CrossMark

Introducing the Tunisia Labor Market Panel Survey 2014

Ragui Assaad[1], Samir Ghazouani[2,3], Caroline Krafft[4*] and Dominique J. Rolando[5]

* Correspondence:
cgkrafft@stkate.edu
[4]Department of Economics, St.
Catherine University, 2004 Randolph
Ave, Saint Paul, MN 55105, USA
Full list of author information is
available at the end of the article

Abstract

This paper introduces the Tunisia Labor Market Panel Survey (TLMPS) of 2014, the first round of a publicly available nationally representative longitudinal household survey. We provide a description of the sample and questionnaires. We discuss a number of data collection issues, such as non-response, as well as what was done to address these issues. The construction of sample weights is detailed. A comparison of the TLMPS to other Tunisian datasets is conducted to illustrate the representativeness of the data in terms of key demographic and labor market measures. Key features of the Tunisian labor market and potential avenues for research using the TLMPS are discussed.

JEL Classification: J00, C81, C83

Keywords: Survey data, Public use data, Sample weights, Labor, Tunisia

1 Introduction

The Egypt Labor Market Panel Surveys (ELMPSs) of 1998, 2006, and 2012 and Jordan Labor Market Panel Survey (JLMPS) of 2010 have become well-recognized data sources for labor market studies in the Middle East and North Africa (MENA). These two surveys have been used in numerous research endeavors including peer-reviewed academic publications, dissertations, and international organization reports.[1] As part of the same series of surveys, the Tunisia Labor Market Panel Survey (TLMPS) of 2014 is the first wave of what will eventually become a longitudinal survey of the Tunisian labor market. Being far richer than any currently available data, the TLMPS 2014 is a much-needed addition in a landscape of otherwise scarce publicly accessible data on the Tunisian labor market. The TLMPS 2014 was collected in partnership between the Economic Research Forum (ERF) and the Tunisian National Institute of Statistics (INS).

Similar to its Egyptian and Jordanian counterparts, the TLMPS 2014 is a nationally representative survey that features detailed information on households and individuals, especially in regard to labor market characteristics. As in other countries in the MENA region, Tunisia suffers from high unemployment, particularly for university graduates, youth, and women, and from low female labor force participation (Haouas et al. 2012; World Bank 2014; Assaad et al. 2016a). The survey allows for an in-depth investigation of current employment characteristics as well as analyses of broader labor market dynamics. For instance, analyses have already revealed the particularly long unemployment durations

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 2 of 21

Tunisian youth experience, long even in comparison to other countries in the region (Assaad and Krafft 2016). The survey provides insight into jobs held across the individual's career trajectory and current income, as well as benefits and other non-pecuniary aspects of employment. The TLMPS 2014 also includes retrospective information on educational trajectories, residential mobility patterns, migration history, and marital and fertility history and can thus be used to conduct in-depth analyses of the life course.

In addition, the TLMPS 2014 includes detailed information on the socio-economic status of individuals, such as parental background (education and employment when the individual was 15 years of age), as well as information on the assets and resources of the household. This information permits researchers to explore intergenerational dynamics even after individuals have left their natal households. It also allows for better understanding of issues such as inequality of opportunity in education or in wages, income, and consumption (Krafft and Alawode 2016; Assaad et al. 2016c; Assaad et al. 2016d). The TLMPS also facilitates education research with the inclusion of a battery of questions on educational attainment, achievement, and experiences. Previous research has shown that Tunisia has substantial problems with education equity, quality, accountability, and efficiency (Milovanovitch 2014; Ben-Ayed et al. 2016). These problems can be further explored through analyses of the education data provided by the TLMPS 2014, analyses akin to those undertaken for Egypt and Jordan using the ELMPS and JLMPS (Assaad and Krafft 2015a; Assaad and Saleh 2015; Krafft and Alawode 2016). The TLMPS 2014 includes detailed migration data for both current migrants and return migrants, including information on remittances, timing and duration of migration, countries of destination, job characteristics abroad and upon return, the resources and networks used to facilitate migration, and the legal status of the migrant while abroad.

With the public release of the TLMPS[2] in August 2016, we hope to provide a reliable source of data for researchers investigating a variety of labor market and related topics. Tunisia had a population of close to 11 million in 2014 and has played a critical role in the region in the aftermath of the Arab Spring, being the only country to successfully undergo a democratic transition. However, Tunisia has suffered from poor macroeconomic performance and political and security setbacks in the wake of the revolution (World Bank 2014). Tunisia, like many countries in MENA, suffers from high unemployment, especially among youth (Assaad and Krafft 2016). This is in part due to stagnation in job creation, especially for university graduates and youth, with low firm dynamism contributing to the slow rate of job creation (Rijkers et al. 2014). Youth in Tunisia prefer to go through long unemployment spells in order to obtain good jobs, usually in the public sector, rather than settle for mediocre to poor jobs in the informal economy (Stampini and Verdier-Chouchane 2011). While there have been attempts to address these challenges, including an entrepreneurship track for the final year of undergraduate studies (Premand et al. 2012) and active labor market policies targeting graduates (Broecke 2013), these policies tend to be mostly ineffective (World Bank 2014; Krafft and Assaad 2015). Thus, empirical research on labor markets in Tunisia is vital to address growing concerns over youth unemployment and poor job quality in a context where educational attainment has been rising very rapidly. The TLMPS is an essential new resource to conduct such research.

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 3 of 21

The paper proceeds as follows: Section 2 discusses the design of the survey, including information on sampling and questionnaire design. Section 3 describes the data collection process and includes discussion of non-response and the calculation of sample weights to correct for non-response problems. This section also provides an overview of potential data problems. Section 4 compares the results of the TLMPS to those from other Tunisian sources, such as the quarterly Survey of Population and Employment (ENPE) carried out by INS. Finally, Section 5 concludes the paper by summarizing potential uses for the TLMPS and presents plans for future work.

2 The survey design

2.1 Sample

The initial sample frame included around 5160 households drawn from a larger sample that is regularly used to conduct the quarterly survey on population and employment in Tunisia. This larger sample contained 18,000 households as of the last quarter of 2012. The drawing of the sample was done in two stages. In the first stage, 258 enumeration areas were randomly drawn according to the principle of probability proportional to size from the list of enumeration areas drawn up in the 2004 census. This first sampling stage was carried out using 46 strata comprised of the urban/rural components of each of Tunisia's governorates.[3] The final sample was made up of 253 clusters (out of a possible 40,377 nationally). In the second stage, 20 households were supposed to be drawn at random from each cluster. This procedure was, however, not strictly followed in the field, as discussed in Section 3.

2.2 Questionnaires

The survey incorporates questionnaires to be administered at both the household and individual levels. At the household level, there was a general household questionnaire, as well as a questionnaire specifically about current migration, transfers, and agricultural and non-agricultural enterprises. At the individual level, there was a detailed questionnaire for working age individuals (15+) and an abbreviated version of the questionnaire for those 6–14 years old. The modules included in each questionnaire are listed in Table 1.

The main household questionnaire and the migration/enterprise questionnaire were designed to be answered by the most knowledgeable individual in the household, usually the head or the spouse of the head. Along with information on the characteristics of the dwelling, access to public services, and ownership of durables, the household questionnaire includes a full household roster with information on basic demographic characteristics, such as age, sex, and relationship to the head of household. The migration/enterprise questionnaire includes information on any family members currently abroad, remittances, and other transfers, such as child support and pensions. Data were gathered on both non-agricultural and agricultural enterprises, including assets used and net revenues.

The ELMPS and JLMPS had a single questionnaire for all individuals regardless of age. However, in Tunisia, a distinct questionnaire for individuals 6–14 was designed in order to more carefully incorporate measures of child labor. In Tunisia, as of 2011, among children 5–14 years of age, around 3 % were engaged in child labor, and more

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 4 of 21

Table 1 Sections found in TLMPS 2014 questionnaires

Household questionnaire	Adult questionnaire	Migration questionnaire
Household demographics	Parent's characteristics[a]	Current migration
Housing, services, and facilities	Siblings' characteristics[a]	Remittances
	Health	Other sources of income
	Education[a]	Non-agricultural activities
	Employment[a]	Employment outside of household
	Unemployment	Expenditures
	Job characteristics[a]	Assets (including agricultural assets)
	Secondary jobs	Revenue of enterprise
	Job mobility	Livestock
	Subsistence and domestic work	Capital equipment
	Marriage	Harvest and disposal of crops
	Fertility	Other agricultural income
	Women's status	
	Female employment	
	Earnings (first and secondary jobs)	
	Return migration	
	Information technology	
	Savings and borrowing	

[a]Also available in child questionnaire

5–11-year-olds (3.2 %) than 12–14-year-olds (1.2 %) were engaged in child labor. Specifically, among 5–11-year-olds, 0.3 % were working for a wage outside the home, 1.5 % engaged in non-wage work outside the home, and 1.9 % worked for a family enterprise (children could engage in multiple forms of work) (Ministry of Development and International Cooperation et al. 2013).

The results from the TLMPS captured just 13 children 6–14 out of 1147 with data engaged in market work (reference period 3 months) (1.0 % when weighted). This rose to 29 children 6–14 (1.5 % when weighted) when subsistence (own consumption) work was included. Although these rates are lower than those detected in 2011, they are not improbably so, being similar to the rate found for 12–14-year-olds, and the social and economic milieu has also evolved. We therefore do not recommend using the TLMPS to describe working children, using either the market or subsistence definition, as the sample of working children is small. This small sample is unlikely to be due to the design of the questionnaire, given the special attention to this issue and the low rates in other studies. Instead, child labor in Tunisia is relatively rare, and studies of child labor will require either a much larger or specially targeted sample. As very little child labor was detected even with this special questionnaire design, in future LMPSs, we plan to revert to a single questionnaire with a few additional questions targeted to children 6–14. Substantially, more domestic labor (housework or chores), a rate of 26 %, was detected in the TLMPS, with expected differences by sex and age. However, this result was using questions identical to those for adults and therefore does not justify a separate questionnaire.

The working-age (15+) questionnaire includes a variety of modules on labor market experience and outcomes and related issues. On the labor market side, it elicits

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 5 of 21

information on the current labor market status of the individual, detailed job characteristics (for the employed), wage earnings and non-wage benefits (for wage workers), and participation in domestic and subsistence work. Those who work were asked about both primary and secondary jobs (if any). The questionnaire also includes a detailed labor market history starting from the first labor market status after leaving school and moving forward towards the present for those who ever worked.[4] Further, there is a detailed section on return migration for those who ever worked abroad.

The labor market intersects with a number of other important life experiences, such as education, fertility, and marriage, which are also captured in the TLMPS individual questionnaire. For instance, there are modules on family background (parents and siblings), educational experiences, health, and residential mobility. For women, a section is devoted to fertility issues, the status of women in the household, and work-family issues such as child care and maternity leave. Data were also collected from both men and women on marriage and decisions around marriage, such as the incidence of kin marriage and living arrangements at marriage. Finally, there are modules on financial decision-making, with specific questions about savings and borrowing, as well as on the use of information technology.

3 Data collection, non-response, data problems, and the calculation of sample weights

3.1 Data collection

An important aspect of data collection was the use of tablets and digitized versions of the questionnaires. These digitized questionnaires were produced using software tailored specifically for this project.[5] For INS, this was a major innovation and the first time tablets rather than paper questionnaires were used to record data in a household survey. A number of challenges in the fielding and data processing stages, which we discuss below, arose from the process of transitioning from a paper to digital questionnaire model. Prior to data collection, the software and the questionnaires were tested. The pretest was conducted over a period of 3 days in the governorate of Ben Arous and covered about 100 households.

For the purpose of fieldwork, 25 teams were appointed by INS from its own field staff. Each team was composed of three interviewers and one supervisor. A training session, which lasted 10 days, was organized in advance of fielding. This session included Tunisian ERF members in charge of the project, INS staff, and an expert from the Egyptian Central Agency for Public Mobilization and Statistics (CAPMAS), the fielding partner for the ELMPS. As part of the training, enumerators received a manual with detailed information on the questions and the design of the questionnaire. During the training, all the trainees had to implement applied exercises, partly using the tablets, to familiarize themselves with the new digitized data collection process.

Fieldwork started in February of 2014, and the majority of it was completed within 1 month. However, due to difficulties in getting households to respond in certain areas, and the need to pause fielding while the Tunisian Population Census was underway, fieldwork continued until November of 2014. Further, due to problems in fielding, a number of households from the initial fielding were recontacted by phone in the spring of 2015 to complete their interviews.[6] Given the length of the questionnaire and

because of the need to interview the individual him or herself, quite often more than one interview session was needed per household. Once the data were collected, they were transmitted daily and stored in central servers. The different questionnaires were saved as different files and linked in central processing based on household identifiers.

3.2 Non-response

There were several different problems with non-response during the fielding. First, households often refused to respond entirely. Second, in completing the household survey, some individuals were not captured and some households refused or failed to answer the migration/enterprise questionnaire. In this section, we discuss the patterns of non-response, which are incorporated into the weights, discussed below.

3.2.1 Non-response of the entire household

While the initial goal was to collect data from 5160 households, time pressures reduced the intended sample to 4986 households. Of the 4986 households initially selected, interviews were completed with only 4521, generating an overall household non-response rate of 9.3 %. Additionally, because several clusters were found not to have the requisite 20 households at the end of the data collection stage, additional households were added to some clusters to improve the response rate, leading to wide variation in the number of the households per cluster. The minimum number of households interviewed in a cluster was 8 and the maximum was 34. The mean was 19.7, and the median was 20, with the interquartile range going from 17 to 22 households.

  After this additional work to add households to the sample, non-response rates at a cluster level ranged from 0 % (complete response), which occurred for 29 % of clusters, to a maximum of 62.5 %. The mean non-response at the cluster level was 10.2 %, the median was 6.7 %, the 75th percentile was 13.3 %, and the 90th percentile was 24.8 %. This household non-response is incorporated in the weights at a cluster level, with the households that did respond within a cluster representing those that did not.

3.2.2 Non-response to child, adult, and migration/enterprise questionnaires

As well as problems with non-response on the household level, there were problems with completing the child, adult, and migration/enterprise questionnaires. We developed weights to account for non-response to each of these questionnaires in their entirety. However, individuals often stopped answering partway through a questionnaire, suffered from incorrect skips, or other data problems, such that data is sometimes missing for a particular question within a questionnaire that contains some data. Additional data imputation techniques, implemented on a question-by-question basis, are required for these problems and are discussed below. Here, we limit our discussion to remedies for non-response to questionnaires in their entirety. Table 2 specifically shows the number of individuals (or households) who should have answered a questionnaire and divides this into the number who did respond and those who did not and then presents the non-response rate. We determine whether or not an individual should have answered a question based on the age given for him or her in the household roster.

Table 2 Non-response rates for child, adult, and migration/enterprise questionnaires

	Child questionnaire (individuals)	Adult questionnaire (individuals)	Migration/enterprise questionnaire (households)
Should have answered	2305	12,514	4521
Did answer	2078	11,738	4336
Did not answer	227	776	185
Non-response rate (unweighted)	9.8	6.2	4.1
Non-response rate (weighted)	11.7	7.8	6.1

Source: Authors' calculations based on TLMPS 2014

We begin by looking at child non-response. Of the 2305 children who should have answered the child questionnaire for children 6–14, there were 2078 who did and 227 who did not, producing an unweighted non-response rate of 9.8 % and a weighted non-response rate of 11.7 %. For adults (15+), of the 12,514 individuals who should have answered this questionnaire, 11,738 did and 776 did not, yielding an unweighted non-response rate of 6.2 % and a weighted non-response weight of 7.8 %. With regard to the household-level migration/enterprise questionnaire, of the 4521 households who should have answered, 4336 did and 185 did not. The unweighted non-response rate was therefore 4.1 %, and the weighted non-response rate was 6.1 %. One thing to note is that areas with higher weights (likely areas with higher household non-response) were also more likely to not respond to other questionnaires.

3.3 Calculation of sample weights

3.3.1 Household weights[7]

The structure of the sample drove the creation of the initial household weights. A stratified two-stage sample was used in fielding the TLMPS. The strata were governorates divided into urban/rural components. Tunisia has 24 governorates, of which two (Tunis and Monastir) did not have rural components, generating 46 strata. Because the 2014 population census had not yet been completed, the sample frame based on the 2004 census with appropriate updates was used to obtain the TLMPS 2014 sample. Specifically, in the first stage, 258 clusters were drawn from within their respective strata according to the principle of probability proportional to size. These were later reduced to 253 clusters due to fielding problems. These clusters are the primary sampling units (PSUs) of the survey. In the second stage, 20 households were randomly selected from each cluster after the sample frame of the cluster was updated by means of a relisting.

Setting aside, for the moment, all the subsequent fielding challenges, this sampling strategy implies a straightforward weight on the household level. For the first stage of sampling, there is a weight, $w_{d,s}$, for each cluster, d, in each stratum, s. This weight is defined as

$$w_{d,s} = \frac{\sum_{j=1}^{D} m_j}{m_d} * \frac{1}{a_s} \tag{1}$$

where m_j is the number of households in cluster j, with j ranging from 1 to D, the total number of clusters within the stratum. s ranges from 1 to 46, and a_s is the number of clusters selected in stratum s. Essentially, this initial weight gives a multiplier to each cluster so that its households represent the stratum and then divides by the number of

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 8 of 21

clusters selected within the stratum. The number of households b_d in each selected cluster was updated in 2014 and then the weights were adjusted as follows:

$$w'_{d,s} = w_{d,s} * \frac{b_d}{m_d} \tag{2}$$

In this way, the growth or contraction of each district between 2004 and 2014 is reflected in the first-stage sampling weights.

For the second sampling stage where households are selected in each cluster, the weight of each household, h, within each cluster d and stratum s is calculated based on r_d, the number of households responding in the cluster, as follows:

$$w_{h,d,s} = w'_{d,s} * \frac{b_d}{r_d} \tag{3}$$

After implementing these household weights, there remained small differences between the expanded estimates of the number of households from the cluster sample and the 2014 census population numbers, amounting to around 0.1 million households out of 2.7 million. This difference was adjusted for using the 2014 census household count at the stratum level. Even after applying these adjustments at the household level, we obtain an overall population estimate of 9.76 million compared to the 2014 population census estimate of 10.96 million. Specifically, we found that young adults and males were particularly likely to be under-represented in our sample compared to the census estimates after we applied the household weights to individuals in our household roster. Since we are primarily interested in individual labor market phenomena, rather than household outcomes, we further adjusted our weights to correctly represent the number of individuals in each stratum by 5-year age group, g, and sex, x, using the data from the 2014 census. Specifically, we generated for each individual, i, the following weight:

$$w_{i,x,g,h,d,s} = \frac{w_{x,g,h,d,s}}{\sum_{i=1}^{k} w_{x,g,h,d,s}} * n_{s,x,g} \tag{4}$$

where $n_{s,x,g}$ was the number of individuals in a stratum of a particular age group and sex based on the census and k was the number of individuals observed in our sample. Essentially, we took an individual's share in the sample of the same age group, sex, and stratum based on the household weights and multiplied it by the stratum, age, and sex-specific population as measured in the 2014 population census.[8] This yielded age group, sex, and strata-representative individual statistics. These weights, which we refer to as household *roster* weights, were therefore no longer equal among household members, since they had different inputs into Eq. (4), but could be expanded to individual-representative statistics. Some differences in household and individual sampling may be due to issues we have identified in other surveys with distinguishing individual households in settings where extended families may share some but not all of the characteristics of the official definition of a household (Assaad and Krafft 2013). However, the absence of individuals entirely from the roster in a pattern similar to what we see in the individual non-response models, below, suggests that some individuals were dropped off the roster when they were difficult to locate.

To both test and account for any systematic patterns on non-response, we use the household roster data on individuals and their households to predict the probability of non-response to the various questionnaires (as summarized in Table 2). We use logit models for the probability of non-response, and these models are the basis of the adjustments that are incorporated into the weights. Table 3 presents the odds ratios for the different logit models. Models are estimated separately for the individual adult questionnaire, the individual child questionnaire, and the migration/enterprise questionnaire (household level). The models demonstrate clearly that there are some systematic patterns, particularly demographic and geographic patterns that need to be accounted for in adjusting the household roster weights for the individual questionnaires and the household weights for the migration questionnaires. Note that the models do not include individuals (or households) with missing information in the household questionnaire, since household characteristics are critical inputs into the model. Those households with data missing on critical predictors were assigned the mean probability of non-response for a particular outcome.

For adults, household composition was particularly important, with individuals in households with more working age males more likely to fail to respond and those with small children less likely to fail to respond. Similar patterns are observed for the migration/enterprise questionnaire, but not the child non-response model. Geographic differences were often substantial, but there are somewhat mixed patterns across questionnaires. Relationship to head and age group were not significant, but age groups do show systematic patterns, as noted earlier. In the migration/enterprise questionnaire, households with heads aged 45–49, 55–59, and 70–74 were less likely to fail to respond than those with heads younger than 25. There were not strong systematic differences in terms of either durable goods or housing asset index quintiles.

The household weights discussed above are the starting point for the migration/enterprise questionnaire weights that adjust for non-response. Similarly, the household roster weights are the starting point for the individual (child and adult) weights that adjust for non-response. For those individuals and households with adequate household information in the household questionnaire, we predicted the probability of non-response (attrition), using the models reported above. These probabilities, $\Pr(A_{ih})$, are denoted here for individual i, but they are calculated in a similar way for the migration/enterprise questionnaire. Adjustment factors, r_{ih}, are then calculated as

$$r_{ih} = 1/(1 - \Pr(A_{ih})) \tag{5}$$

These adjustment factors are used as multipliers to modify the weights of individuals or households who did respond to represent those who did not. Essentially, the final weights are specific to the questionnaire, q, and are denoted as $w_{q,i,x,g,h,d,s}$ and are calculated as follows:

$$w_{q,i,x,g,h,d,s} = w_{i,x,g,h,d,s} * r_{ih} \tag{6}$$

It is important to keep in mind that for the migration/enterprise questionnaire these weights are not age group or sex-specific and are the same for all household members.

Table 3 Logit models (odds ratios) for probability of non-response of adults, children, and households (for the migration/enterprise questionnaire)

Dependent variable:	Pr (adult did not respond)	Pr (child did not respond)	Pr (household did not respond)
Household composition			
Mean no. of children 0 to 5 in household	0.590***	0.981	0.458***
	(0.086)	(0.137)	(0.109)
Mean no. of children 6 to 14 in household	0.892	1.020	1.089
	(0.086)	(0.115)	(0.139)
Mean no. of WA males in household	1.210**	1.063	1.371*
	(0.073)	(0.146)	(0.172)
Mean no. of WA females in household	1.122	0.857	1.373**
	(0.073)	(0.118)	(0.144)
Mean no. of elderly males in household	1.287	1.117	1.273
	(0.291)	(0.545)	(0.856)
Mean no. of elderly females in Household	0.815	0.785	0.596
	(0.185)	(0.300)	(0.251)
Region (Greater Tunis omit.)			
North East	0.724	3.976***	0.481*
	(0.168)	(1.304)	(0.179)
North West	0.619	1.334	0.264
	(0.395)	(0.681)	(0.201)
Center East	2.110**	1.847	1.415
	(0.479)	(0.597)	(0.448)
Center West	1.192	2.951*	0.361
	(0.385)	(1.260)	(0.216)
South East	0.205**	0.267**	0.072*
	(0.116)	(0.135)	(0.076)
South West	0.128**	0.927	0.292
	(0.086)	(0.474)	(0.198)
Residence (urban omit.)			
Rural	0.818	0.866	0.868
	(0.284)	(0.617)	(0.507)
Rural and region interactions			
North East	1.784	0.040**	0.948
	(0.706)	(0.042)	(0.685)
North West	0.167*	0.896	0.110
	(0.140)	(0.771)	(0.154)
Center East	0.805	0.939	0.677
	(0.307)	(0.732)	(0.428)
Center West	2.357	0.423	1.930
	(1.058)	(0.350)	(1.597)
South East	(.)	3.351	(.)
	(.)	(3.462)	(.)
South West	4.132	0.499	3.185
	(3.556)	(0.563)	(3.296)

Table 3 Logit models (odds ratios) for probability of non-response of adults, children, and households (for the migration/enterprise questionnaire) *(Continued)*

Dependent variable:	Pr (adult did not respond)	Pr (child did not respond)	Pr (household did not respond)
Relationship to head (adult: head omit.)			
Spouse	1.028		
	(0.722)		
Son/daughter	4.356		
	(3.842)		
Other/missing	4.259		
	(3.821)		
Relationship to head (child: son/daughter omit.)			
Not son/daughter		1.309	
		(0.540)	
Age group (adult: 15–19 omit., migr.: <25 omit.)			
20–24	1.471		
	(0.501)		
25–29	0.736		0.466
	(0.263)		(0.483)
30–34	1.449		0.202
	(0.523)		(0.210)
35–39	1.102		0.341
	(0.458)		(0.311)
40–44	1.789		0.209
	(0.864)		(0.196)
45–59	1.126		0.084**
	(0.515)		(0.076)
50–54	0.908		0.093**
	(0.432)		(0.084)
55–59	1.208		0.055**
	(0.615)		(0.052)
60–64	0.631		0.091*
	(0.309)		(0.088)
65–69	0.852		0.194
	(0.630)		(0.207)
70–74	0.899		0.056**
	(0.564)		(0.060)
75+	0.493		0.077*
	(0.326)		(0.090)
Female interacted with age group (adult: 15–19 omit.)			
20–24	1.113		
	(0.529)		
25–29	1.135		
	(0.608)		
30–34	0.810		
	(0.564)		

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 12 of 21

Table 3 Logit models (odds ratios) for probability of non-response of adults, children, and households (for the migration/enterprise questionnaire) *(Continued)*

Dependent variable:	Pr (adult did not respond)	Pr (child did not respond)	Pr (household did not respond)
35–39	0.671		
	(0.482)		
40–44	0.298		
	(0.213)		
45–59	0.615		
	(0.438)		
50–54	0.269		
	(0.196)		
55–59	0.327		
	(0.248)		
60–64	0.381		
	(0.299)		
65–69	0.507		
	(0.491)		
70–74	0.196		
	(0.188)		
75+	0.377		
	(0.343)		
Age group (child: 6–9 omit.)			
10–14		1.187	
		(0.328)	
Female interacted with age group (child: 6–9 omit.)			
10–14		0.760	
		(0.285)	
Sex (male omit.)			
Female	6.836	1.451	1.427
	(7.050)	(0.398)	(1.027)
Marital status (single omit.)			
Married	0.411		0.544
	(0.487)		(0.465)
Widowed/divorced	3.158		1.642
	(2.575)		(1.160)
Spouse present	7.618		3.509
	(8.203)		(2.831)
Residency status			
Absent	1.495		0.782
	(0.653)		(1.021)
Housing factor quintile (lowest omit.)			
Second	0.913	1.120	1.148
	(0.174)	(0.331)	(0.403)
Third	0.994	1.289	1.050
	(0.212)	(0.403)	(0.424)

Table 3 Logit models (odds ratios) for probability of non-response of adults, children, and households (for the migration/enterprise questionnaire) *(Continued)*

Dependent variable:	Pr (adult did not respond)	Pr (child did not respond)	Pr (household did not respond)
Fourth	0.914	0.623	0.898
	(0.213)	(0.236)	(0.358)
Highest	0.888	1.446	1.124
	(0.222)	(0.468)	(0.457)
Durables factor quintile (lowest omit.)			
Second	0.788	1.261	0.857
	(0.158)	(0.353)	(0.315)
Third	0.735	0.969	0.564
	(0.147)	(0.327)	(0.235)
Fourth	0.685	1.358	0.500
	(0.182)	(0.439)	(0.222)
Highest	1.625*	1.570	1.736
	(0.389)	(0.532)	(0.768)
Rel. to head and sex interactions			
Spouse and female	0.732		
	(0.712)		
Son/daughter and female	0.139*		
	(0.135)		
Rel. other/miss and female	0.583		
	(0.594)		
Marital status and sex interactions			
Married and female	1.845		
	(2.460)		
Div./widow. and female	0.224		
	(0.218)		
Spouse present and sex interactions			
Spouse present and female	0.234		1.684
	(0.297)		(1.891)
Pseudo R-sq.	0.121	0.087	0.158
N (Obs.)	10,858	2214	4116

Source: Authors' calculations based on TLMPS 2014
Notes: (.) denotes a variable that perfectly predicted responding; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

3.4 Data problems

In addition to problems with non-response, a number of data problems occurred in fielding, in part due to the transition from paper to digital questionnaires. In this section, we describe some of the data collection problems and detail our recommendations for managing them.

First, there were some poorly designed skip patterns in the questionnaire, which were particularly problematic. For instance, there were problems with skipping data on the highest year of school for certain levels of schooling, or only asking for costs of marriage for individuals who lived with their extended families after they married, a non-representative sample. Additionally, the tablets did not have the skip patterns fully

programmed but instead just included indications of which questions should be answered. Thus, an individual might identify as a wage worker in the job characteristics section but not answer the wage questions. Likewise, an individual identified as self-employed might answer the wage questions. These problems were pervasive through-out the data. In the created variables in the publicly available data set, we followed a rule of assuming that earlier information was the most accurate; thus, a self-employed individual should not have answered the wage question and would have no (created variable) data on wages. However, we left the raw data directly corresponding to ques-tions as it was collected to allow researchers to make their own preferred data cleaning decisions. In part because of the problems with skip patterns individuals often gave contradictory information, such as a male answering the fertility questions. Again, we left the raw data as is but assumed earlier data were correct in creating new variables in the public release dataset. The lessons from transitioning to digital processes, par-ticularly the importance of programming skip patterns and validation rules, will be in-corporated into future surveys such as the upcoming 2016 JLMPS.

Non-response to individual questions, as well as questionnaires, was a problem. It tended to worsen over the course of the survey, with more data missing on later ques-tions. We recommend researchers examine whether non-response was random or not for particular variables of interest using other questions. In our analyses for questionnaire non-response, discussed above, we found that non-response tended to be largely random. A further non-response problem occurred because of how some of the questions were programmed. In certain (check-box style) questions, it was not possible to distinguish be-tween a no, a missing or other non-response, and individuals for whom the question was not applicable. These questions must be treated with particular caution. For example, when individuals were asked their highest year of schooling completed within a level, everyone who did not answer the question was given a zero, but zero would also be a valid response for someone who was just starting school. In cleaning the data, when there was a zero response, we set it to missing if it was given by an individual who, per the preceding raw variables and skip patterns, should not have answered the question. However, remaining zero responses (or "no" responses for some variables) might actually represent missing data and should be treated with some caution.

In managing the data problems in research, it is also important to think about how missing data will affect the variable or statistic of interest. For instance, randomly missing responses to a yes/no question on attending pre-primary education will not bias an esti-mate of the percentage of children who attended pre-primary education. However, among women who have given birth, if data is missing on 10 % of their births, this will bias total fertility rates downward. In what follows, we provide an example using fertility data of the extent of bias and how to correct that bias (as much as possible).

The total fertility rate (TFR) is calculated based on the annual probability of giving birth in various age brackets, namely the age-specific fertility rate (ASFR). This annual probability is calculated relative to all women (married, childbearing, or not) in the age bracket, typically based on births in the 5 years preceding a survey. In the TLMPS, there is a section in the individual questionnaire on fertility for ever-married women ages 18–59. The women are asked first whether they have ever given birth, then their number of live births are recorded, and then data about each birth is recorded (sex, twin status, month, year of birth). Among women who had ever given birth and who

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 15 of 21

had data on the number of births, approximately 10 % had no data about even one specific birth. Further, among women who did have data on at least one birth, the ratio of total births (based on the number of births) to births with data was 1.03.

Thus, there are multiple points where missing data will bias the TFR downwards. First, the non-response to the individual questionnaire may make it erroneously appear that women in various age brackets, as identified in terms of age, sex, and marital status in the household questionnaire, are not having children at all. This can be corrected by focusing on the sample that did respond to the individual questionnaire and using the expansion weights for individual questionnaire respondents. The failure to report any births and the under-reporting of births will both bias TFR downward. Assuming that non-response in this section is random, it is possible to correct for both of these issues. First, for women who gave birth, did report some births, but potentially under-reported, we can multiply the births they did report separately by a ratio of the total number of births they stated to the number of births they reported separately (1.03 on average). Second, we can further multiply the number of live births for women who report data for specific births by the (weighted) ratio of all women who reported giving birth to women with specific birth data (1.11). Without correcting for question-naire non-response, no births at all, and under-reporting of specific births, the calcu-lated TFR is 1.81. Correcting for just questionnaire non-response generates a TFR of 1.91. Implementing all of the corrections generates a TFR of 2.11, which is much closer to recent fertility rates reported by the INS, which were 2.13 in 2010, 2.15 in 2011, and 2.20 in 2012 (Assaad et al. 2016b).
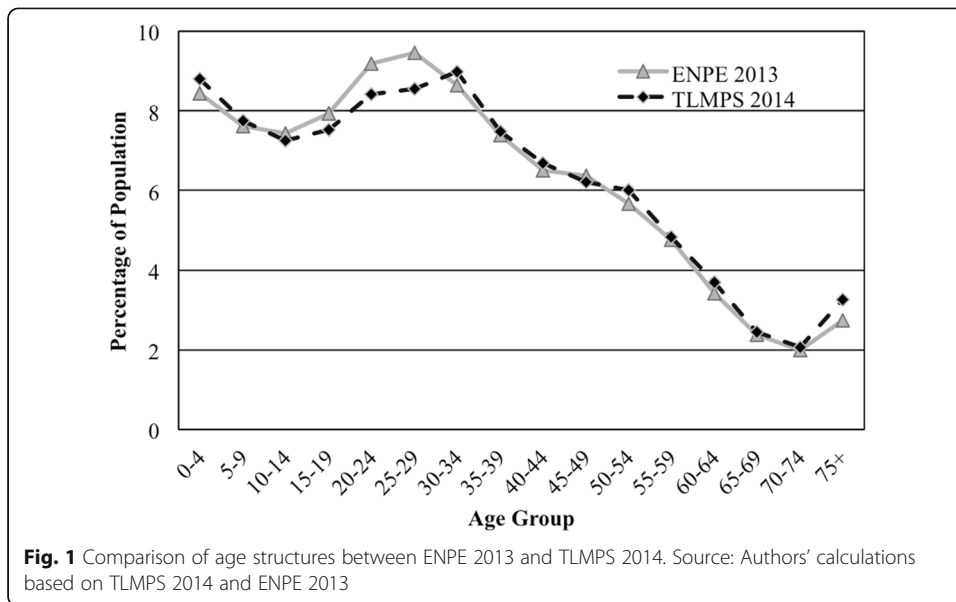
4 Comparison of TLMPS results to other surveys

To assess whether the TLMPS accurately represents individuals and households in Tunisia after incorporating sample and non-response weights, we compared the 2014 TLMPS to the 2013 Enquete Nationale sur la Population et l'Emploi (ENPE). The ENPE is a labor force survey collected by INS focusing on labor market characteristics and demographics. The 2013 ENPE is made up of a nationally representative sample of 472,244 individuals.

4.1 Demographic and labor market characteristics

Figure 1 provides a comparison of age distributions as depicted by the 2013 ENPE and the 2014 TLMPS using the household roster weights. Because these incorporate 5-year age groups, the TLMPS in fact represents the 2014 census exactly. Comparing to the ENPE, the age distributions generally followed a similar pattern. The TLMPS had a bigger sample share (8.8 %) between 0 and 4 years old than the ENPE (8.4 %). The sample shares are higher for the ENPE than the TLMPS for the 15–29 age groups. The TLMPS had similar sample shares as the ENPE for individuals of aged 30–74. The TLMPS captures a slightly larger proportion of adults 75+.

Figure 2 displays the distribution of household sizes reported in the ENPE 2013 and the TLMPS 2014. The mean household size for the ENPE was 3.9, while the TLMPS produced a mean of 3.6. This relates to our earlier observation that expanding by the number of households did not result in the full national individual population, and individual roster weights must be used. The TLMPS reported a bigger proportion of households of one and two members than the ENPE did (10.4 % TLMPS and 7.2 %

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 16 of 21

Fig. 1 Comparison of age structures between ENPE 2013 and TLMPS 2014. Source: Authors' calculations based on TLMPS 2014 and ENPE 2013

ENPE for one member; 20.1 % TLMPS and 16.6 % ENPE for two members). The ENPE had 2.4 percentage points more for households with four members than the TLMPS sampled. Small differences favoring larger houses in the ENPE are found for households of 5+.

We now focus on comparing demographic and labor market characteristics for the working age population (aged 15–64) as sampled by the 2013 ENPE and the 2014 TLMPS. As seen in Table 4, the samples show strong similarities. In terms of sampled gender, the ENPE 2013 was 51 % female as was the TLMPS 2014. Similarly, 68 % of individuals had urban residence in the ENPE 2013 compared to 69 % in the 2014 TLMPS. As far as educational levels sampled, the TLMPS and ENPE sampled within a percentage point of each other for having no, primary, secondary, or higher education. As far as employment status is concerned, the TLMPS 2014 sampled a bit differently from the ENPE 2013 with 42 % employed in the TLMPS versus 43 % in the ENPE and 7 % of the
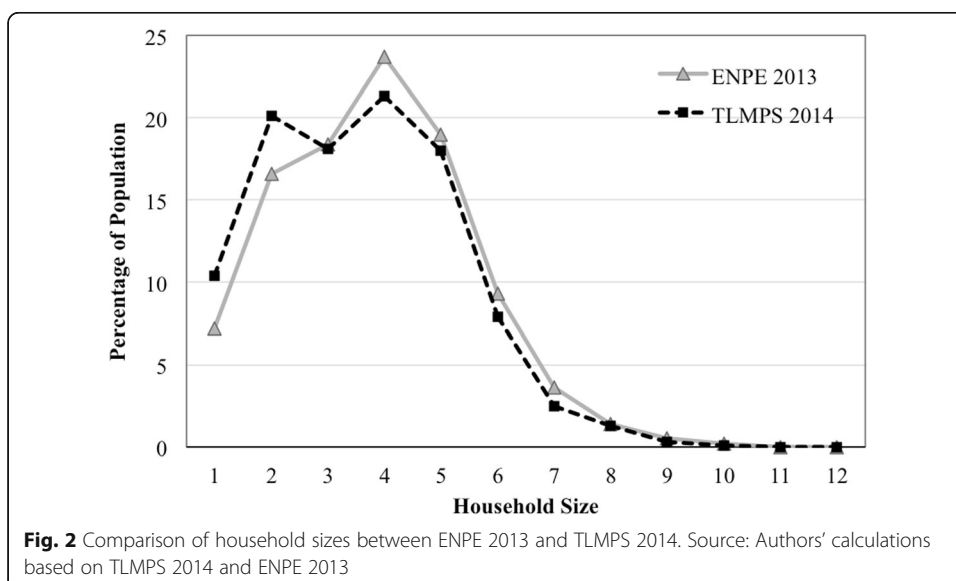

Fig. 2 Comparison of household sizes between ENPE 2013 and TLMPS 2014. Source: Authors' calculations based on TLMPS 2014 and ENPE 2013

Table 4 Distributions for working age population (age 15–64) ENPE 2013 and TLMPS 2014

	ENPE 2013	TLMPS 2014
Gender		
Male	49.3	49.0
Female	50.7	51.0
Residence		
Urban	68.3	68.8
Rural	31.7	31.2
Education level		
None	14.1	15.0
Primary	32.1	32.6
Secondary	38.0	37.6
Higher education	15.7	14.8
Employment status		
Employed	43.3	41.7
Unemployed	8.3	6.6
Inactive	48.4	51.7
Marital status		
Single	42.7	40.7
Married	54.0	56.5
Widow/er	2.2	1.8
Divorced	1.0	1.1

Source: Authors' calculations based on TLMPS 2014 and ENPE 2013

population unemployed in the TLMPS versus 8 % in the ENPE; thus, the shares of inactive were also slightly different. Finally, marital status in both surveys showed some minor differences in sampling. Around 41 % of those sampled in the TLMPS 2014 were single compared to 43 % in the ENPE 2013. However, these differences in proportion were less apparent for widowers and divorcees with 2 and 1 %, respectively, in both surveys.
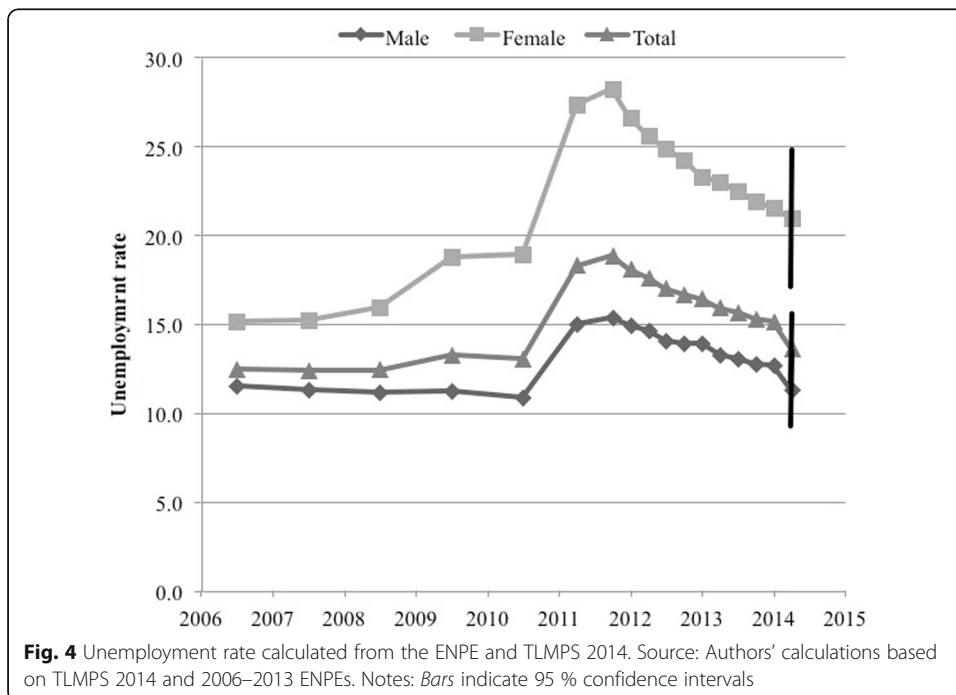
4.2 Comparisons with time trends

A particularly important focus of the survey is providing accurate information on individuals' labor force status. In this sub-section, we compare our key estimates of labor market indicators, specifically labor force participation, employment, and unemployment rates, with the time trends in these indicators from the ENPE. We also provide confidence intervals around our estimates as an important element of comparisons. We plot the trend of the labor force participation rate between 2006 and the first quarter of 2014 in Fig. 3. Starting in 2011, the data became quarterly rather than annual. The second quarter of 2014 estimate is from the TLMPS.

The estimates produced from the TLMPS 2014 followed the trend displayed by the ENPE. The male labor force participation rate was slightly above the trend while labor force participation was slightly below the trend for females. The estimate in the first quarter of 2014 for females was within the 95 % confidence interval of the TLMPS estimate. The total labor force participation rate from the TLMPS followed the trend from the ENPE more closely, and the 95 % confidence interval encompassed the previous quarter ENPE estimate.

Fig. 3 Labor force participation calculated from the ENPE and TLMPS 2014. Source: Authors' calculations based on TLMPS 2014 and 2006–2013 ENPEs. Notes: *Bars* indicate 95 % confidence intervals
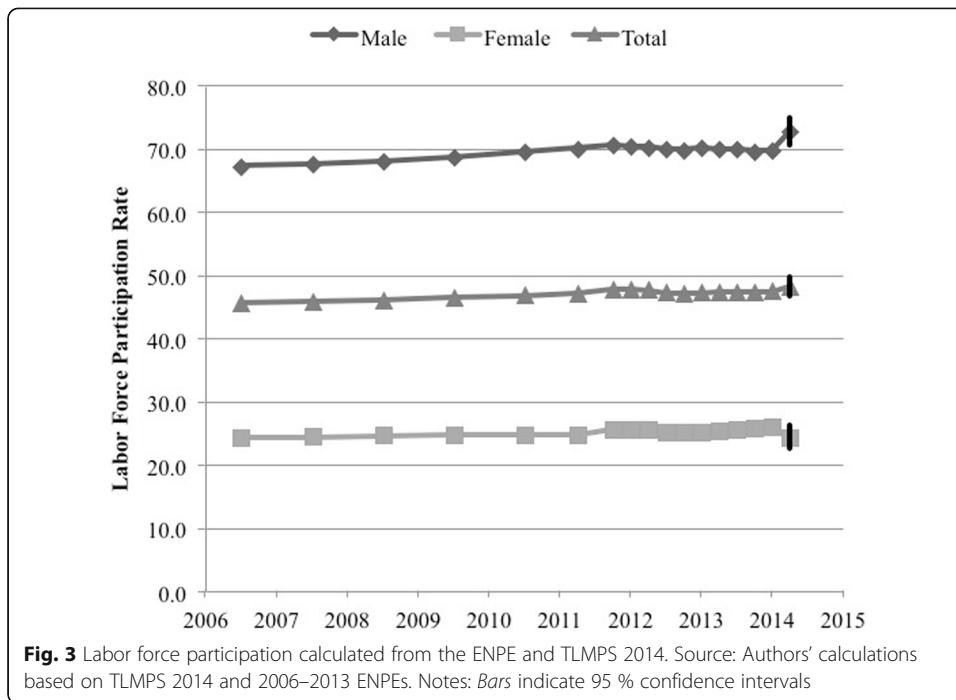
Likewise, Fig. 4 shows the unemployment rate from the ENPE and the TLMPS 2014. The unemployment rate from the TLMPS for females followed the trend for the ENPE quite closely. For males, however, the TLMPS measure was slightly below the previous quarter but within the 95 % confidence interval and following the trend. As for the total between both groups, the TLMPS 2014 estimates was slightly below the trend but the trend once again fell within the 95 % confidence interval.

Fig. 4 Unemployment rate calculated from the ENPE and TLMPS 2014. Source: Authors' calculations based on TLMPS 2014 and 2006–2013 ENPEs. Notes: *Bars* indicate 95 % confidence intervals

Assaad *et al. IZA Journal of Labor & Development* (2016) 5:15

Page 19 of 21

Finally, we compared the employment rate for the TLMPS 2014 to the trend from the ENPE (Fig. 5). The male employment rate from the TLMPS was slightly above the trend and outside the 95 % confidence interval. For females, the TLMPS employment rate was below the trend but the previous quarter was within the confidence interval, while the total employment rate for both groups was slightly above the trend but the previous quarter fell within the 95 % confidence interval.

5 Conclusions

In making cogent analyses of labor markets, and social phenomena in general, employing rich, high-quality data is vital. This rings especially true for Tunisia where the availability of rich household survey data is limited. It is this limitation that we hope the TLMPS 2014 will help alleviate, at a time when an accurate understanding of Tunisia's economic and social challenges is particularly vital. Through the public release of the TLMPS, researchers will be able to undertake new and more complex analyses of the Tunisian labor market. In addition to labor market research, the TLMPS 2014 data can be usefully exploited for other research areas, such as education, migration, demography, and life course studies. Research in MENA as a whole will have a new asset, particularly since ERF will be releasing an integrated (harmonized) version of the TLMPS 2014 along with the ELMPS 1998, 2006, and 2012, the 1988 Egypt special Labor Force Sample Survey, and the 2010 JLMPS.

This paper has described the vast potential of the TLMPS to be used as a data source for topics of social science research ranging from labor market dynamics to fertility decisions. While there were challenges in data collection, we have illustrated their nature and how to account for them to generate credible results. As with the TLMPS, the ELMPS and JLMPS are available through the ERF online data portal (www.erfdataportal.com) and follow a similar design. Future rounds of all these

Fig. 5 Employment rate calculated from the ENPE and TLMPS 2014. Source: Authors' calculations based on TLMPS 2014 and 2006–2013 ENPEs. Notes: *Bars* indicate 95 % confidence intervals
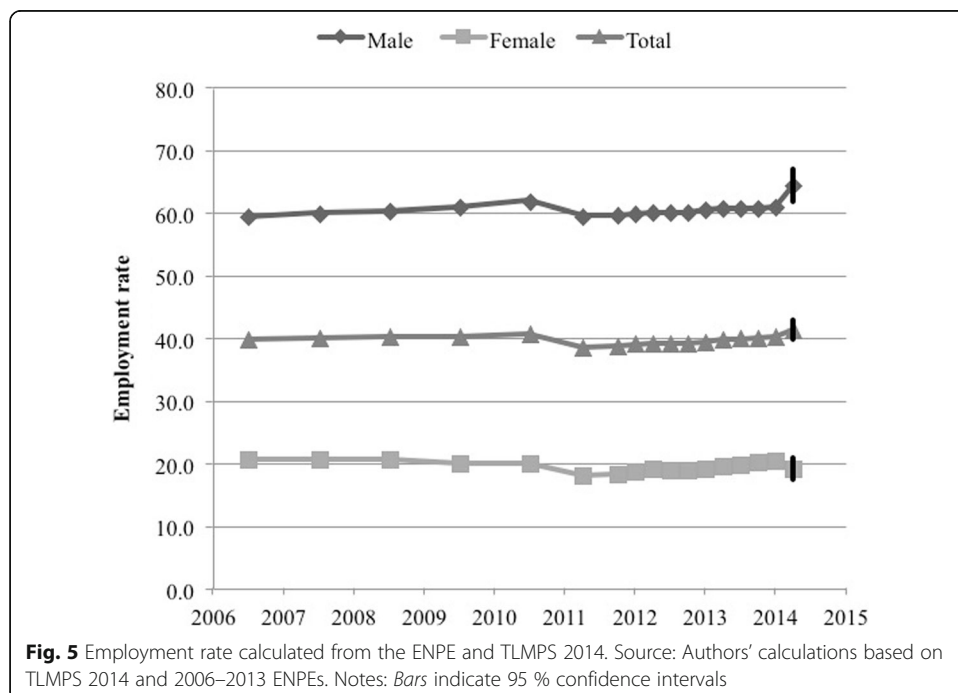
surveys are planned, including a 2016 round of the JLMPS, a 2018 round of the ELMPS, and follow-up on the TLMPS in 2020. It is our intent to continue collecting, organizing, and distributing similar Labor Market Panel Surveys (LMPS) for these three countries, in order to contribute to the potential for meaningful social science research in MENA.

Endnotes

[1]E.g., Assaad and Krafft 2015b; Assaad 2002, 2009, 2014; Belhaj Hassine 2011; Gatti et al. 2014; Herrera and Badr 2013; Sieverding 2012; Silva et al. 2013; UNDP and Institute of National Planning 2010; Yassin 2015.

[2]The TLMPS is publicly available through the ERF Open Access Micro-Data Initiative (OAMDI) through the ERF data portal: http://www.erfdataportal.com/index.php/catalog

[3]Tunisia has 24 governorates, but the governorates of Tunis and Monastir do not have a rural component.

[4]We have determined that moving from first job forward in time rather than current job backward yields better measures of dynamics, by comparing rounds of ELMPSs that used these different methodologies (Assaad et al. 2016e).

[5]The specific operating system environment for the software employed was Android 4.2.

[6]The households with updated data were only those for whom telephone numbers were available and for whom there were substantial problems in data collected during their original face-to-face visits.

[7]Mr. Yamen Helel at INS provided initial weights and documentation.

[8]When a group appeared in the census but not in the TLMPS sample (for instance, if the TLMPS sampled no 55–59-year-old women in rural Tozeur), the census population for that group was assigned to the nearest age group of the same sex and location.

Author details
[1]Humphrey School of Public Affairs, University of Minnesota, 301 19th Ave. S, Minneapolis, MN 55455, USA. [2]University of Manouba, Ecole Supérieure de Commerce de Tunis (ESCT), Campus Universitaire de Manouba, 2010, Manouba, Tunisia. [3]University of Tunis-El Manar, Faculté des Sciences Economiques et de Gestion de Tunis (FSEGT), Laboratory of Research "Prospective, Stratégies et Développement Durable" (PS2D), Campus Universitaire El Manar, B.P. 248 - El Manar II, 2092 Tunis, Tunisia. [4]Department of Economics, St. Catherine University, 2004 Randolph Ave, Saint Paul, MN 55105, USA. [5]Department of Applied Economics, University of Minnesota, 1994 Buford Ave, Saint Paul, MN 55108, USA.

References
Assaad R (ed) (2002) The Egyptian labor market in an era of reform. American University in Cairo Press, Cairo, Egypt

Assaad R (ed) (2009) The Egyptian labor market revisited. American University in Cairo Press, Cairo, Egypt

Assaad R (ed) (2014) The Jordanian labour market in the new millenium. Oxford University Press, Oxford, UK

Assaad R, Krafft C (2013) The Egypt labor market panel survey: introducing the 2012 round. IZA J Labor Dev 2:1–30

Assaad R, Krafft C (2015a) Is free basic education in Egypt a reality or a myth? Int J Educ Dev 45:16–30. doi: 10.1016/j.ijedudev.2015.09.001

Assaad R, Krafft C (eds) (2015b) The Egyptian labor market in an era of revolution. Oxford University Press, Oxford, UK

Assaad R, Krafft C (2016) Labor market dynamics and youth unemployment in the Middle East and North Africa: evidence from Egypt, Jordan and Tunisia. Economic Research Forum Working Paper Series No. 993. Economic Research Forum, Cairo, Egypt

Assaad R, Saleh M (2015) Does improved local supply of schooling enhance intergenerational mobility in education? Evidence from Jordan. Toulouse School of Economics Working Papers No. TSE-549. Toulouse School of Economics, Toulouse

Assaad R, Ghazouani S, Krafft C (2016a) The composition of labor supply and unemployment in Tunisia (forthcoming). Economic Research Forum Working Paper Series. Economic Research Forum, Cairo, Egypt

Assaad R, Ghazouani S, Krafft C (2016b) Marriage, fertility, and women's agency and decision making in Tunisia (forthcoming). Economic Research Forum Working Paper Series. Economic Research Forum, Cairo, Egypt

Assaad R, Krafft C, Roemer J, Salehi-Isfahani D (2016c) Inequality of opportunity in income and consumption in Egypt. Economic Research Forum Working Paper Series No. 1002. Economic Research Forum, Cairo, Egypt

Assaad R, Krafft C, Roemer J, Salehi-Isfahani D (2016d) Inequality of opportunity in income and consumption: the Middle East and North Africa Region in comparative perspective. Economic Research Forum Working Paper Series No. 1003. Economic Research Forum, Cairo, Egypt

Assaad R, Krafft C, Yassin S (2016e) Comparing retrospective and panel data collection methods to assess labor market dynamics. Economic Research Forum Working Paper Series No. 994. Economic Research Forum, Cairo, Egypt

Belhaj Hassine N (2011) Inequality of opportunity in Egypt. World Bank Econ Rev 26:265–295. doi: 10.1093/wber/lhr046

Ben-Ayed O, Lahmar H, Kammoun R (2016) Class-time utilization in business schools in Tunisia. Int J Educ Dev 47:86–96. doi:10.1016/j.ijedudev.2016.01.001

Broecke S (2013) Tackling graduate unemployment in North Africa through employment subsidies: a look at the SIVP programme in Tunisia. IZA J Labor Policy 2:1–19

Gatti R, Angel-Urdinola DF, Silva J, Bodor A (2014) Striving for better jobs: the challenge of informality in the Middle East and North Africa. World Bank, Washington, DC

Haouas I, Sayre E, Yagoubi M (2012) Youth unemployment in Tunisia: characteristics and policy responses. Top Middle East North African Econ 14:395–415

Herrera S, Badr K (2013) Heterogeneity in returns to investment in education in Egypt. Middle East Dev J 5:1–43. doi:10.1142/S179381201350020X

Krafft C, Alawode H (2016) Inequality of opportunity in higher education in the Middle East and North Africa. St. Catherine University, St. Paul, MN

Krafft C, Assaad R (2015) Promoting successful transitions to employment for Egyptian youth. Economic Research Forum, Cairo, Egypt

Milovanovitch M (2014) Trust and institutional corruption: the case of education in Tunisia. Edmond J. Safra Working Papers, No. 44. Harvard University, Edmond J. Safra Center for Ethics, Cambridge, MA

Ministry of Development and International Cooperation, National Institute of Statistics, UNICEF (2013) Survey of the situation of children and women in Tunisia—multiple indicator cluster survey 2011-2012: final report (French).

Premand P, Brodmann S, Almeida R, et al (2012) Entrepreneurship training and self-employment among university graduates: evidence from a randomized trial in Tunisia. World Bank Policy Research Working Paper No. 6285. World Bank, Washington, DC

Rijkers B, Arouri H, Freund C, Nucifora A (2014) Which firms create the most jobs in developing countries? Evidence from Tunisia. World Bank Policy Research Working Paper No. 7068. World Bank, Washington, DC

Sieverding M (2012) Gender and generational change in Egypt. University of California, Berkeley

Silva J, Levin V, Morgandi M (2013) Inclusion and resilience: the way forward for social safety nets in the Middle East and North Africa. World Bank, Washington, DC

Stampini M, Verdier-Chouchane A (2011) Labor market dynamics in Tunisia: the issue of youth unemployment. Rev Middle East Econ Financ 7:1–35. doi:10.2202/1475-3693.1394

UNDP, Institute of National Planning (2010) Egypt human development report 2010. UNDP and Institute of National Planning, Egypt

World Bank (2014) The unfinished revolution: bringing opportunity, good jobs and greater wealth to all Tunisians. World Bank, Washington, DC

Yassin S (2015) Labor market search frictions in developing countries. Universite Paris I - Pantheon Sorbonne, Paris