Generally, these rates will be understated, and possibly increasingly so as we go back in time, confounding any measurement of trends. In contrast, trends describing job-to-job transitions can be captured more reliably using retrospective data.

Another conclusion we derived from analyzing the reporting of recalled marriage costs is that retrospective questions eliciting monetary amounts are unreliable at best. Even when asked to report the nominal amount paid at the time, at least some respondents tend to inflate the amount to their equivalent value at the time of the survey. It thus becomes impossible to ascertain monetary trends over time when some of the data is inflated and some of it is not.

This research has also produced valuable lessons about how to use existing retrospective data from the ELMPS or other similar surveys. It is tempting to create annualized retrospective panel data from the labor market history module of the questionnaire and use those to calculate various transition rates. However, the labor market history module of the questionnaire is only applicable to people who have ever worked, excluding people at risk of transitioning to employment who may not have ever worked. Moreover, because of the limitations on the number of states that the questionnaire inquires about (up to four), the retrospective data may not reach up to the current state. To correct for the possible biases that can result from this, the labor market history data must be augmented by information from the current employment or unemployment sections of the questionnaire and from the life events calendar, which can potentially include more transitions to and from employment. Finally, it needs to be augmented by adding individuals who have never worked but who are currently either unemployed or out of the labor force. In adding data from the current section of the survey, it is important to correct for the fact that current spells may last for less than six months and may therefore not be comparable to spells captured in the retrospective data. Individuals currently unemployed for less than six months should potentially be reclassified to their previous status to ensure compatibility of definitions.

Finally, this experience has allowed us to derive some important lessons on how to improve questionnaire design to collect more accurate retrospective data. First, in comparing the retrospective data from 2012 to the data from previous rounds, we determined it is preferable to ask questions about the individual's labor market trajectory in chronological rather than in reverse chronological order. It elicits better information about labor market entry and in particular about any initial unemployment spells prior to first employment. Second, we suspect that many

respondents (and possibly interviewers) interpreted status to mean job, contributing to the underreporting of non-employment spells. In future versions of these labor market panel surveys we recommend that separate questions be used for non-employment spells and for employment spells. The questionnaire should elicit first information about the non-employment spell just after exit from school, if any, and determine whether it was an unemployment spell or an out of the labor force spell. As before, spells of under six months would be ignored, but both the start and end date of the spell in terms of month and year should be recorded if possible. This would be followed by questions about the first employment spell and its characteristics, the next non-employment spell, if any, and so on until the fifth or even sixth non-employment and employment spells lasting more than six months are reached, if relevant. This strategy would increase the potential number of spells for which information is available to ten or twelve, instead of the maximum of four statuses in the current survey design. Another improvement to the questionnaire would be to ask those who have never worked for a period of more than six months prior to the interview and are currently inactive about whether they have ever sought employment and about the timing and length of the spell in which they were seeking employment, at least for the first time. Even though these changes will not eliminate recall bias, they could potentially reduce bias that results from questionnaire design.

Given budgetary and data availability constraints, the retrospective panels are currently the primary source of data in the MENA region that allow researchers study labor market dynamics. Having discussed the errors encountered in retrospective data, it is important to note that it is possible to use some remedies to attenuate these measurement errors and eventually produce unbiased (or possibly less biased) results. A possible solution would be to match biased moments obtained from retrospective data with more accurate moments obtained from auxiliary contemporaneous cross-sectional data. Of course, this could be obtained from the same dataset or an external data source, so long as comparability between the data sets is verified. In this case, one assumes that the information obtained from the contemporaneous data is the most accurate. Assumptions about the (functional) form of the “forgetting rate” or information loss in the retrospective data would also be required. Langot and Yassin (2015) correct the ELMPS aggregate labor market transition rates between employment, unemployment and inactivity states, obtained from the retrospective panels, using this methodology. They assume that the most recent year of the latter panels are the most accurate and that people report more distant events less accurately. The measurement error has a functional form that increases exponentially as one goes back in time. This methodology can allow the re-construction of corrected separation and job finding time series that can be used in the analysis of the macroeconomic trends of the labor market. This can even be extended to make use of the micro-level information available about the labor market transitions. Using the aggregate measurement errors estimated for the different types of transitions, one could distribute these errors in the form of weights to the individuals in the survey (Yassin, n.d.). Again, assumptions need to be made on how to attribute weights to the individuals. Yassin (n.d.) discusses two ways of doing so: (1) a naïve method, where all individuals are assumed to be corrected similarly i.e. proportional weights and (2) a differentiated method, where weights are predicted based on the probability an individual would make a certain type of transition. All the approaches assumed that the information in retrospective panels is correct, just a little bit over reported or under-reported with respect to the true contemporaneous points (i.e. true moments). Another possible solution, with a different assumption, would be to estimate the alignment rate, possibly the rate of telling the truth, and eventually creating a weight such that individuals who report the truth have higher weights. This requires however the availability of both micro-level

contemporaneous and retrospective information for the same individuals. In our case, it could be applied to the ELMPS but not to other datasets, for instance the Jordan Labor Market Panel Survey (JLMPS) and the Tunisia Labor Market Panel Survey (TLMPS), where for the time being only one wave of the panel is available. Drawbacks of how representative the sample becomes after the creation of such weights need to be also discussed.

To conclude, we believe that panel data with retrospective modules to fill in the gaps between waves of the panel are the best data we can hope for to study labor market dynamics. Some advanced countries have moved to continuous administrative data to study such phenomena. However, given the low administrative capacity of most developing countries and the high rates of labor market informality, such methods will not become practical in developing country contexts for a very long time. It is absolutely crucial in collecting such panel data that the individuals who split from their original households to form new households are followed, something that was done in the ELMPS, but is often ignored in short-run labor force survey panels. Proper accounting for panel attrition is also necessary. In the absence of such panel data, a great deal can be learned from properly designed retrospective questions, so long as researchers are aware of the limitations of these data. As a general rule, distinctions that are hard to make in contemporaneous data, like differences between regular and irregular employment, formality and informality, illiteracy and literacy, and non-employment and home-based self-employment for women, are going to be even harder to make retrospectively. Shorter spells and more frequent events in an individual's labor market trajectory are more likely to be forgotten. We have attempted in this work to highlight some of these problems, but we are in no way suggesting that analyses based on retrospective data are worthless. We are simply advising that proper caution needs to be exercised in interpretation and have provided some pointers as to where the potential pitfalls might lie.

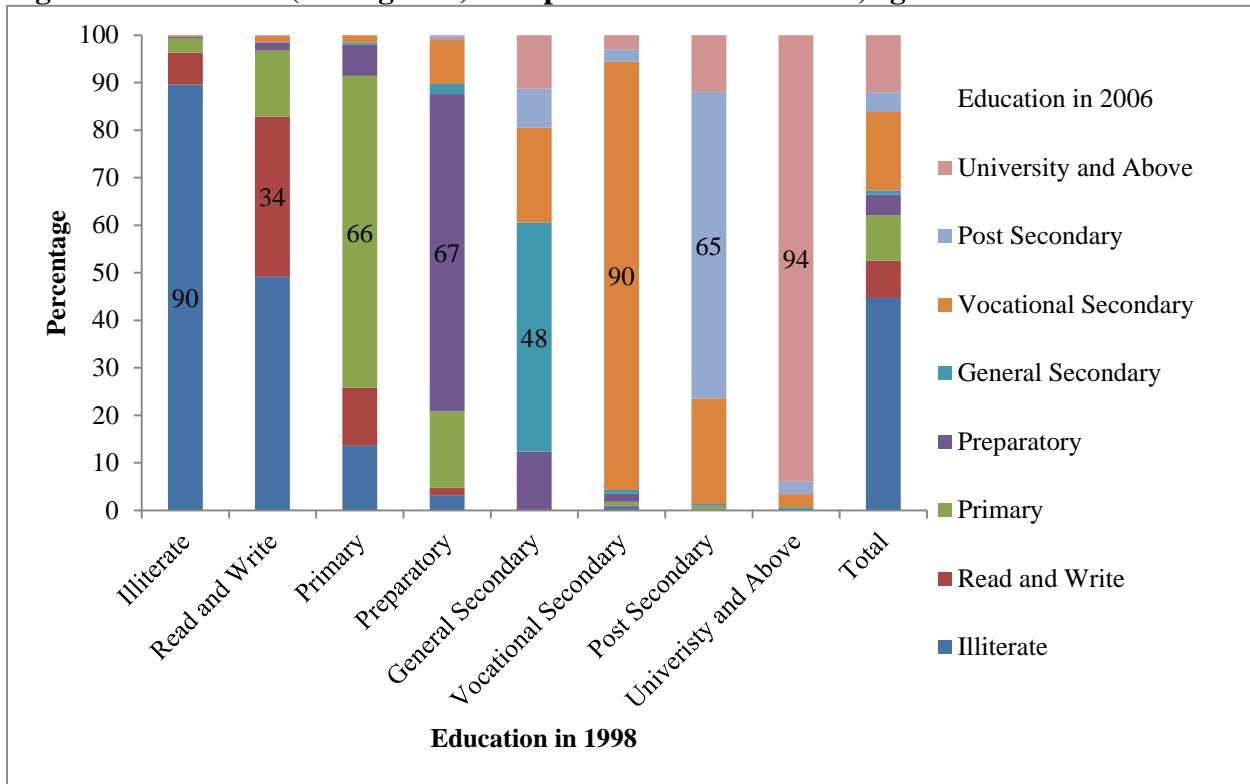
## References

- Anker, R., & Anker, M. (1995). Measuring Female Labour Force with Emphasis on Egypt. In N. F. Khoury & V. M. Moghadam (Eds.), *Gender and Development in the Arab World: Women's Economic Participation: Patterns and Policies* (pp. 148–176). United Nations University Press.
- Artola, C., & Bell, U.-L. (2001). *Identifying Labour Market Dynamics Using Labour Force Survey Data*. ZEW Discussion Paper No. 01-44. Mannheim.
- Assaad, R. (1997). The Employment Crisis in Egypt: Current Trends and Future Prospects. *Research in Middle East Economics*, 2(1), 39–66.
- Assaad, R., & El-Hamidi, F. (2009). Women in the Egyptian Labor Market: An Analysis of Developments, 1988-2006. In R. Assaad (Ed.), *The Egyptian Labor Market Revisited* (pp. 219–258). Cairo, Egypt: The American University in Cairo Press.
- Assaad, R., & Krafft, C. (2013). The Egypt Labor Market Panel Survey: Introducing the 2012 Round. *IZA Journal of Labor & Development*, 2(8), 1–30.
- Assaad, R., & Krafft, C. (2014). *Youth Transitions in Egypt: School, Work, and Family Formation in an Era of Changing Opportunities*. Silatech Working Paper No. 14-1. Doha, Qatar: Silatech.
- Assaad, R., & Krafft, C. (2015a). An Empirical Analysis of the Economics of Marriage in Egypt, Morocco, and Tunisia. In C. Monga & J. Y. Lin (Eds.), *The Oxford Handbook of Africa and Economics: Policies and Practices* (Vol. 2). Oxford, UK: Oxford University Press.
- Assaad, R., & Krafft, C. (2015b). The Economics of Marriage in North Africa: A Unifying Theoretical Framework. In C. Monga & J. Y. Lin (Eds.), *The Oxford Handbook of Africa and Economics: Contexts and Concepts* (Vol. 1). Oxford, UK: Oxford University Press.
- Assaad, R., & Krafft, C. (2015c). The Structure and Evolution of Employment in Egypt: 1998-2012. In R. Assaad & C. Krafft (Eds.), *The Egyptian Labor Market in an Era of Revolution* (pp. 27–51). Oxford, UK: Oxford University Press.
- Assaad, R., & Ramadan, M. (2008). *Did Housing Policy Reforms Curb the Delay in Marriage Among Young Men in Egypt? Middle East Youth Initiative Policy Outlook No. 1*. Washington, DC.
- Assaad, R., & Roushdy, R. (2009). Methodological Appendix 3: An Analysis of Sample Attrition in the Egypt Labor Market Panel Survey 2006. In R. Assaad (Ed.), *The Egyptian Labor Market Revisited* (pp. 303–316). Cairo, Egypt: American University in Cairo Press.
- Beckett, M., Da Vanzo, J., Sastry, N., Panis, C., & Peterson, C. (2001). The Quality of Retrospective Data An Examination of Long-Term Recall in a Developing Country. *The Journal of Human Resources*, 36(3), 593–625.
- Beegle, K., Carletto, C., & Himelein, K. (2012). Reliability of Recall in Agricultural Data. *Journal of Development Economics*, 98(1), 34–41.
- Beegle, K., De Weerd, J., Friedman, J., & Gibson, J. (2012). Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania. *Journal of Development Economics*, 98(1), 3–18.

- Black, D., Berger, M., & Scott, F. (2000). Bounding Parameter Estimates with Nonclassical Measurement Error. *Journal of the American Statistical Association*, 95(451), 739–748.
- Blossfeld, H.-P., Golsch, K., & Rohwer, G. (2007). *Event History Analysis With Stata*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement Error in Survey Data. *Handbook of Econometrics*, 5, 3705–3843.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2012). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. CRC Press.
- de Nicola, F., & Giné, X. (2014). How Accurate Are Recall Data? Evidence from Coastal India. *Journal of Development Economics*, 106, 52–65.
- Dercon, S., & Shapiro, J. S. (2007). *Moving On, Staying Behind, Getting Lost: Lessons on Poverty Mobility from Longitudinal Data: Global Poverty Research Group Working Paper Series No. 75*.
- Dhillon, N., & Yousef, T. (2009). *Generation in Waiting: The Unfulfilled Promise of Young People in the Middle East*. Washington, DC: Brookings Institution Press.
- Finn, A., & Ranchood, V. (2013). *Genuine Fakes: The Prevalence and Implications of Fieldworker Fraud in a Large South African Survey. Southern Africa Labour and Development Research Unit Working Paper Series No. 115*. Cape Town.
- Fuller, W. A. (1987). *Measurement Error Models*. New York, NY: John Wiley & Sons.
- Gibson, J., & Kim, B. (2010). Non-Classical Measurement Error in Long-Term Retrospective Recall Surveys. *Oxford Bulletin of Economics and Statistics*, 72(5), 687–695.
- Grosh, M., & Glewwe, P. (Eds.). (2000). *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study*. Washington, DC: The World Bank.
- Heitjan, D., & Rubin, D. B. (1990). Inference From Coarse Data Via Multiple Imputation With Application to Age Heaping. *Journal of the American Statistical Association*, 85(410), 304–314.
- Hiroyuki, N., Sawada, Y., & Mari, T. (2010). *Asking Retrospective Questions in Household Surveys: Evidence from Vietnam. RIETI Discussion Paper Series No. 10-E-008*.
- Judge, G., & Schechter, L. (2009). Detecting Problems in Survey Data Using Benford's Law. *Journal of Human Resources*, 44(1), 1–24.
- Kherfi, S. (2015). Determinants of Unemployment Duration in Egypt. In R. Assaad & C. Krafft (Eds.), *The Egyptian Labor Market in an Era of Revolution* (pp. 90–107). Oxford, UK: Oxford University Press.
- Langot, F., & Yassin, S. (2015). *Reforming Employment Protection in Egypt: An Evaluation Based on Transition Models with Measurement Errors. Economic Research Forum Working Paper Series No. 918*. Cairo, Egypt.
- Langsten, R., & Salem, R. (2008). Two Approaches to Measuring Women's Work in Developing Countries : A Comparison of Survey Data from Egypt. *Population and Development Review*, 34(2), 283–305.

- Magnac, T., & Visser, M. (1999). Transition Models with Measurement Errors. *Review of Economics and Statistics*, 81(3), 466–474.
- Puetz, D. (1993). Improving Data Quality in Household Surveys. In J. Von Braun & D. Puetz (Eds.), *Data Needs for Food Policy in Developing Countries: New Directions for Household Surveys* (pp. 173–185). Washington, DC: IFPRI.
- Roberts, J. M., & Brewer, D. D. (2001). Measures and Tests of Heaping in Discrete Quantitative Distributions. *Journal of Applied Statistics*, 28(7), 887–896.
- Salem, R. (n.d.). Imagined Crises: Assessing Evidence of Delayed Marriage and Never-Marriage in Contemporary Egypt. In K. Celello & H. Kholoussy (Eds.), *Domestic Tensions, National Anxieties: Global Perspectives on Marriage Crisis (Forthcoming)*. Oxford, UK: Oxford University Press.
- Salem, R. (2014). Trends and Differentials in Jordanian Marriage Behavior: Timing, Spousal Characteristics, Household Structure and Matrimonial Expenditures. In R. Assaad (Ed.), *The Jordanian Labour Market in the New Millennium* (pp. 189–217).
- Salem, R. (2015). Changes in the Institution of Marriage in Egypt from 1998 to 2012. In R. Assaad & C. Krafft (Eds.), *The Egyptian Labor Market in an Era of Revolution* (pp. 162–181). Oxford, UK: Oxford University Press.
- Singerman, D. (2007). *The Economic Imperatives of Marriage: Emerging Practices and Identities Among Youth in the Middle East. Middle East Youth Initiative Working Paper No. 6.*
- Singerman, D., & Ibrahim, B. (2003). The Costs of Marriage in Egypt: A Hidden Variable in the New Arab Demography. In N. S. Hopkins (Ed.), *Cairo Papers in Social Science, Volume 24 Numbers 1/2: The New Arab Family* (pp. 80–166). Cairo, Egypt: American University in Cairo Press.
- Yassin, S. (n.d.). *Constructing Labor Market Transitions Weights in Retrospective Data: An Application to Egypt and Jordan. Economic Research Forum Working Paper Series (Forthcoming)*. Cairo, Egypt.

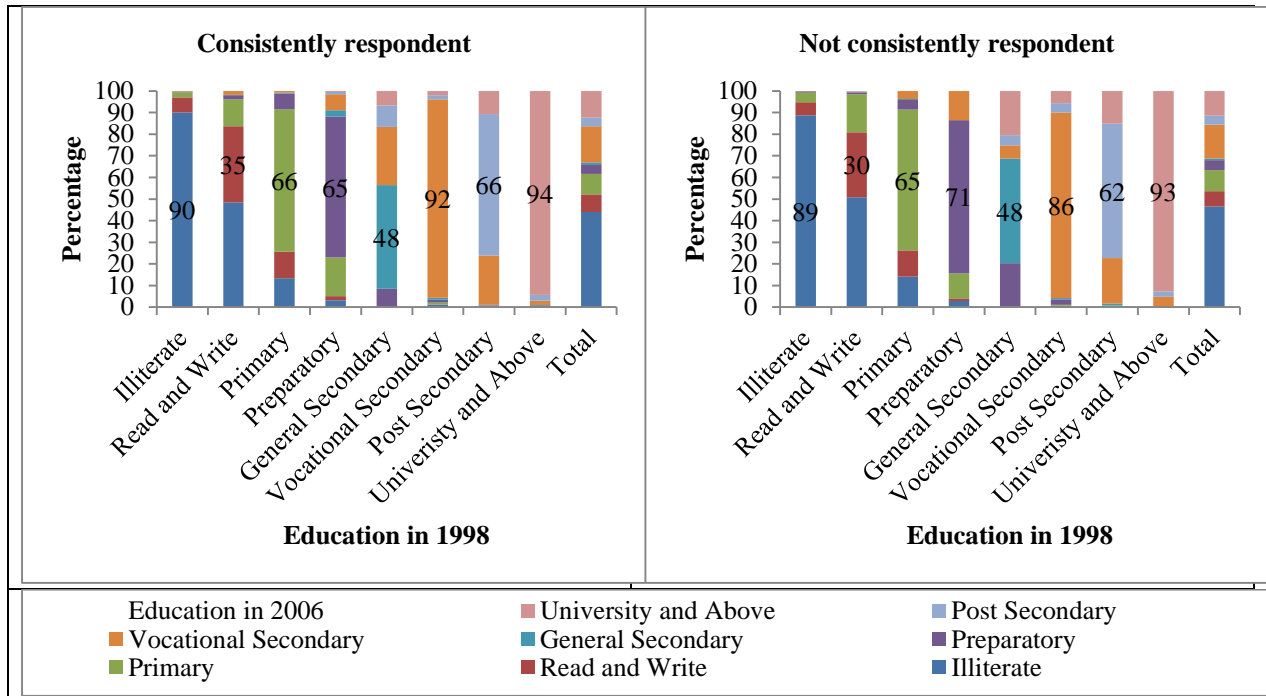
**Figure 1: Education (8 categories) as reported in 1998 vs. 2006, ages 30-54 in 1998**



Notes: See Table 4 for values underlying figure.

Source: Authors' calculations based on ELMPS 1998 and ELMPS 2006

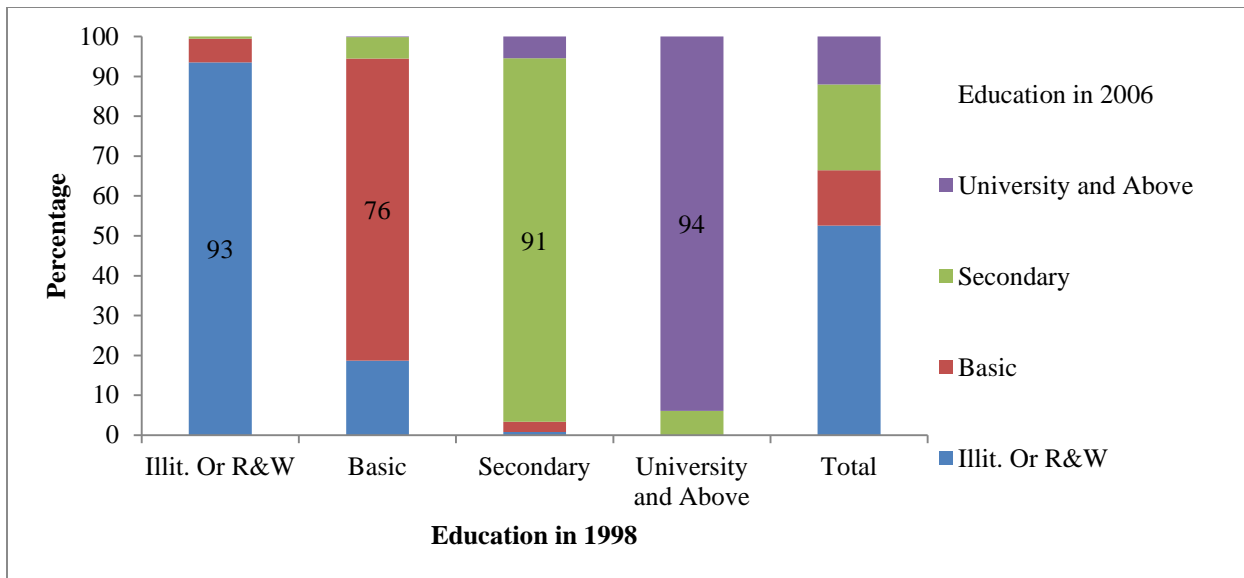
**Figure 2: Education (8 categories) as Reported in 1998 vs. 2006, by Respondent, Ages 30-54 in 1998**



Notes: See Table 5 for values underlying figure.

Source: Authors' calculations based on ELMPS 1998 and ELMPS 2006

**Figure 3: Education (4 categories) as Reported in 1998 vs. 2006, Ages 30-54 in 1998**

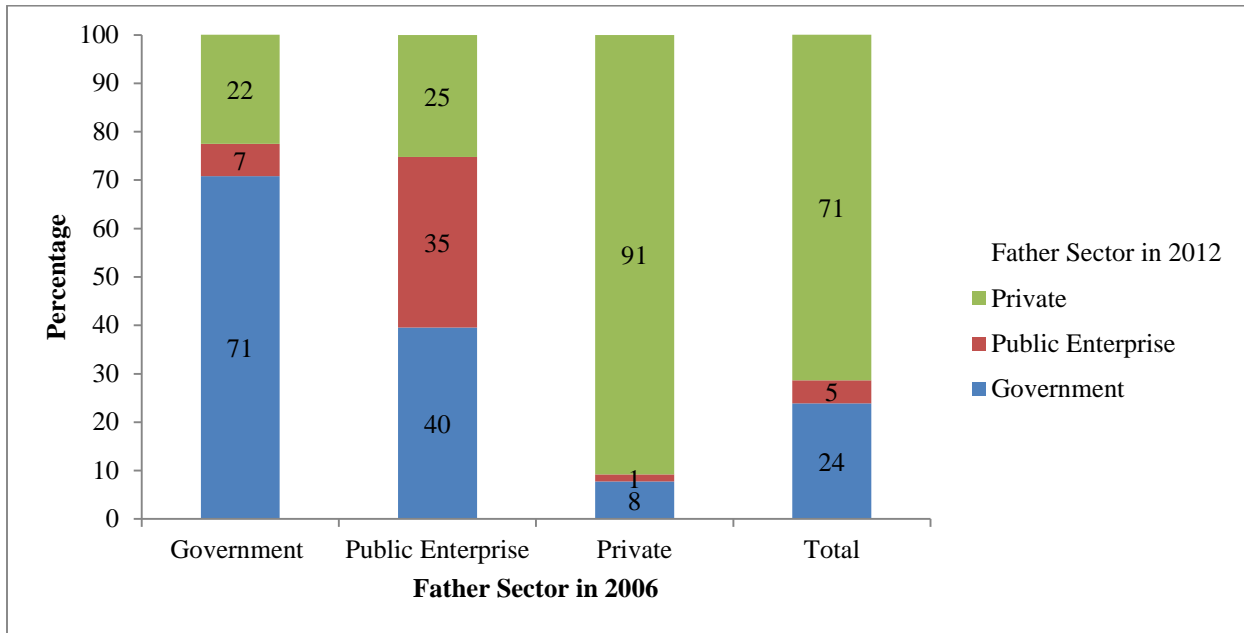


Notes: See Table 6 for values underlying figure.

Source: Authors' calculations based on ELMPS 1998 and ELMPS 2006



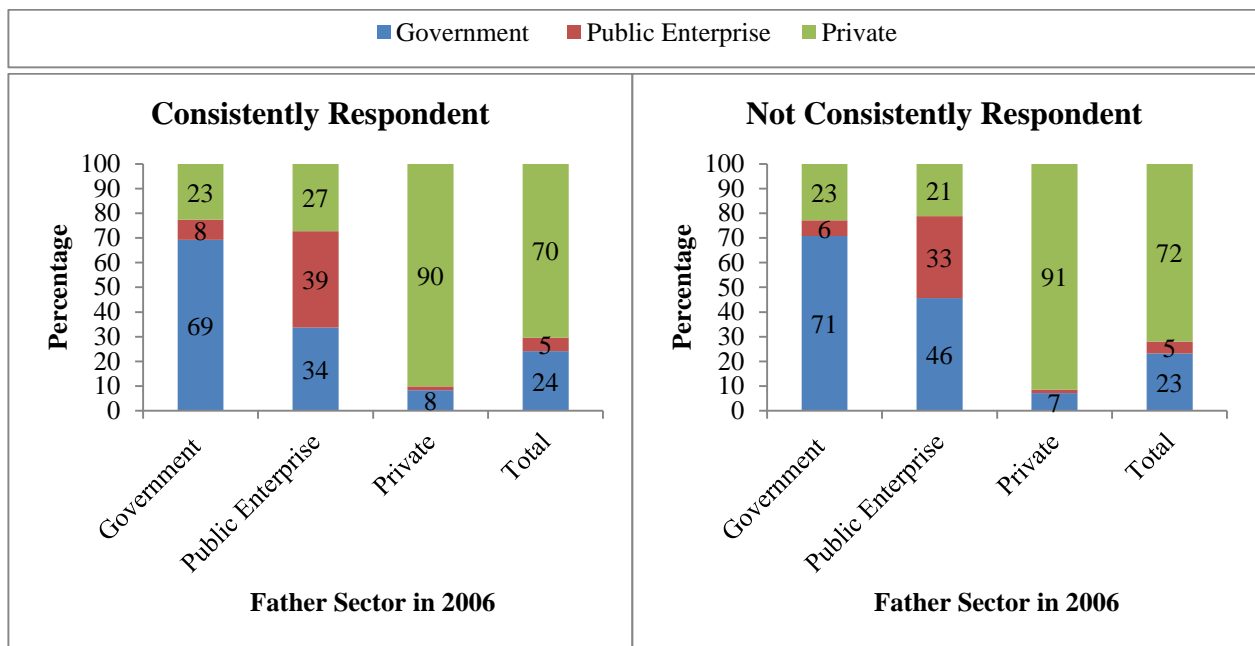
**Figure 4: Father’s Sector of Work When Age 15, as Reported in 2006 versus 2012, Father not in Household in 2006 or 2012, Age 30-54 in 2006**



Notes: See Table 7 for values underlying figure.

Source: Authors’ calculations based on ELMPS 2006 and ELMPS 2012

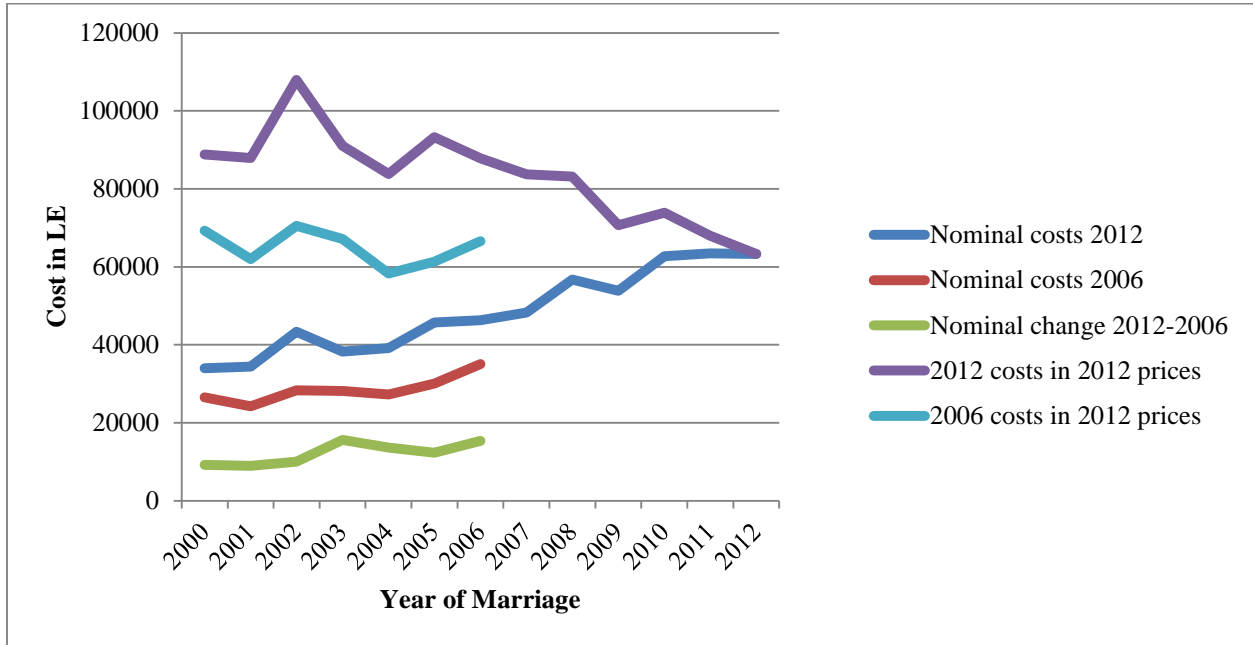
**Figure 5: Father’s Sector of Work When Individual Was Age 15, As Reported in 2006 versus 2012, by Respondent, Father Not in Household in 2006 or 2012, age 30-54 in 2006**



Notes: See Table 8 and Table 9 for values underlying figure.

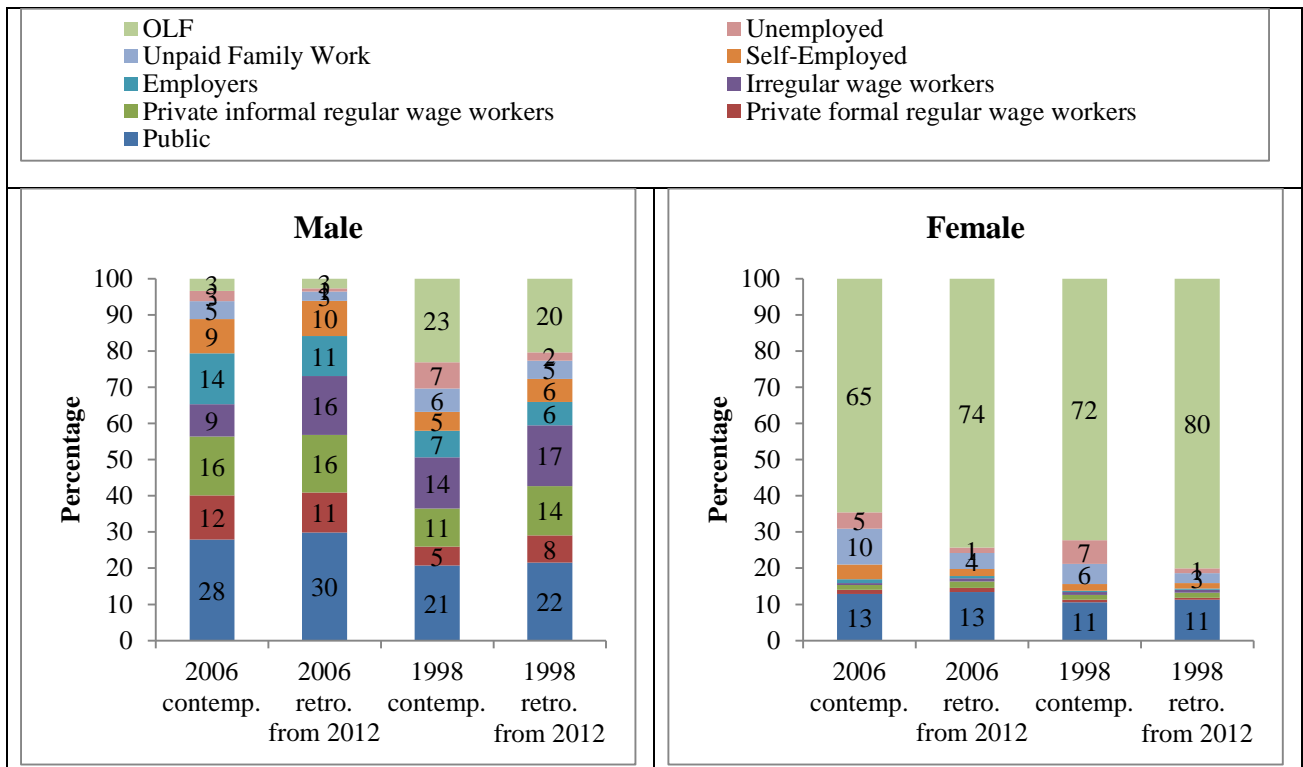
Source: Authors’ calculations based on ELMPS 2006 and ELMPS 2012

**Figure 6: Total Marriage Costs, As Reported in 2006 versus 2012, Individuals in Both Waves, Answering Marriage Section in 2012 (may not yet have been married in 2006)**



Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

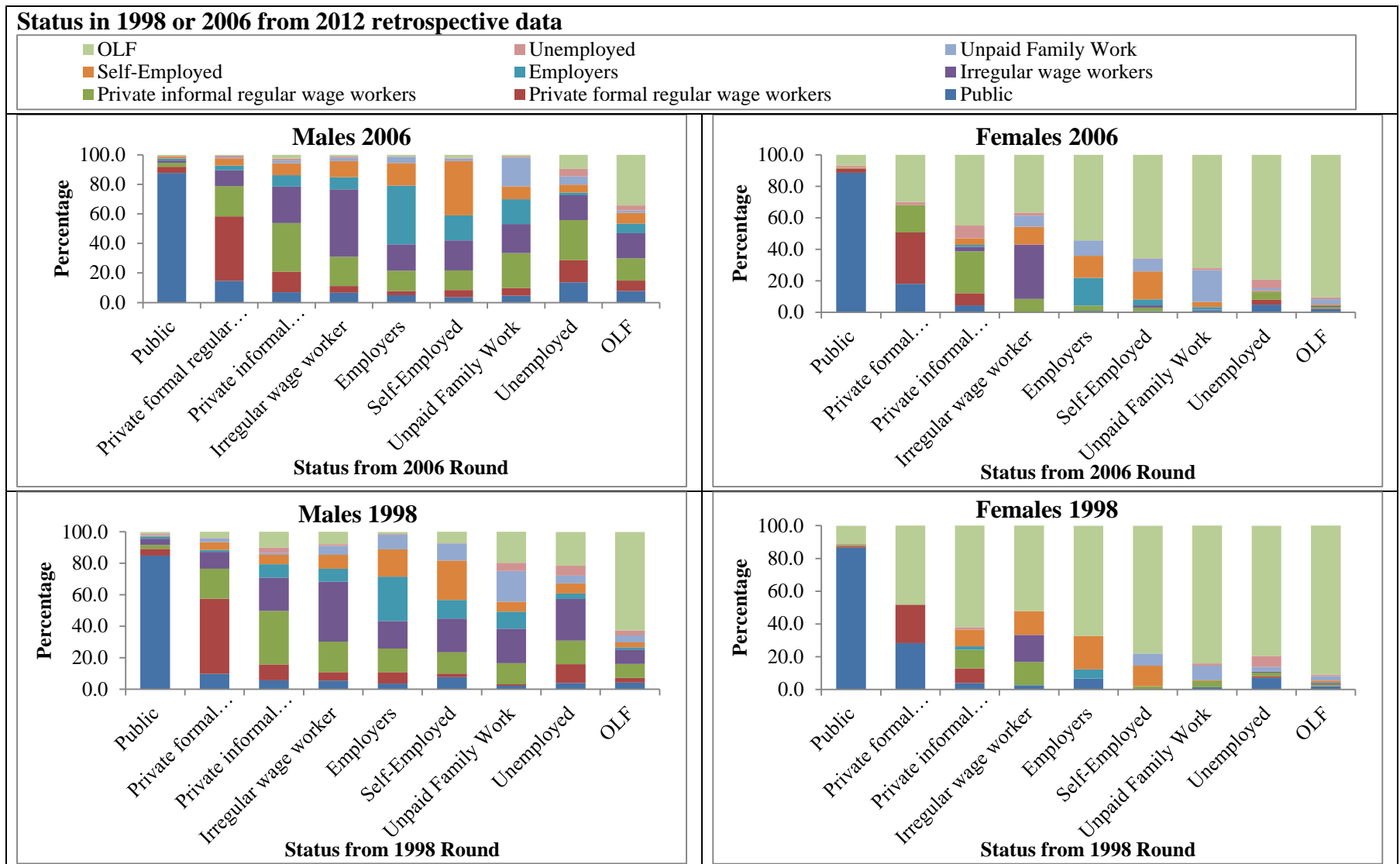
**Figure 7: Labor market Status, As Reported Contemporaneously For 1998 and 2006 and as Reported Retrospectively for Those Years from 2012 data, by sex, individuals ages 30-54 in 2012 present in both waves**



Notes: See Table 10 and Table 11 for values underlying figure.

Source: Authors' calculations based on ELMPS 1998, ELMPS 2006 and ELMPS 2012

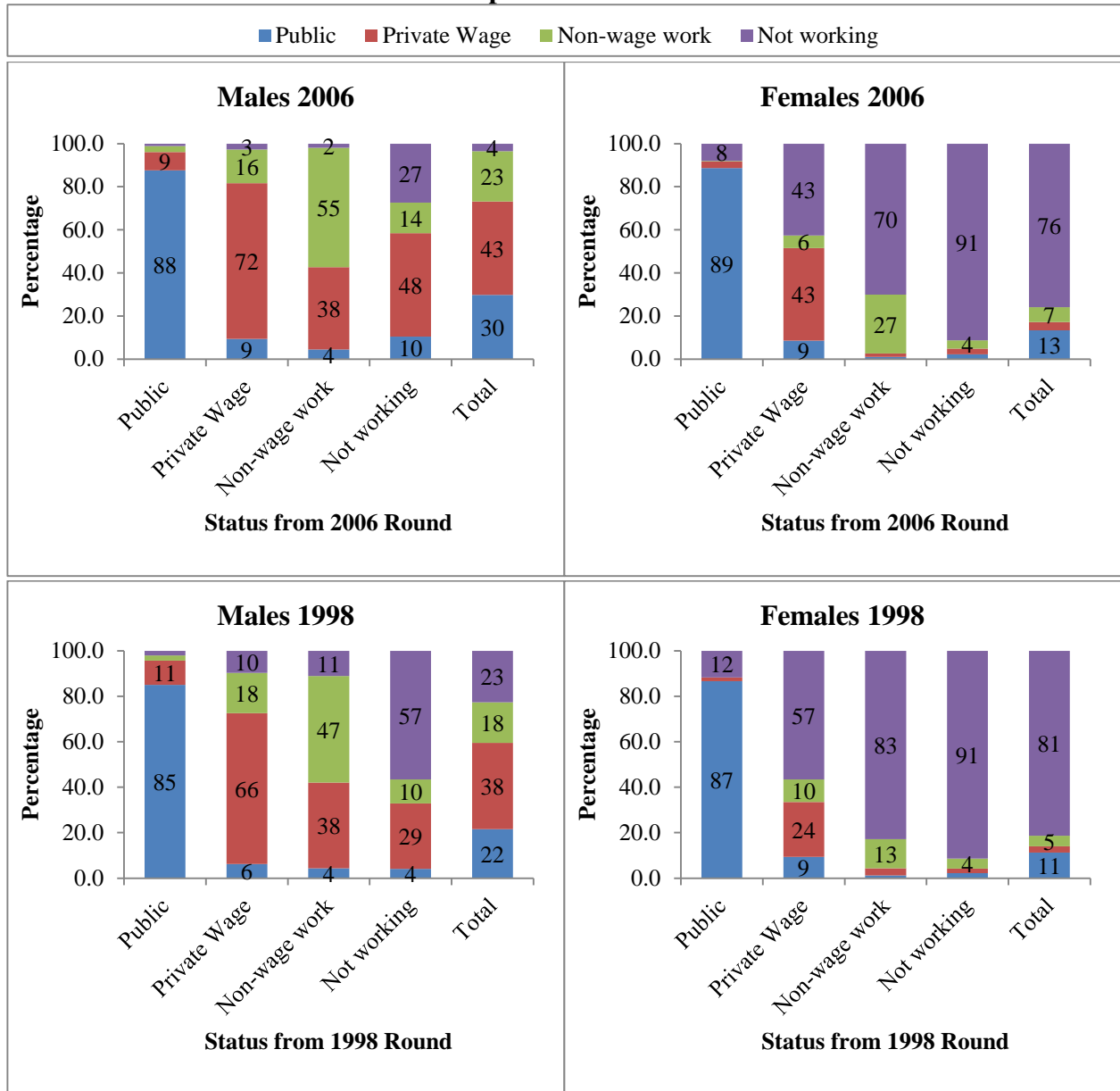
**Figure 8: Labor Market Status, As Reported in 1998 or 2006 versus 2012 Retrospective Data for 1998 or 2006, by Sex, Individuals Ages 30-54 in 2012 Present in Both Waves**



Notes: See Table 12, Table 13, Table 14, and Table 15 for values underlying figure.  
 Source: Authors' calculations based on ELMPs 1998, ELMPs 2006 and ELMPs 2012

**Figure 9: Collapsed Labor Market Status, As Reported in 1998 or 2006 Versus 2012 Retrospective Data for 1998 Or 2006, by Sex, Individuals Ages 30-54 In 2012 Present in Both Waves**

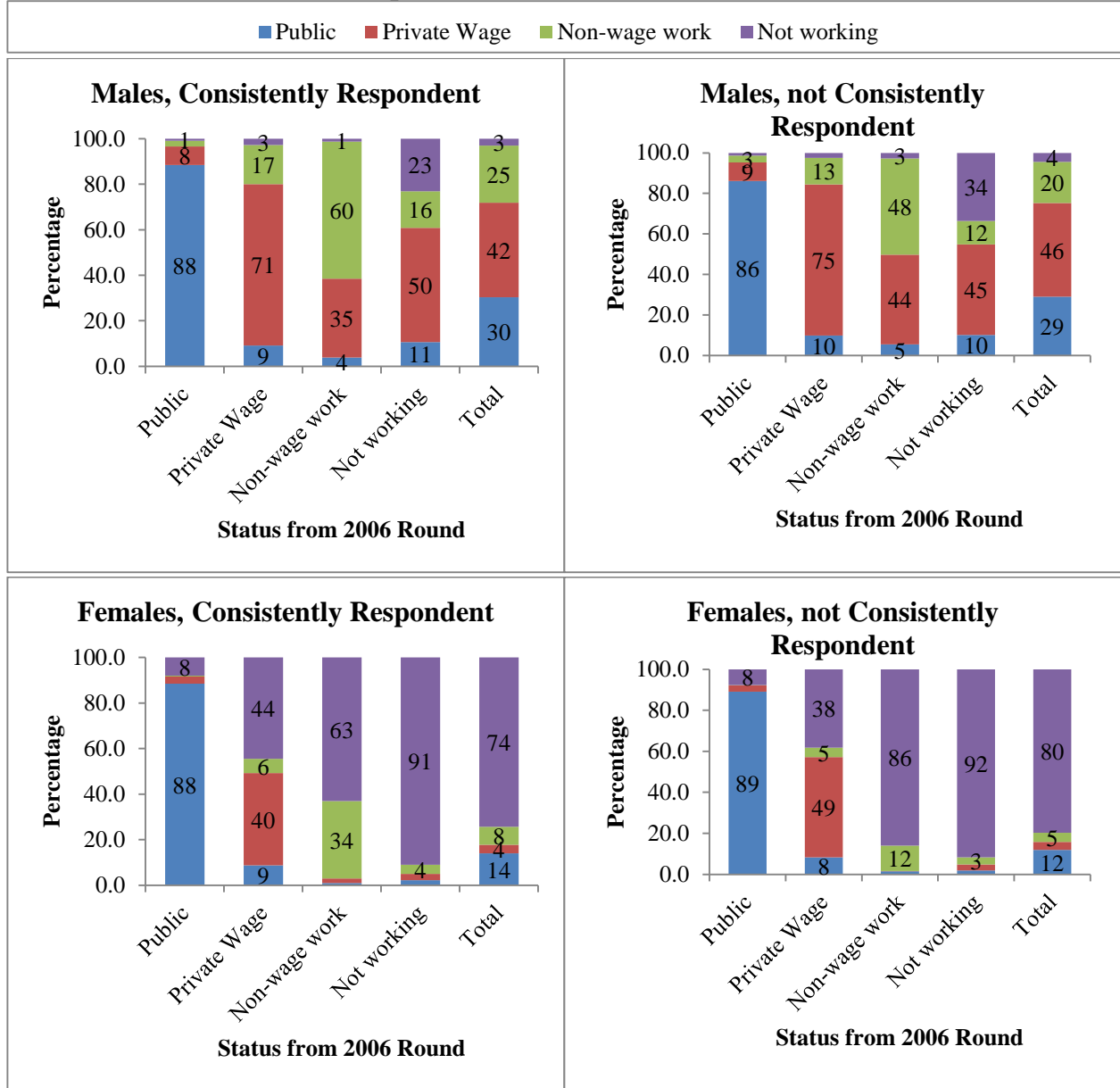
**Status in 1998 or 2006 from 2012 retrospective data**



Notes: See Table 16, Table 17, Table 18, Table 19 for values underlying figure.  
 Source: Authors' calculations based on ELMPs 1998, ELMPs 2006 and ELMPs 2012

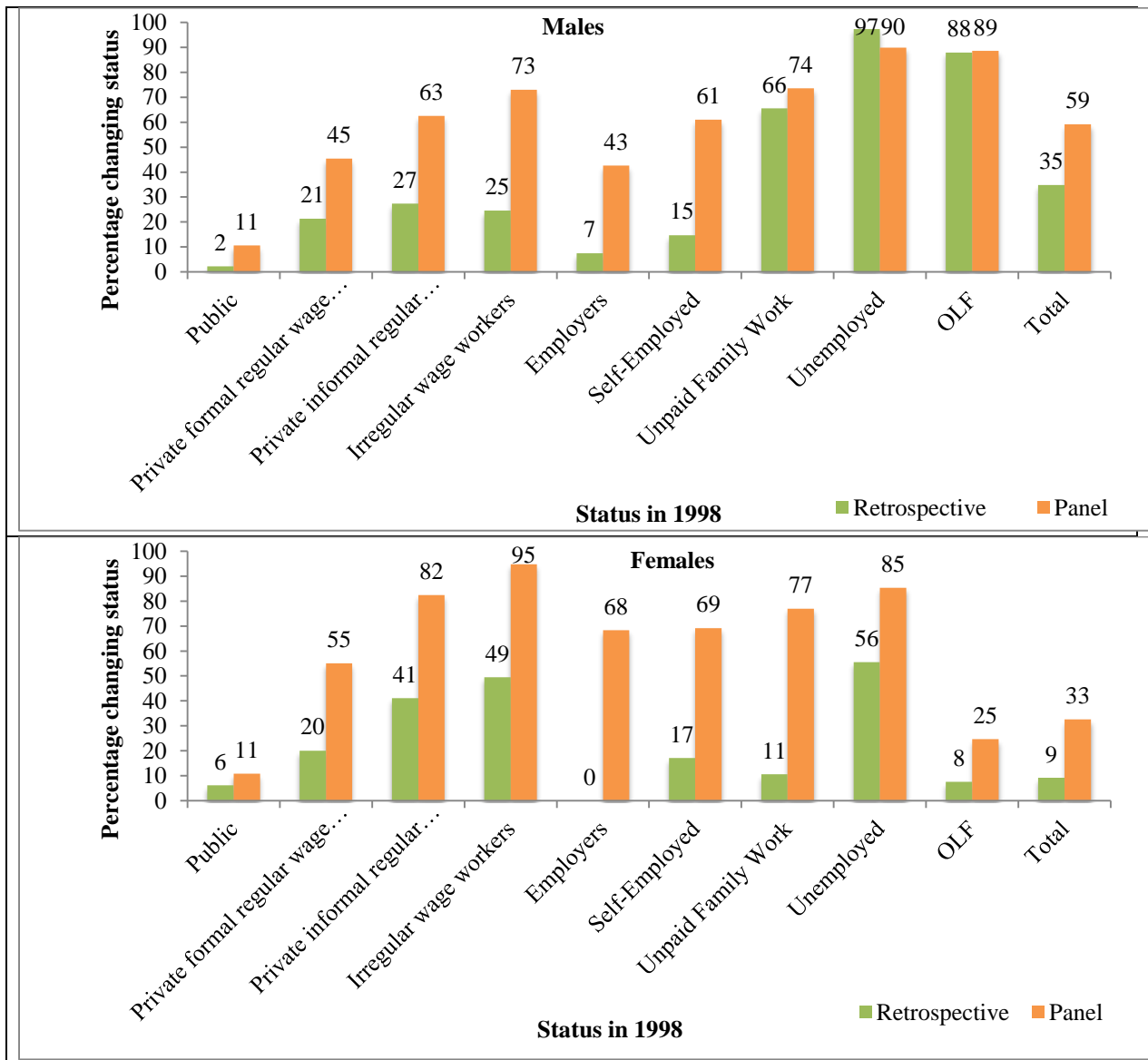
**Figure 10: Collapsed Labor Market Status, as Reported in 2006 versus 2012 Retrospective Data for 2006, by Sex and Respondent, Individuals Ages 30-54 in 2012 Present in Both Waves**

**Status in 2006 from 2012 retrospective data**



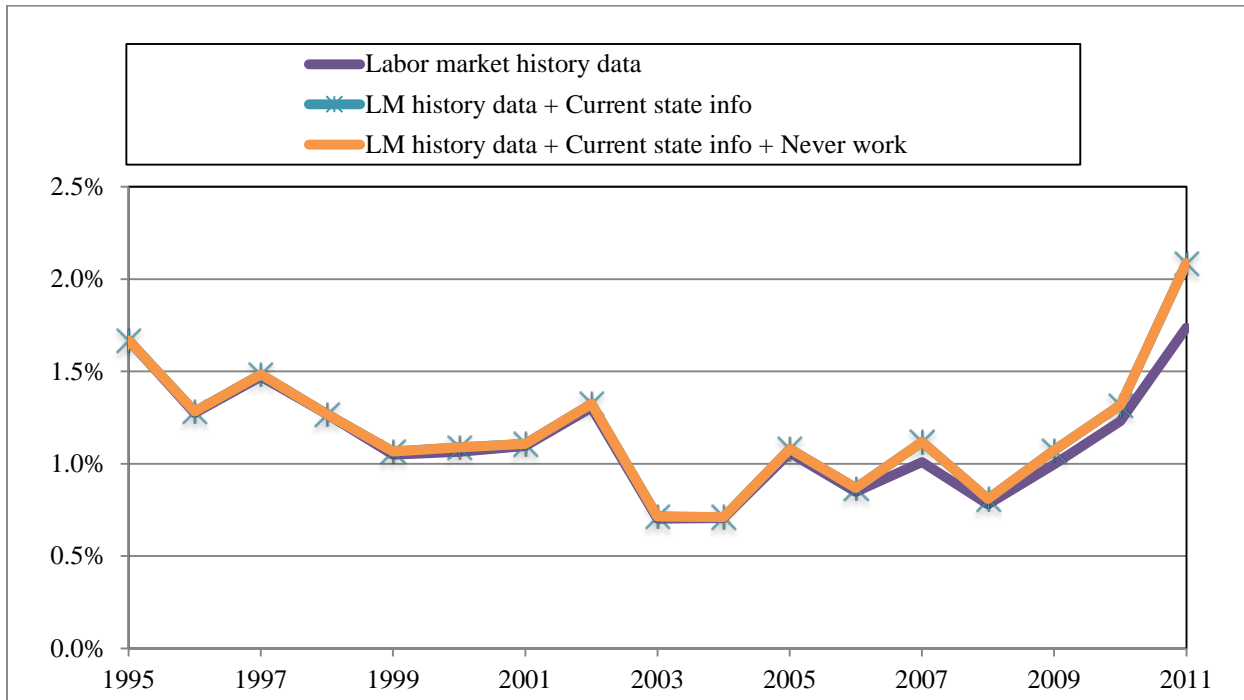
Notes: See Table 16, Table 17, Table 18, Table 19 for values underlying figure.  
 Source: Authors' calculations based on ELMPS 1998, ELMPS 2006 and ELMPS 2012

**Figure 11: Rates of Status Change in Panel Data for 1998 To 2006 Versus Rates of Status Change in Retrospective Data from 2012 For Changes from 1998 To 2006 by Sex and Status in 1998**



Notes: Based only on individuals in all three waves. Status in 1998 is from either retrospective or panel data depending on whether transition rates are being examined for retrospective or panel data. See Table 20 for values underlying figure.  
 Source: Authors' calculations based on ELMPS 1998, ELMPS 2006 and ELMPS 2012

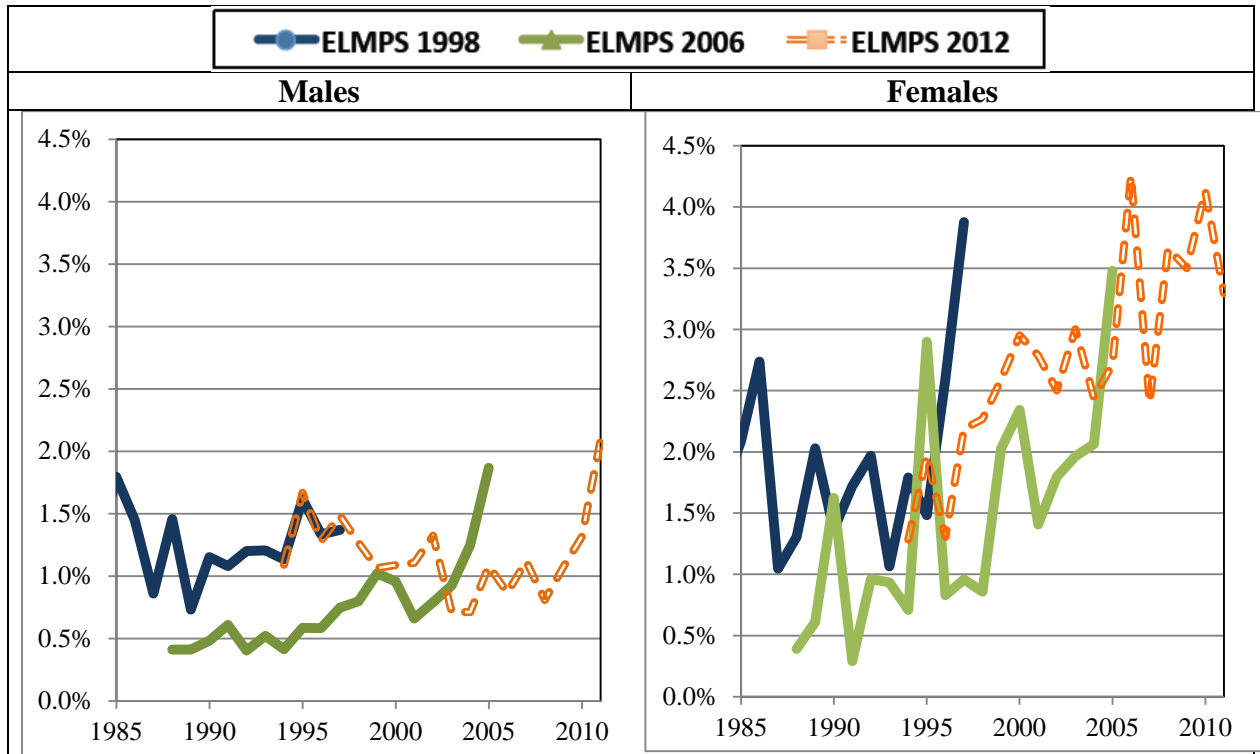
**Figure 12: Comparing Retrospective Panels for Employment to Non-Employment Separation Rates, Males 15-54 Years of Age, 1995-2011**



Source: Authors' calculations based on ELMPS 2012

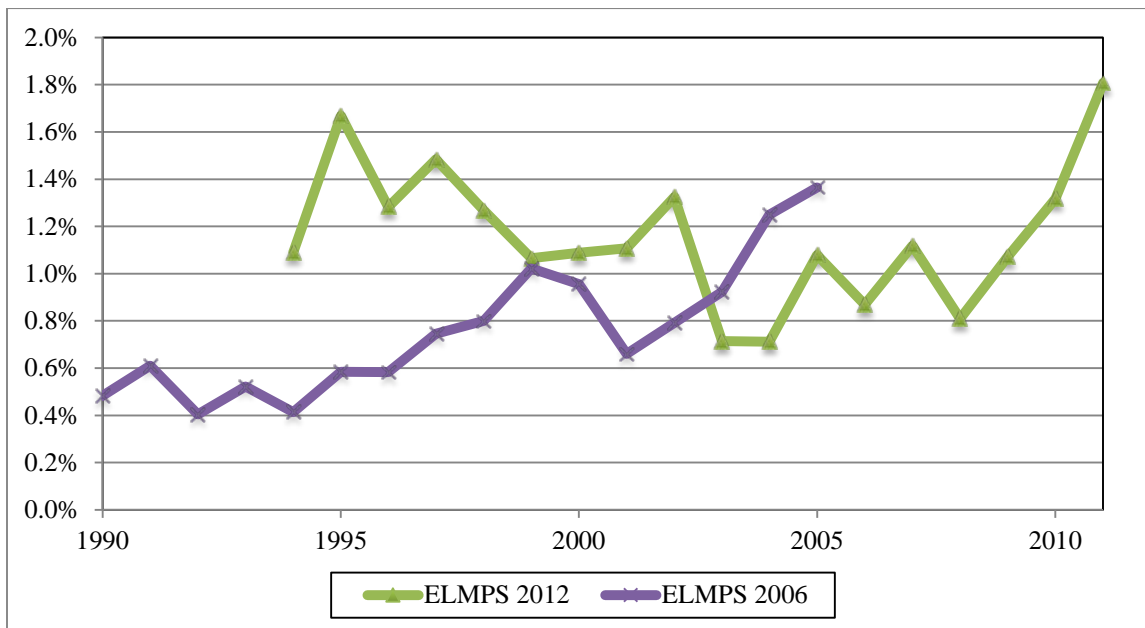


**Figure 13: Employment to Non-Employment Separation Rates by Sex and Wave, Augmented Panel, 15-54 Years of Age, 1985-2011**



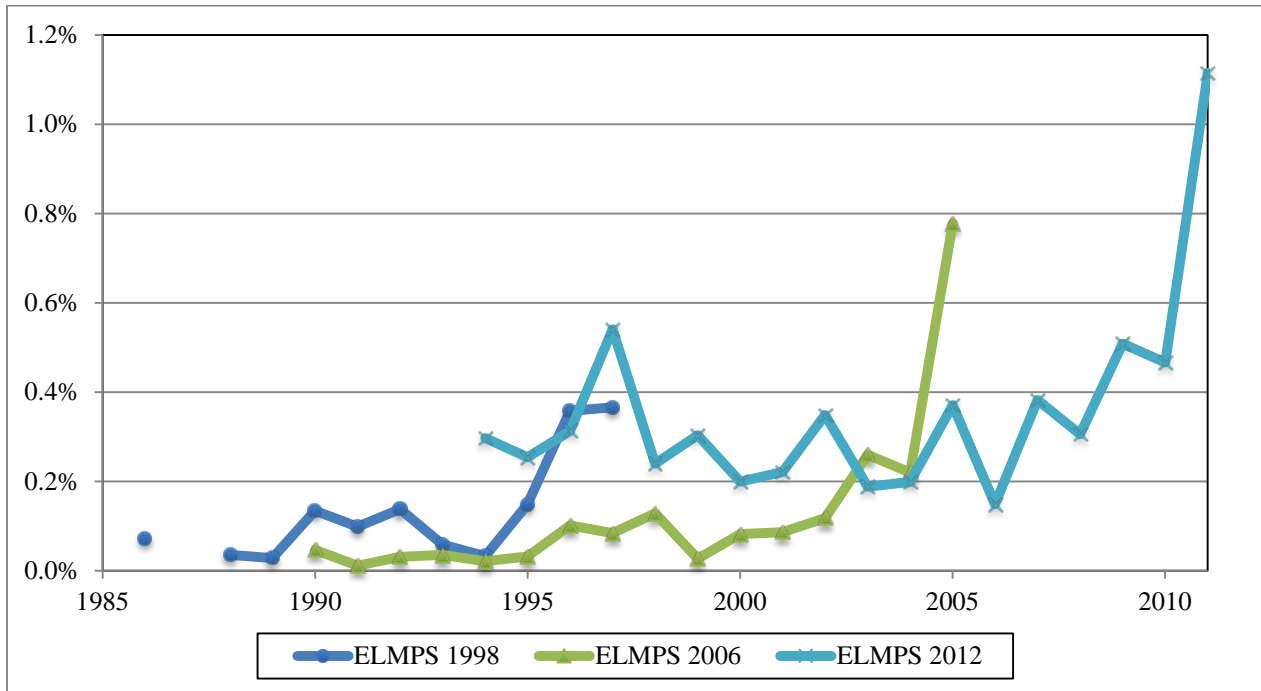
Source: Author's calculations based on ELMS 1998, ELMPS 2006, and ELMPS 2012

**Figure 14: Employment to Non-Employment Separation Rates by Wave, Males Ages 15-54, Using Augmented Panels With Unemployment Spells Greater Than or Equal to Six Months**



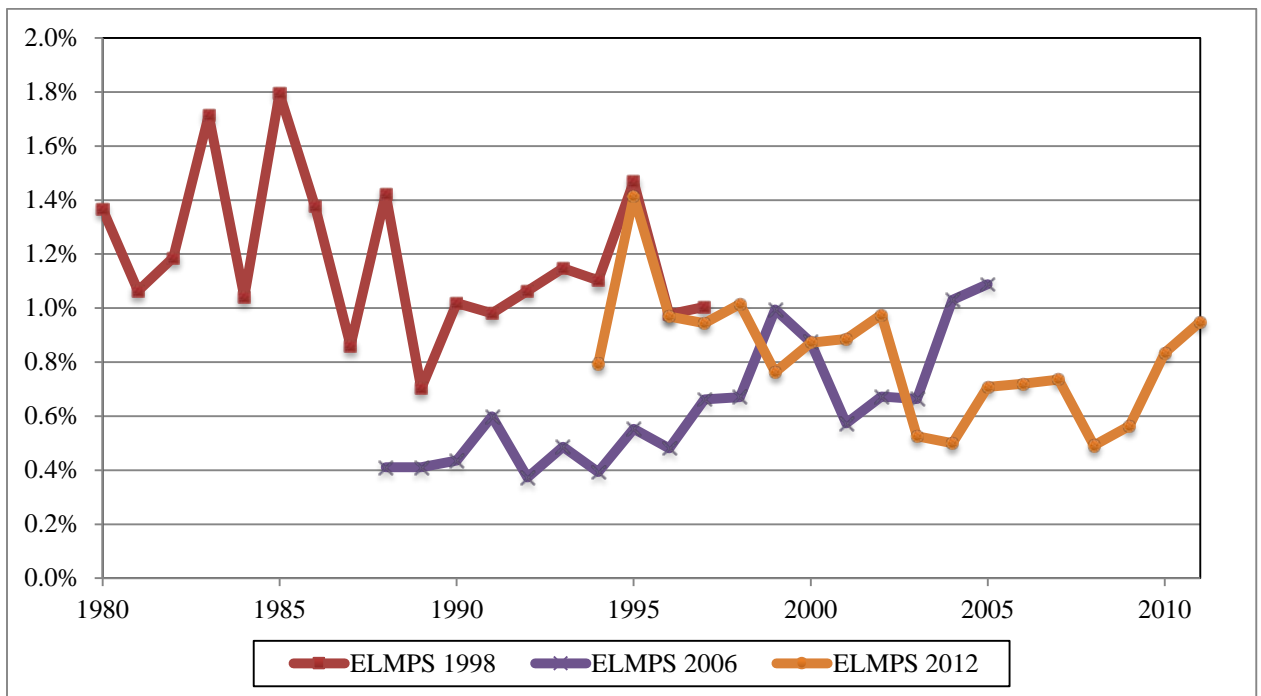
Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Figure 15: Employment to Unemployment Separation Rates by Wave, Males Ages 15-54, Using Augmented Panel Data**



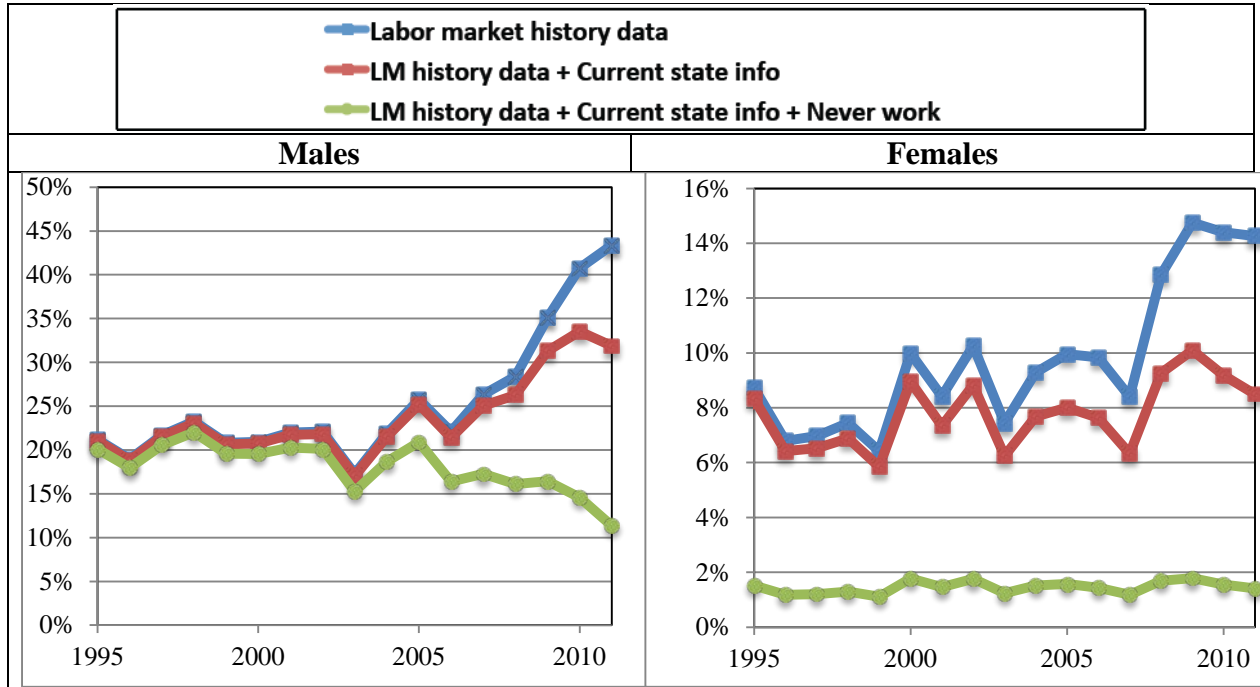
Source: Author's calculations based on ELMS 1998, ELMPS 2006, and ELMPS 2012

**Figure 16: Employment to Inactivity Separation Rates by Wave, Males Ages 15-54, Using Augmented Panel Data**



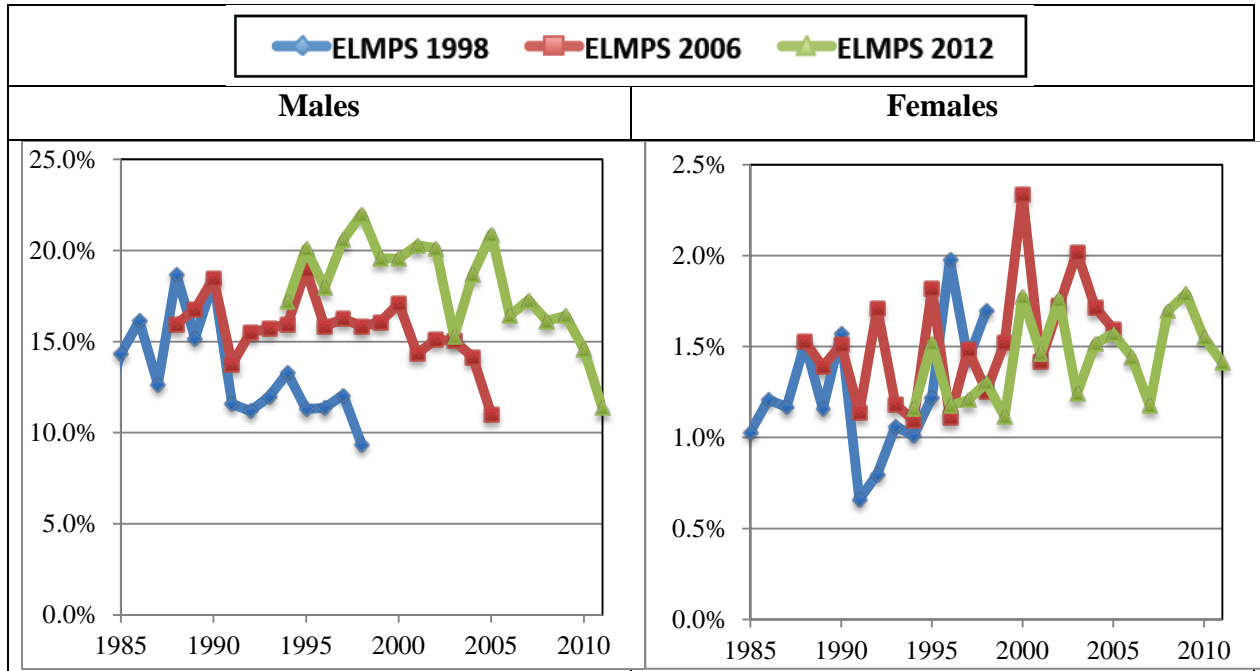
Source: Author's calculations based on ELMS 1998, ELMPS 2006, and ELMPS 2012

**Figure 17: Comparing Retrospective Panels for Non-Employment To Employment Job Finding Rates by Sex, Ages 15-54, 1995-2011**



Source: Author's calculations based on ELMPS 2012

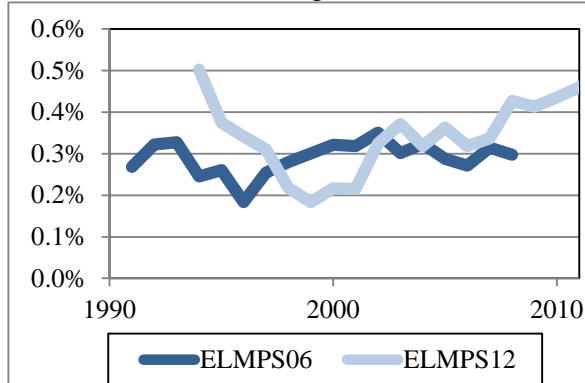
**Figure 1: Non-Employment to Employment Job Finding Rates by Sex and Wave, Ages 15-54, Using Augmented Panel Data and Incorporating Those Who Never Worked**



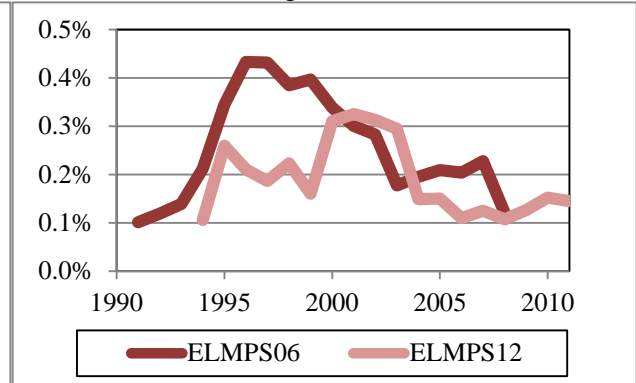
Source: Author's calculations based on ELMS 1998, ELMPS 2006, and ELMPS 2012

**Figure 19: Job-to-Job Transitions by Wave, Male Workers, Ages 15-54 Calculated from Retrospective Panels**

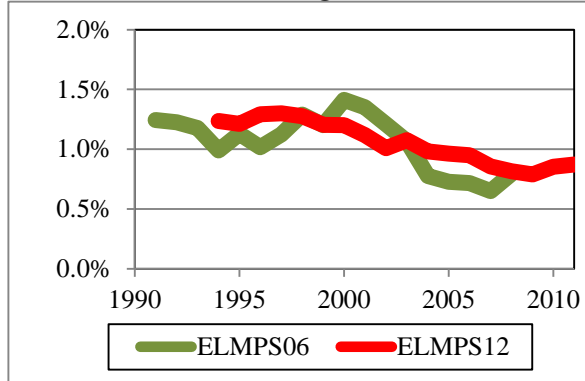
(a) Public to Private wage work



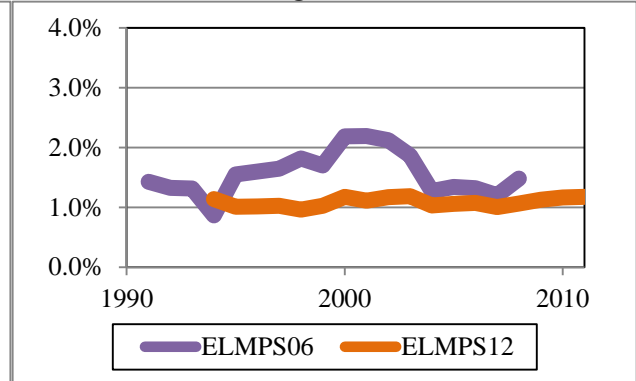
(b) Public to Non-wage work



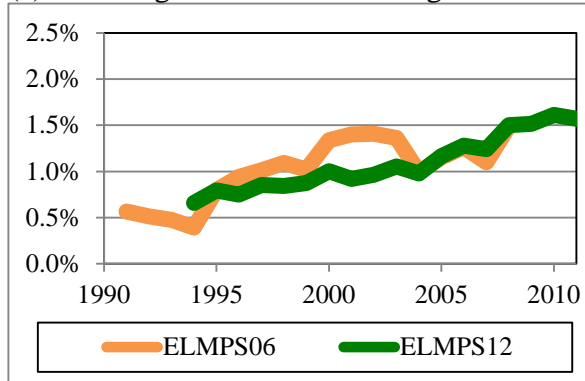
(c) Private to Public wage work



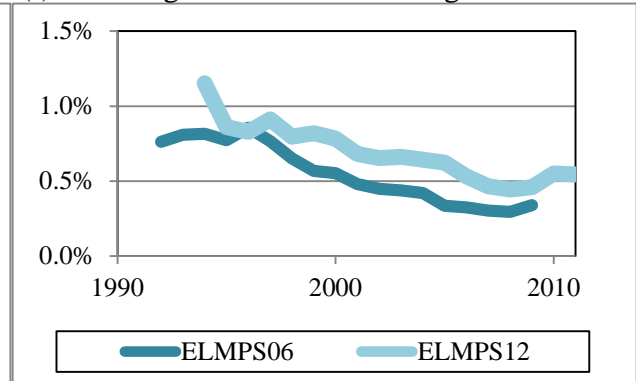
(d) Private to Non-wage work



(e) Non-wage work to Public wage work

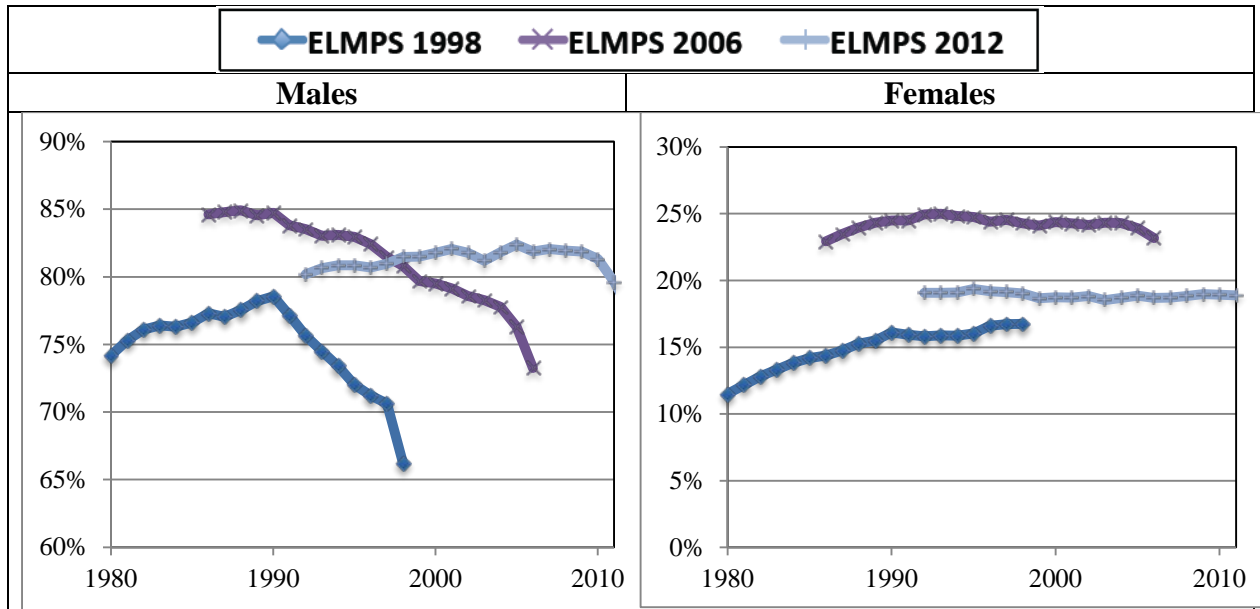


(f) Non-wage work to Private wage work



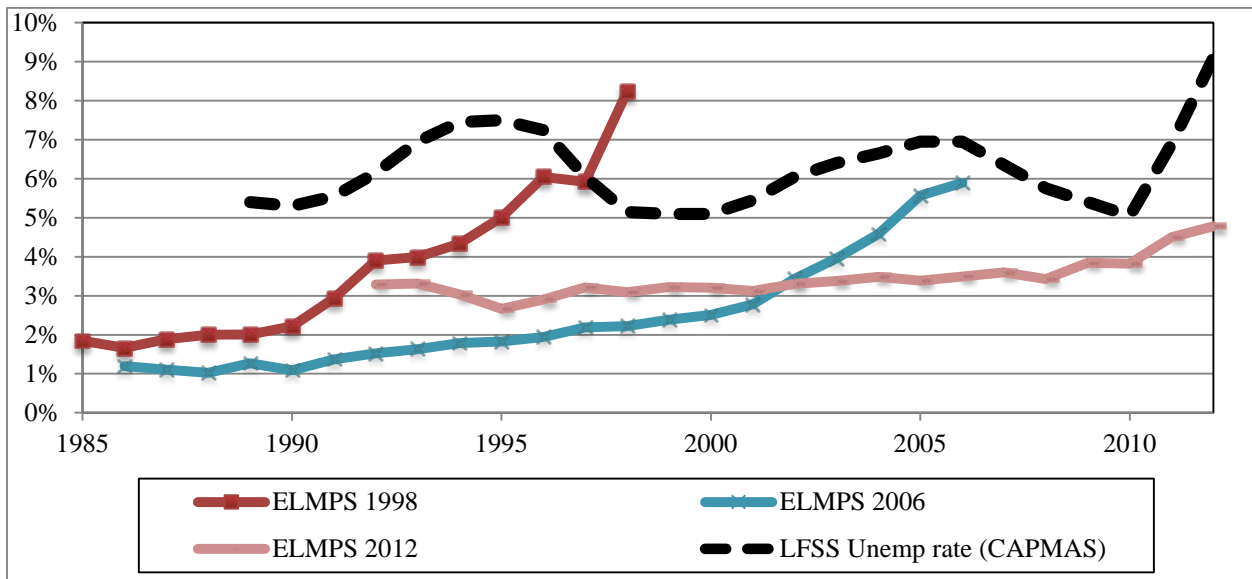
Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Figure 20: Employment-to-Population Ratios by Sex and Wave, Ages 15-54, Calculated from Augmented Panels Including the Never Worked**



Source: Authors' calculations based on ELMS 1998, ELMPS 2006, and ELMPS 2012

**Figure 21: Unemployment Rates, Males, Ages 15-64, Calculated from Augmented Retrospective Panels Including the Never Worked Compared to Those Reported in the LFSS**



Source: Authors' calculations based on ELMS 1998, ELMPS 2006, ELMPS 2012, and LFSS data<sup>7</sup>

<sup>7</sup> Based on CAPMAS's bulletin of the Labor Force Sample Survey for 1989-2011

**Table 1: Patterns of Unemployment Reporting as Reported in 1998 or 2006 versus 2012 Retrospective Data for 1998 or 2006, Individuals Reporting Contemporaneous Unemployment in 2006 or 1998 and Present in 2012**

Comparison to retrospective data	Dist. 1998	Mean current unemp. dur. mos. 1998	% less than six months 1998	Dist. 2006	Mean current unemp. dur. mos. 2006	% less than six months 2006
Aligned	9	23	26	13	16	31
Unemployed within one year +/-	1	0	22	5	5	24
Unemployed within two-five years +/-	7	12	33	12	9	38
Unemployed more than five years +/-	11	8	31	7	7	46
Never unemployed but have a fourth status	6	15	35	4	27	31
Never unemployed no fourth status	65	22	41	59	22	29
Total	100	19	37	100	18	32
N	261	261	261	443	443	443

Source: Authors' calculations based on ELMPS 1998, ELMPS 2006 and ELMPS 2012

**Table 2: Patterns of Unemployment Reporting as Reported in 2006 versus 2012 Retrospective Data for 2006, by Gender and Whether Consistently Respondent, Individuals Reporting Contemporaneous Unemployment in 2006 and present in 2012**

	Male			Female			Total		
	Not consist. resp.	consist. resp.	Total	Not consist. resp.	consist. resp.	Total	Not consist. resp.	consist. resp.	Total
Aligned	8	7	8	29	25	26	13	13	13
Unemployed within one year +/-	4	5	4	4	6	6	4	5	5
Unemployed within two-five years +/-	10	14	13	4	12	10	9	13	12
Unemployed more than five years +/-	7	7	7	20	5	9	10	6	7
Never unemployed but have a fourth status	4	4	4	3	4	4	3	4	4
Never unemployed no fourth status	67	63	65	40	48	46	61	59	59
Total	100	100	100	100	100	100	100	100	100
N (Obs.)	120	194	314	35	94	129	155	288	443

Source: Authors' calculations based on ELMPS 1998, ELMPS 2006 and ELMPS 2012

**Table 3: Probit Model Marginal Effects for the Probability of Alignment of Reporting between Contemporaneous 1998 or 2006 and 2012 Retrospective Data by Sex, Individuals in 2006 or 1998 and Present in 2012, ages 30-54 in 2012**

	1998 Male	1998 Female	2006 Male	2006 Female
<b>Reference Case Probability:</b>	0.927	0.918	0.811	0.796
<b>Own Education (Univ. omitted)</b>				
Illit. or R&W	-0.185*** (0.042)	0.017 (0.028)	-0.011 (0.028)	0.051* (0.022)
Basic	-0.225*** (0.040)	0.021 (0.032)	-0.021 (0.028)	0.049 (0.027)
Secondary	-0.196*** (0.030)	-0.047 (0.026)	-0.026 (0.023)	0.028 (0.021)
<b>Age group in 2012 (30-34 omit.)</b>				
35-39	-0.132*** (0.033)	-0.042* (0.018)	-0.012 (0.021)	-0.002 (0.017)
40-44	-0.136*** (0.041)	-0.067** (0.022)	-0.008 (0.024)	0.017 (0.017)
45-49	-0.049 (0.042)	-0.041 (0.022)	0.000 (0.026)	0.033* (0.016)
50	-0.021 (0.044)	-0.037 (0.021)	0.031 (0.027)	0.042* (0.018)
<b>Region (Gr. Cairo omitted)</b>				
Alex. and Suez Canal	0.027 (0.043)	0.032 (0.027)	-0.003 (0.032)	0.022 (0.024)
Urban Lower	0.067 (0.040)	0.052 (0.027)	0.046 (0.029)	0.037 (0.022)
Urban Upper	0.057 (0.038)	0.030 (0.024)	0.071* (0.029)	0.028 (0.021)
Rural Lower	0.008 (0.039)	0.030 (0.026)	0.025 (0.026)	0.007 (0.020)
Rural Upper	0.066 (0.041)	0.038 (0.025)	0.058* (0.028)	0.012 (0.021)
<b>Consist. Respondent (Not consist. omit.)</b>				
Consist. resp.	-0.011 (0.027)	0.001 (0.018)	0.005 (0.016)	0.018 (0.012)
<b>Panel (1998 or 2006) employment status (public wage omit.)</b>				
Private wage	-0.105** (0.037)	-0.604*** (0.088)	-0.137*** (0.025)	-0.459*** (0.052)
Non-wage	-0.305*** (0.042)	-0.718*** (0.052)	-0.290*** (0.025)	-0.663*** (0.033)
Not working	-0.276*** (0.041)	0.005 (0.036)	-0.547*** (0.037)	-0.017 (0.026)
<b>2012 Employment chars.</b>				
Not employed in 2012	-0.054 (0.061)	0.224*** (0.041)	-0.202*** (0.044)	0.005 (0.024)
Irregular in 2012	-0.070 (0.036)	-0.127 (0.100)	-0.082*** (0.021)	-0.269** (0.084)
Informal in 2012	-0.010 (0.029)	-0.111*** (0.030)	0.008 (0.021)	0.052 (0.043)
<b>N(Obs.)</b>	2408	2465	4540	4656

Source: Authors' calculations based on ELMPS 1998, ELMPS 2006 and ELMPS 2012

**Table 4: Education (8 Categories) as Reported in 1998 vs. 2006, Ages 30-54 in 1998**

<b>Response in 2006</b>									
<b>Response in 1998</b>		<b>Illiterate</b>	<b>Read and Write</b>	<b>Primary</b>	<b>Preparatory</b>	<b>General Secondary</b>	<b>Vocational Secondary</b>	<b>Post Secondary</b>	<b>University and Above</b>
	Illiterate	89.6	6.7	3.0	0.4	0.0	0.3	0.0	0.0
	Read and Write	49.1	33.7	13.9	1.6	0.2	1.4	0.1	0.1
	Primary	13.6	12.2	65.7	6.5	0.5	1.6	0.0	0.0
	Preparatory	3.1	1.7	16.1	66.8	2.1	9.2	0.7	0.4
	General Secondary	0.0	0.0	0.0	12.4	48.1	20.0	8.3	11.2
	Vocational Secondary	0.9	0.2	0.7	1.7	0.8	90.1	2.6	3.0
	Post Secondary	0.0	0.2	0.7	0.3	0.2	22.2	64.5	12.0
	University and Above	0.0	0.0	0.0	0.0	0.7	2.7	2.7	93.9
<b>Total</b>	<b>44.7</b>	<b>7.8</b>	<b>9.5</b>	<b>4.4</b>	<b>0.9</b>	<b>16.5</b>	<b>4.1</b>	<b>12.1</b>	

Source: Authors' calculations based on ELMS 1998 and ELMPS 2006

**Table 5: Education (8 categories) as Reported in 1998 vs. 2006, by Respondent, Ages 30-54 in 1998**

<b>Consistently respondent</b>									
<b>Response in 2006</b>									
<b>Response in 1998</b>		<b>Illiterate</b>	<b>Read and Write</b>	<b>Primary</b>	<b>Preparatory</b>	<b>General Secondary</b>	<b>Vocational Secondary</b>	<b>Post Secondary</b>	<b>University and Above</b>
	Illiterate	90.0	7.0	2.4	0.3	0.0	0.4	0.0	0.0
	Read and Write	48.5	35.2	12.5	1.9	0.2	1.8	0.0	0.0
	Primary	13.4	12.2	65.9	7.2	0.5	0.8	0.0	0.0
	Preparatory	3.2	1.9	17.9	65.1	3.0	7.4	1.0	0.6
	General Secondary	0.0	0.0	0.0	8.4	48.0	26.9	10.1	6.6
	Vocational Secondary	1.0	0.2	0.9	1.4	0.9	91.8	1.9	1.9
	Post Secondary	0.0	0.3	0.6	0.4	0.0	22.6	65.5	10.7
	University and Above	0.0	0.0	0.0	0.0	0.9	2.0	2.7	94.3
<b>Total</b>	<b>44.0</b>	<b>8.1</b>	<b>9.5</b>	<b>4.2</b>	<b>0.9</b>	<b>16.9</b>	<b>4.1</b>	<b>12.3</b>	
<b>Not consistently respondent</b>									
<b>Response in 2006</b>									
<b>Response in 1998</b>		<b>Illiterate</b>	<b>Read and Write</b>	<b>Primary</b>	<b>Preparatory</b>	<b>General Secondary</b>	<b>Vocational Secondary</b>	<b>Post Secondary</b>	<b>University and Above</b>
	Illiterate	88.6	6.2	4.4	0.6	0.0	0.2	0.0	0.0
	Read and Write	50.8	30.0	17.7	0.8	0.0	0.2	0.4	0.2
	Primary	14.1	12.1	65.1	4.5	0.4	3.8	0.0	0.0
	Preparatory	2.8	1.2	11.6	70.9	0.0	13.5	0.0	0.0
	General Secondary	0.0	0.0	0.0	20.3	48.4	6.1	4.7	20.5
	Vocational Secondary	0.5	0.2	0.3	2.5	0.7	85.7	4.4	5.7
	Post Secondary	0.0	0.0	0.8	0.0	0.8	21.1	62.1	15.2
	University and Above	0.0	0.0	0.0	0.0	0.0	4.8	2.6	92.7
<b>Total</b>	<b>46.6</b>	<b>7.1</b>	<b>9.7</b>	<b>4.7</b>	<b>0.8</b>	<b>15.7</b>	<b>4.2</b>	<b>11.4</b>	

Source: Authors' calculations based on ELMS 1998 and ELMPS 2006



**Table 6: Education (4 categories) as Reported in 1998 vs. 2006, Ages 30-54 in 1998**

<u>Response in 2006</u>		<u>Illit. Or R&amp;W</u>	<u>Basic</u>	<u>Secondary</u>	<u>University and Above</u>
<u>Response in 1998</u>	Illit. Or R&W	93.5	5.9	0.6	0.0
	Basic	18.7	75.8	5.4	0.1
	Secondary	0.8	2.5	91.2	5.5
	University and Above	0.0	0.0	6.1	93.9
	Total	52.6	13.9	21.5	12.1

Source: Authors' calculations based on ELMS 1998 and ELMPS 2006

**Table 7: Father's Sector of Work When Age 15, As Reported in 2006 Vs. 2012, Father not in Household in 2006 or 2012, Ages 30-54 in 2006**

		<u>Father's Sector in 2012</u>		
		<u>Government</u>	<u>Public Enterprise</u>	<u>Private</u>
<u>Father's Sector in 2006</u>	Government	70.8	6.7	22.5
	Public Enterprise	39.6	35.2	25.2
	Private	7.8	1.5	90.8
	Total	23.9	4.7	71.4

Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Table 8: Father's Sector of Work When Age 15, As Reported in 2006 vs. 2012, Father Not in Household in 2006 or 2012, Consistently Respondent, Ages 30-54 in 2006**

		<u>Father's Sector in 2012</u>		
		<u>Government</u>	<u>Public Enterprise</u>	<u>Private</u>
<u>Father's Sector in 2006</u>	Government	69.2	8.2	22.6
	Public Enterprise	33.7	39.1	27.3
	Private	8.3	1.5	90.2
	Total	24.2	5.4	70.4

Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Table 9: Father's Sector of Work When Age 15, As Reported in 2006 vs. 2012, Father not in Household in 2006 Or 2012, not Consistently Respondent, Ages 30-54 in 2006**

		<u>Father's Sector in 2012</u>		
		<u>Government</u>	<u>Public Enterprise</u>	<u>Private</u>
<u>Father's Sector in 2006</u>	Government	70.8	6.4	22.8
	Public Enterprise	45.6	33.2	21.2
	Private	7.1	1.5	91.4
	Total	23.3	4.8	72.0

Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Table 10: Labor Market Status, as Reported Contemporaneously for 1998 and 2006 and as Reported Retrospectively for Those Years from 2012 Data, Male Respondents Ages 30-54 in 2012 Present in Both Waves**

	<b>2006 contemp.</b>	<b>2006 retro. from 2012</b>	<b>1998 contemp.</b>	<b>1998 retro. from 2012</b>
Public	27.9	29.9	20.7	21.5
Private formal regular wage workers	12.3	11.0	5.3	7.5
Private informal regular wage workers	16.3	16.0	10.5	13.6
Irregular wage workers	8.9	16.3	14.2	16.8
Employers	14.1	11.1	7.3	6.4
Self-Employed	9.5	9.7	5.2	6.4
Unpaid Family Work	5.0	2.6	6.5	5.1
Unemployed	2.8	0.9	7.3	2.3
OLF	3.4	2.7	23.1	20.3
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Source: Authors' calculations based on ELMS 1998, ELMPS 2006 and ELMPS 2012

**Table 11: Labor Market Status, as Reported Contemporaneously for 1998 And 2006 and As Reported Retrospectively for Those Years from 2012 Data, Female Respondents Ages 30-54 in 2012 Present in Both Waves**

	<b>2006 contemp.</b>	<b>2006 retro. from 2012</b>	<b>1998 contemp.</b>	<b>1998 retro. from 2012</b>
Public	12.9	13.4	10.6	11.3
Private formal regular wage workers	1.1	1.2	0.7	0.5
Private informal regular wage workers	1.3	1.7	1.3	1.4
Irregular wage workers	0.6	0.9	0.8	0.8
Employers	1.0	0.7	0.3	0.3
Self-Employed	4.0	1.9	1.9	1.5
Unpaid Family Work	10.0	4.4	5.5	2.8
Unemployed	4.5	1.5	6.6	1.4
OLF	64.5	74.3	72.3	80.0
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Source: Authors' calculations based on ELMS 1998, ELMPS 2006 and ELMPS 2012

**Table 12: Labor Market Status, As Reported in 2006 versus 2012 Retrospective Data for 2006, Male Respondents Ages 30-54 in 2012 Present in Both Waves**

	Public	Private formal regular wage	Private informal regular wage	Irregular wage worker	Employers	Self-Employed	Unpaid Family Work	Unemployed	OLF
Public	87.6	4.3	2.6	1.6	1.3	1.3	0.2	0.3	0.8
Private formal regular wage	14.8	43.6	20.4	10.8	3.2	4.9	0.6	1.2	0.5
Private informal regular wage	7.0	14.0	32.8	24.6	7.7	8.1	1.9	1.2	2.7
Irregular wage worker	6.6	4.6	19.8	45.7	8.3	10.8	2.3	1.0	1.1
Employers	5.0	2.9	13.8	17.7	39.8	15.3	4.2	0.2	1.2
Self-Employed	3.6	4.9	13.2	20.4	16.9	36.7	1.7	0.4	2.1
Unpaid Family Work	4.6	5.3	23.6	19.7	16.8	8.7	19.4	0.9	1.0
Unemployed	13.7	14.9	27.1	16.8	1.8	5.4	5.7	5.1	9.5
OLF	7.9	6.9	15.2	16.8	6.6	7.2	1.7	3.4	34.3
Total	29.9	11.0	16.0	16.3	11.1	9.7	2.6	0.9	2.7

Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Table 13: Labor Market Status, As Reported in 1998 Versus 2012 Retrospective Data for 1998, Male Respondents Ages 30-54 in 2012 Present in Both Waves**

	Public	Private formal regular wage	Private informal regular wage	Irregular wage worker	Employers	Self-Employed	Unpaid Family Work	Unemployed	OLF
Public	85.0	4.2	2.6	4.0	1.4	0.4	0.4	1.1	1.0
Private formal regular wage	9.7	47.8	19.0	10.8	1.2	4.8	2.5	0.2	3.9
Private informal regular wage	5.6	10.2	34.0	21.1	8.6	6.0	1.1	3.5	10.0
Irregular wage worker	5.5	5.4	19.4	38.1	8.2	8.7	5.7	1.2	7.8
Employers	3.6	7.3	15.0	17.4	28.3	17.3	9.3	0.4	1.4
Self-Employed	7.8	2.2	13.5	21.4	11.8	25.0	10.9	0.2	7.1
Unpaid Family Work	2.2	0.9	13.4	21.9	10.8	6.4	19.6	5.1	19.6
Unemployed	3.9	12.0	15.1	26.4	3.5	6.5	4.9	6.3	21.5
OLF	4.2	3.1	8.7	9.1	1.4	3.2	4.5	3.1	62.7
Total	21.5	7.5	13.6	16.8	6.4	6.4	5.1	2.3	20.3

Source: Authors' calculations based on ELMS 1998 and ELMPS 2012

**Table 14: Labor Market Status, As Reported in 2006 versus 2012 Retrospective Data for 2006, Female Respondents Ages 30-54 in 2012 Present in Both Waves**

	Public	Private formal regular wage	Private informal regular wage	Irregular wage worker	Employers	Self-Employed	Unpaid Family Work	Unemployed	OLF
Public	88.7	2.6	0.6	0.0	0.0	0.3	0.0	1.1	6.8
Private formal regular wage	18.1	32.7	17.3	0.0	0.0	0.0	0.0	2.1	29.9
Private informal regular wage	4.4	7.8	26.5	3.0	1.2	4.0	0.0	8.2	45.0
Irregular wage worker	0.0	0.0	8.5	34.5	0.0	11.2	7.1	1.9	36.8
Employers	0.9	0.0	3.3	0.0	17.7	14.0	9.7	0.0	54.4
Self-Employed	0.9	0.0	2.2	1.3	3.9	17.7	8.1	0.3	65.7
Unpaid Family Work	1.2	0.0	0.0	0.5	1.6	3.3	20.2	1.2	72.0
Unemployed	4.9	3.1	4.8	0.0	0.0	0.7	1.8	5.5	79.2
OLF	2.0	0.4	1.1	0.8	0.2	0.9	2.9	1.3	90.4
Total	13.4	1.2	1.7	0.9	0.7	1.9	4.4	1.5	74.3

Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Table 15: Labor Market Status, As Reported in 1998 versus 2012 Retrospective Data for 1998, Female Respondents Ages 30-54 in 2012 Present in Both Waves**

	Public	Private formal regular wage	Private informal regular wage	Irregular wage worker	Employers	Self-Employed	Unpaid Family Work	Unemployed	OLF
Public	88.7	2.6	0.6	0.0	0.0	0.3	0.0	1.1	6.8
Private formal regular wage	18.1	32.7	17.3	0.0	0.0	0.0	0.0	2.1	29.9
Private informal regular wage	4.4	7.8	26.5	3.0	1.2	4.0	0.0	8.2	45.0
Irregular wage worker	0.0	0.0	8.5	34.5	0.0	11.2	7.1	1.9	36.8
Employers	0.9	0.0	3.3	0.0	17.7	14.0	9.7	0.0	54.4
Self-Employed	0.9	0.0	2.2	1.3	3.9	17.7	8.1	0.3	65.7
Unpaid Family Work	1.2	0.0	0.0	0.5	1.6	3.3	20.2	1.2	72.0
Unemployed	4.9	3.1	4.8	0.0	0.0	0.7	1.8	5.5	79.2
OLF	2.0	0.4	1.1	0.8	0.2	0.9	2.9	1.3	90.4
Total	13.4	1.2	1.7	0.9	0.7	1.9	4.4	1.5	74.3

Source: Authors' calculations based on ELMS 1998 and ELMPS 2012

**Table 16: Collapsed Labor Market Status, as Reported in 2006 versus 2012 Retrospective Data for 2006, Male Respondents Ages 30-54 In 2012 Present in Both Waves**

	Public	Private Wage	Non-wage work	Not working
Public	87.6	8.5	2.8	1.1
Private Wage	9.4	72.2	15.7	2.7
Non-wage work	4.5	38.2	55.5	1.9
Not working	10.5	47.9	14.3	27.4
Total	29.9	43.3	23.3	3.6

Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Table 17: Collapsed labor market status, as reported in 1998 versus 2012 retrospective data for 1998, male respondents ages 30-54 in 2012 present in both waves**

	Public	Private Wage	Non-wage work	Not working
Public	88.7	3.1	0.3	7.9
Private Wage	8.6	43.0	5.8	42.6
Non-wage work	1.1	1.5	27.4	70.0
Not working	2.2	2.7	3.9	91.3
Total	13.4	3.8	7.0	75.8

Source: Authors' calculations based on ELMS 1998 and ELMPS 2012

**Table 18: Collapsed Labor Market Status, as Reported in 2006 versus 2012 Retrospective Data for 2006, Female Respondents Ages 30-54 in 2012 Present in Both Waves**

	Public	Private Wage	Non-wage work	Not working
Public	85.0	10.7	2.3	2.0
Private Wage	6.3	66.3	17.7	9.7
Non-wage work	4.3	37.8	46.8	11.2
Not working	4.1	28.8	10.5	56.7
Total	21.5	38.0	17.8	22.6

Source: Authors' calculations based on ELMPS 2006 and ELMPS 2012

**Table 19: Collapsed Labor Market Status, As Reported In 1998 versus 2012 Retrospective Data for 1998, Female Respondents Ages 30-54 in 2012 Present in Both Wave**

	Public	Private Wage	Non-wage work	Not working
Public	86.7	1.6	0.0	11.7
Private Wage	9.4	24.2	9.9	56.6
Non-wage work	1.2	3.2	12.8	82.9
Not working	2.2	2.0	4.3	91.5
Total	11.3	2.7	4.6	81.4

Source: Authors' calculations based on ELMS 1998 and ELMPS 2012

**Table 20: Rates of Status Change in Panel Data for 1998 to 2006 versus Rates of Status Change in Retrospective Data from 2012 for Changes from 1998 to 2006 by Sex and Status in 1998**

	Male		Female	
	Retrospective	Panel	Retrospective	Panel
Public	2	11	6	11
Private formal regular wage workers	21	45	20	55
Private informal regular wage workers	27	63	41	82
Irregular wage workers	25	73	49	95
Employers	7	43	0	68
Self-Employed	15	61	17	69
Unpaid Family Work	66	74	11	77
Unemployed	97	90	56	85
OLF	88	89	8	25
Total	35	59	9	33

Source: Authors' calculations based on ELMS 1998, ELMPS 2006 and ELMPS 2012