

**A Pre-Modern Middle-Eastern Population Brought to Light:
Digitization of the 1848 and 1868 Egyptian Individual-Level Census Records**

Mohamed Saleh*

University of Southern California

(May 2011: Preliminary and incomplete)

Abstract

This paper describes my digitization project of two nationally representative samples of the 1848 and 1868 Egyptian censuses from the original manuscripts at the National Archives of Egypt, and introduces an application of the data in the field of Egyptian economic history. Our knowledge about pre-modern Middle Eastern societies has been limited by the lack of data. The 1848 and 1868 Egyptian individual-level census records provide two detailed snapshots of the Egyptian population in its early attempts to make the transition into a modernized society. Carried out during the reigns of Muhammad Ali (1805-1848) and Ismail (1863-1879) respectively, these censuses are perhaps the earliest in the Middle East and among the earliest in any non-Western country to include information on all segments of society including females, children, and slaves, and on a wide range of demographic and socioeconomic variables.

Keywords: Egypt historical census; Egypt modern economic history; historical demography; Muhammad Ali; Middle Eastern economic history.

* The author is a PhD candidate at Department of Economics, University of Southern California (E-mail: msaleh@usc.edu). I sincerely thank Dora Costa and Jeffrey Nugent for their great support and advice. I gratefully acknowledge the financial support for the project that I received from Minnesota Population Center, Economic History Association, and University of Southern California. I also gratefully acknowledge the logistic support of the National Archives of Egypt that was crucial to undertake the data collection project. I sincerely thank Dr. Saber Arab and Dr. Rifaat Hilal for allowing me to undertake the data collection project and for providing me with the personnel and the equipments. I thank Dr. Emad Hilal, Hisham Shibl, Julie Iskander, and all the members of the data entry team for their valuable support and cooperation, without which the project would not have come to fruition. I sincerely thank Ragui Assaad for his great help. I also thank Naomi Lamoreaux, Leah Boustan, Steven Ruggles, Robert McCaa, Roger Owen, Khaled Fahmy, Kenneth Cuno, Dimitris Pipinis, and the participants at the UCLA Economic History Proseminar and the International Commission of Historical Demography Conference for their valuable comments and suggestions. All errors are mine.

I. Introduction

“It is therefore necessary for Us to enumerate exactly the people of Our country so that it may be a cause of its progress in civilization,” Muhammad Ali Pasha (1845) (Cuno and Reimer 1997, 213, italics mine)

“The conducting of the census is a matter of benefit for the homeland, and whoever understands its merit will strive body and soul for the sake of its execution,” Census Order (1847) (Cuno and Reimer 1997, 215, italics mine)

In 1816, Muhammad Ali Pasha, the autonomous Ottoman viceroy of Egypt, initiated an ambitious manufacturing program in what constituted one of the earliest state-led development experiments in the world. The early program, which failed in the end to transform Egypt into an industrial nation, was so intriguing to historians that it remained one of the most central topics in the social and economic history of modern Egypt and the Middle East at large. Nevertheless, despite the vast historical literature on this topic (El-Gritli 1952; M. Fahmy 1954; El-Hitta 1967; Marsot 1984; Owen 2002), little is known with respect to the size of the workforce in these early Egyptian factories, let alone the occupational, ethnic, age, or religious distributions of the workers. Such detailed picture, which is indispensable to evaluate the social impact of the manufacturing experiment on the populace, was unattainable either because the data did not exist at all, or even if they did exist they were too sporadic to form a complete picture of the topic.

In this context, the Egyptian individual-level census records from 1848 and 1868 offer two extraordinary data sources on the Egyptian population in this period. By virtue of the wide range of information contained in the census records, one is able to form a detailed picture of the social impact of Muhammad Ali’s state-led development, as well as of numerous other central topics in Egyptian and Middle Eastern social and economic history, such as education and slavery, that are

currently not fully understood because of the lack of data. The two censuses, that were carried out during the reigns of Muhammad Ali (1805-48) and Ismail (1863-79) respectively, are preserved in 6,592 handwritten registers at the National Archives of Egypt (henceforth, NAE). They include information on a wide range of variables including geographic location, name, ownership of dwelling, type of dwelling, relationship to household head, age, gender, religion, ethnicity, nationality, place of origin, occupation, and place of work, among other variables. The data are recorded systematically for every individual in the household including females, children, and slaves.

In an international perspective the 1848 and 1868 Egyptian census records are perhaps the oldest modern censuses in the Middle East and among the earliest from any non-Western country. Several salient features make these two censuses “modern” and distinguish them from the earlier Ottoman enumerations (in 15th and 16th centuries) or the other contemporary Ottoman censuses during the period: First, while the target population of the Ottoman enumerations was the taxpaying population or the population eligible for military conscription (adult males), the Egyptian census records enumerated all segments of the population including females, children, and slaves. Second, in the same spirit of contemporary Western censuses, the Egyptian census records kept standardized individual-level lists and not a mere “count” of heads (unlike the earlier Ottoman enumerations). Third, the two Egyptian censuses recorded socio-economic information (such as school enrollment and occupation for males) besides demographic ones (such as location, gender, age, and place of origin). This shows an interest on part of the Egyptian government in the “human capital” of the population.

The digitization of these individual-level census records is significant for researchers in various fields. The census records open the door for cliometricians to employ quantitative methods in

studying Egyptian and Middle Eastern economic and social history at the micro- or individual-level. So far, the study of history of this region has been mostly based on qualitative methods, generally because of the lack of data. While qualitative research is indispensable in understanding various historical phenomena in the Middle East, the introduction of quantitative methods will definitely enrich the historical research and will allow tackling a number of research questions that cannot otherwise be addressed.

In addition to this, little is known of the demographic and socio-economic characteristics of pre-modern populations in the Middle East and the Islamic World. Such knowledge is hindered by the lack of unbiased sources of information that can provide *representative* snapshots of the population. The available sources (such as tax, cadastral, and court registers) are usually confined to specific segments of the population (perhaps the wealthier strata), and are hence of limited use for the historical demographer who is interested in the overall characteristics of the population. By the inclusion of all segments of the population, the census records will allow the study of the basic demographics of this pre-modern Middle Eastern population: fertility, mortality, marriage, and immigration, as well as socio-economic conditions.

Finally, a significant feature of the census records is that they come from the pre-colonial period unlike comparable historical censuses from other developing countries that were conducted under colonial administration (e.g. British India 1872). They hence provide an early snapshot of a pre-modern and a pre-colonial population in its early attempts towards modernization. When combined with later post-colonial censuses, this source allows examining the impact of colonization on the colonized population.

This paper describes the digitization project of two nationally representative samples of the 1848 and 1868 census records, which I undertook with the help of data entry assistants at the NAE.

The project, that took place over the period 2009- 10, resulted in the creation of two datasets from the 1848 and 1868 censuses each with about 80,000 individual observations. The two datasets are currently being integrated into the historical censuses of the North Atlantic Population Project (NAPP) at Minnesota Population Center (MPC) and will be disseminated on the web, with free access to all, as soon as the integration is done. Upon their public dissemination, the two datasets will hopefully boost quantitative research on the history of the Middle East.

The rest of this paper is organized as follows. Section II describes the census registers: the historical context in which the Egyptian censuses were undertaken, the format of the census registers, and the enumeration methodology. Section III describes the sampling strategy and the details of the digitization project. Section IV discusses how the digitized samples can shed light on Muhammad Ali's manufacturing experiment, as an example for the research questions that one can tackle using this new data source. Finally, section V concludes.

II. The Census Registers: Historical Context, Format of the Registers, and Enumeration Methodology

A. Historical Context

In 1845, Muhammad Ali Pasha ordered the undertaking of a nationwide census enumerating every individual in the country (including females, children, and slaves). The census operations started in the rural provinces as early as in 1846 but were then extended to the urban ones in 1847 and were further repeated in 1848 in most of the rural units that were already enumerated over the period 1846- 47. Although a number of enumerations took place in the 1850s and the early 1860s, these remained limited to scattered villages and did not constitute a national census. Egypt had to wait for almost twenty years to have its second wide- scale census in 1865-69

during the reign of Khedive Ismail (1863-79). That second census is incomplete in the sense that it did not cover the entire country. The result of these two census operations was the creation of standardized lists of individual-level census records kept in 6,592 hand-written registers (the vast majority of which belong to the years 1848 and 1868) that were since then preserved at the NAE.

Before the discovery of the 1848 and 1868 censuses in Egypt, the series of Egyptian modern censuses was believed to have started in 1882 (a few months before the British occupation). The 1882 census was widely known because it resulted in a published census report at the village/urban quarter-level although the microdata did not survive¹. The second published census was conducted in 1897, and was then followed by almost regular decennial censuses (1907, 1917, 1927, 1937, 1947, 1960, 1966, 1976, 1986, 1996, and 2006). However, the microdata for the censuses over the period 1882-1976 were destroyed either by choice or by chance. This resulted in the anomalous situation that the only available census microdata in Egypt are either the very early ones from 1848 and 1868 or the most recent ones (1986, 1996, and 2006).

Within the Middle East, the 1848 and 1868 censuses are perhaps the oldest censuses to contain an enumeration of females, children, and slaves. On the one hand, the earlier Ottoman tax registers (*Tahrir Defterleri*) in the 15th and 16th centuries contained an enumeration of only the adult male population with very basic information provided since the main concern of the government was the tax-paying population (Barkan 1957; Cosgel 2004; Hütteroth and Abdulfattah 1977). On the other hand, while the Egyptian censuses were the counterparts of the

¹ There are a number of census registers at the NAE that date back to 1879-82. These registers include individual-level census records as well as preliminary tabulations and might represent part of the largely destroyed microdata of the 1882 census.

censuses undertaken in other parts of the Ottoman Empire from the 1830s onwards (Cuno and Reimer 1997), two features of the Egyptian censuses make them distinct from the Ottoman ones: first, they contained an enumeration of females and children, and second, they contained detailed information on occupation of all males and school enrollment of male children. The contemporary Ottoman censuses lacked these features. According to Karpat (1978), the first Ottoman governmental census took place in certain parts of the empire in 1831-38, but was an enumeration of households rather than individuals. The second census took place in 1844 and was an enumeration of the adult male population. Finally, the third census (1866-73) that was limited to the Danube province was also an enumeration of adult males. In fact, the first census in the Ottoman Empire that contained an enumeration of females took place in 1881/82-1893 (Karpat 1978).

These comparisons naturally lead to the questions: Why did the Egyptian government conduct these censuses? What were the motives behind them? Was the Egyptian government attempting to imitate the experiences of European countries? Answering such questions is in no way definite; nevertheless, it may be useful to speculate on the answers. Indeed, one important incentive for conducting the 1848 and 1868 censuses was “controlling” the individuals. The census lists provided the basis for tax collection, military conscription, and corvée work for public projects just like earlier and contemporary Ottoman enumerations, and they also allowed the government to effectively control internal migration between villages (which was not normally allowed). Nevertheless, enumerating females and children and including socio-economic variables such as school enrollment cannot be explained by purposes of “controlling” the population, and might have been motivated by other concerns about fertility and the potential size of the labor force. Here, one may speculate, following Alleaume and Fargues (1998), that

the Egyptian students who were sent on educational missions to Europe may have played a role in introducing the techniques of the European censuses into Egypt. One may also speculate that the European technicians (mostly French and Italian) whom Muhammad Ali and his successors employed in order to provide advice to the administration may have also played a role. At any rate, the census orders suggest a *conscious* attempt to imitate the experiences of the more advanced nations.

B. Format of the Census Registers

The census registers are organized by province, with each province having a number of registers enumerating its population. In urban provinces (Cairo, Alexandria, and Rosetta), the registers are further classified by district (which are in turn divided into urban quarters or *shiyakhas*)². In rural provinces, the registers are classified into provincial towns, villages, hamlets, and Arab tribes. Besides the aforementioned urban and rural provinces, there are two single-city provinces that have one register each: Al-Arish (in Sinai Peninsula) and Al-Qusayr (at Red Sea coast).

The format of recording in urban provinces is on a dwelling-by-dwelling basis with the heading of each page usually stating the urban quarter and the street name³. For each dwelling, the dwelling type (e.g. ruined hut, tenement house) is recorded, followed by the legal status of the dwelling (e.g. private ownership, religious endowment), the name of the property right holder, and the dwelling number in the street⁴. The households residing in the dwelling are then recorded one by one. Each household starts with the word “residence of” followed by the name of the household head, and the breaks between the households are usually clearly marked. The

² Rosetta is treated as one single district and is hence directly divided into urban quarters.

³ See Appendix A for two scanned pages of the registers.

⁴ I found a governmental order that dates back to 1847 with respect to naming Cairo’s streets and numbering its dwellings that is published in Sami (1928), 2: 547-552. The order might have been related to the ongoing census operations.

household members are then recorded: males followed by females, and within each gender free individuals are followed by servants (who are also free) and then followed by slaves.

In rural provinces, the census records of the provincial towns, villages, and hamlets are classified into: (i) locals, (ii) strangers, who are further divided into legal and illegal immigrants⁵, and (iii) deserters, or those who escaped illegally from the geographic unit. The locals are recorded according to the section (*Hissa*) of the geographic unit they are residing in. The household of the *Hissa*'s headman (*sheikh*) is always recorded first followed by the other households. Strangers and deserters, on the other hand, are classified geographically according to their place of origin (for strangers), or place of destination if known (for deserters). In large provincial towns, the census records are classified by urban quarters and street names (if available), and are hence more similar to the recording in the urban provinces. Finally, Arab tribes⁶ registers classify the members of the tribe by the specific tribal group (*gama'a*) they belong to and/or by their location. No information is recorded on the dwelling's legal status in rural provinces, perhaps because private landownership in rural areas was not yet fully recognized by the state (Cuno 1992).

For every individual in the census, a systematic list of information is usually recorded that includes the following variables:

1. *Name*: Egyptian naming system uses the first name of the individual followed by the name of the father, the paternal grandfather, and so on. The names in the records are usually recorded up to the father's name. Some individuals have their names recorded up to the grandfather's name.

⁵ To control peasants' flight, immigration from one village to another required a governmental permission starting from 1829. People were "illegally" migrating from one village to another in order to avoid paying the excessive taxes, the military conscription, and corvée in public works (Cuno 1992; K. Fahmy 1998).

⁶ For Arab tribes, the recording starts with the household of the headman of the tribe followed by the other households in the tribe. The registers usually state the tribe's name and its location.

Females' first names in urban provinces are usually *not* recorded unless the female is the household head.

2. *Relationship to household head*: This is recorded in detail for most individuals. Some confusion might occur in large rural households where the relationship recorded for an individual might be relating the individual to the household head or to the preceding household member.

3. *Age*: is recorded in full years (as of the next birthday) for every enumerated individual. Age is categorical (with only two options: child and adult) for females in urban provinces. Age for infants is sometimes recorded in months or even days.

4. *Occupation*: This is recorded mainly for males. The occupational titles are pretty detailed and reflect the highly specialized guild system that was prevalent at the time. I found about 3,700 distinct occupational titles in each census⁷. For male children, school enrollment is recorded as an occupation. Unemployed and retired personnel are recorded as such. The name of the work establishment is usually recorded if the person is employed in governmental enterprises. Occupation in rural localities is often missing indicating the default occupation: farming. Similarly, for Arab tribes, occupation is often missing indicating perhaps that they are shepherds.

5. *Nationality*: This includes two categories: inside or outside government's control. The concept of Egyptian nationality seems to have been quite developed in the censuses as nationals of other regions of the Ottoman Empire are recorded as "outside government's control", indicating that they are foreigners.

6. *Ethnicity*: This is usually missing for those inside government's control who are ethnically Egyptians. Sometimes categories such as "Son of an Arab" or "Son of the country" are used to

⁷ The occupational titles "scribe" and "scribe at the customs department" are counted as distinct.

describe these individuals. For the individuals outside government's control, and for slaves (who are mostly inside the governments' control), the ethnicity is usually recorded (e.g. Turkish, Armenian).

7. *Religion*: This includes the following categories: Muslim, Christian, and Jew. Often the denomination within Christianity and Judaism is recorded (e.g. Coptic Christian, Rabbi Jew). Also, the denomination can often be inferred upon combining religion with ethnicity (e.g. Armenian Christian).

8. *Place of origin*: It is recorded up to the village- level, but is *not* exactly the place of birth. In 1848 census, children inherit the place of origin of their father. The wife does *not* take the place of origin of her husband, and it seems that she takes the place of origin of her father instead⁸.

9. *Infirmities*: An infirmity is indicated if the individual has any (e.g. blind).

10. *Marital Status*: is also often recorded for females living with no adult males present in the household.

However, despite the modern features of the Egyptian census records, an element of trial-and-error is also evident. The standardized lists are often not so standard, with some information omitted or mentioned irregularly. Information on religion and nationality are often omitted so long as the "default" category is assumed (e.g. "Muslim" and "Inside government's control"). Confusion on part of the census takers over some fields can be inferred. For instance, age, which is a categorical variable (child or adult) for females in Cairo and Alexandria according to census instructions (Cuno and Reimer 1997), is sometimes recorded in full years for females in these cities. In both censuses, there was often some confusion on whether there should be an age

⁸ In the 1868 census, many children are recorded as from Cairo while their fathers are from another province indicating that the concept of this variable was getting closer to the "place of birth".

restriction when recording the occupation for males or not⁹. Spelling mistakes are numerous, and each scribe seems to have, to some extent, his *own* way of recording information, some being keener on recording details than others. The records also reflect the preliminary stage of the statistical operations at the time. Apart from preliminary tabulations on the age and occupational distributions at the street- or village section (*Hissa*)- level, the census operations in both years did not result in national statistical tabulations in published census reports perhaps due to the unsophisticated techniques the census takers were using.

C. Enumeration Methodology

Enumeration in the censuses seems to have largely followed the “de facto” principle. This means enumerating individuals according to their actual, rather than permanent, place of residence: A person who is not present in his permanent place of residence at the time of enumeration is recorded in that place as “not enumerated”, and is enumerated instead in his actual location¹⁰. Servants and slaves are recorded as part of the household they are residing with. An exception to this principle appears in not enumerating the individuals who are “protected” by foreign consulates. These individuals, although they should have been enumerated according to the de facto principle, and who are mostly (but not always) foreigners in the modern sense of the word, are recorded at the household head level only with the note that they are to be enumerated by their foreign consulates. Another important exception to this principle is military personnel (officers and soldiers) and students in public schools (which were military boarding schools in nature) who were recorded as not enumerated with the note that they are to be enumerated in the

⁹ The occupation for males who are less than 6 years old is sometimes left blank. In other cases, the word “child” or “infant” is recorded in place of the occupation.

¹⁰ The “not enumerated” note is either mentioned explicitly in words or by assigning a “zero” beside the name of the non-enumerated individual meaning that he/she is not added to the enumeration count of the page. A person who is not enumerated usually has only his/her name recorded (and occupational title for the military personnel). Age and other information are normally not mentioned for non-enumerated individuals.

census of the army. The members of their households, however, were recorded and enumerated. Finally, Cairo's notables in the census of 1848 (but fortunately not in 1868) were not enumerated in their actual place of residence and were instead recorded separately in a special register¹¹. The information on these notables in 1848 was scarce and only mentioned the household head's name and the number of males and females residing in his/her household. Notables seem to have been a broadly defined group constituting about 6% of 1848 Cairo's population (Cuno and Reimer 1997).

The 1848 census operations were explicitly mentioned in the census orders of 1845 and 1847 (Sami 1928; Cuno and Reimer 1997). Noticeably, there was no "census form" and the census operations were conducted by the headmen of urban quarters and villages' sections, under the supervision of an officer from the army. There were threats by harsh punishments for the concealment of individuals. Each enumerated individual was issued a "ticket" indicating that he/she has been enumerated. For the 1868 census, the only order that I found so far is very brief and refers to some decree issued by the parliament (Sami 1928). However, the 1868 census operations seem to have been very similar to those of 1848 as far as one can infer from the format of the census registers.

III. Digitization Project and Sampling Strategy

The digitization project of the census records consisted of the creation of a 1% sample of each of the 1848 and 1868 census records (5% in the two major cities, Cairo and Alexandria). I undertook this project over the period 2009-10 with the help of data entry assistants at the NAE.

¹¹ According to the census order (1847) (Cuno and Reimer 1997), each quarter's headman had to send to the Prefect a list of the notables in his urban quarter. The Prefect was to request counts of males and females in the household from the household head, which were to be recorded in the notables' register. This way, the census takers and the quarters' headmen in charge of the census operations could not intervene in the enumeration of the notables or access their houses. Similarly, the *Khedivial Diwan* was responsible for collecting information on foreigners from the consulates.

The result was the creation of two representative samples (about 80,000 individuals each) from the years 1848 and 1868, which are currently being integrated into the NAPP datasets and will be publicly disseminated once the integration is complete.

In what follows, I will first discuss the target population and the sampling frame. Second, I will discuss the sampling mechanism in detail. Third, I will compare the sampling strategy to the U.S. 1850 census sample constructed by IPUMS, and to the unreleased and uncompleted sample of the 1848 Egyptian census described in Alleaume and Fargues (1998) and Fargues (2003).

A. Target Population and Sampling Frame

I define the target population for each of the two samples as the population that was residing in Egypt at some point (the census night) in 1848 or 1868. The sampling frame, however, can be different and I define it as all the available census records in the *surviving* registers from both years. The discrepancy between the two concepts is due to two problems: First, the missing registers (non-coverage problem), and second, the multiple enumerations of some geographic units over the periods 1846-48 and 1865-69 (duplicates problem). In this subsection, I will address each of these two problems in turn. I want to emphasize, however, that the solutions to the two problems discussed in this subsection are the “pre-sampling” solutions that have to do with re-defining the target population and/or the sampling frame. Post-sampling solutions (i.e. creating sample weights) are discussed in section III.B.

1. Missing Registers¹²: Some census registers are missing. They might have been destroyed or are undiscovered yet. With the exception of the cities of Damietta, Suez, and Al-Qusayr (in 1868), which are entirely missing, the surviving urban provinces have almost complete registers

¹² The problem of missing pages within a surviving register is also possible. However, it is generally very limited in both censuses. In such cases, I assumed that the missing pages are random and I oversampled from the surviving pages.

in both years. In order to evaluate the completeness of the surviving registers of the rural provinces in the 1848 and 1868 censuses and the extent to which the surviving census records represent the population of rural Egypt in these two years, I use the first published census of 1882 as the reference point. The 1882 census has a complete list of the geographic units in Egypt to which one can compare the available list from 1848 and 1868. Table (1) shows this comparison at the province-, district-, and village- levels. At the province- level, all the 1882 rural provinces are represented in 1848. Within the “surviving” provinces, Al-Gharbiya and Isna have the most serious missing registers problem at the village-level (coverage rate is 7% and 14% respectively). In 1868, however, the problem seems to be more severe, since 6 out of 14 rural provinces are entirely missing, and 3 provinces (Al-Minya, Girga, and Al-Sharqiya) have extremely low coverage rates. At this point I cannot confirm whether the 1868 census was a nationwide census whose registers did not survive or rather a partial one that was not completed for some unknown reason.

Although the aforementioned comparison gives an idea of the degree of coverage up to the village- level within each province, it does not tell how complete the *individual* records are within each province. After all, there might have been missing individual census records due to under-enumeration. To have an approximate idea on the extent of completeness of the individual-level records in each province (both urban and rural), I again use the 1882 census as the reference point under the assumption that under-enumeration was less of a problem in that year. In particular, in order to find an estimate of the true population size of each province, I calculate the “true” population of each province in 1848 and 1868 based on its share of the total population in 1882 census (I will discuss the validity of the assumptions underlying this calculation in section III.B). I then compare this “estimated” population size to the size of the

population that has been *actually* enumerated in the census registers; the latter I obtained by summing up the population counts from all the geographic units in a given province¹³. In order to do this comparison, a priori information on the size of the entire population in Egypt in both years is required. For 1848, I used the widely accepted figure of 4,476,439 (Alleaume and Fargues 1998). For 1868, I assumed a constant annual growth rate over the period (1848-1882) in order to calculate the population in 1868 (knowing the population in 1848 and 1882). Table (2) shows this comparison for the two censuses. Overall, the results are similar to the coverage rates in table (1), although the enumeration counts in a number of provinces actually exceed the estimated populations especially in 1848. Nevertheless, these provinces have relatively high coverage rates in table (1).

[Insert tables (1) and (2)]

The problem of missing registers results in discrepancy between the target population and the sampling frame. I chose to treat this problem in the pre-sampling phase as follows: If an entire province listed in the 1882 census is missing in the 1848 or 1868 census (or has only a very small number of its units represented in the registers¹⁴), I re-defined the target population to exclude this province. If, on the other hand, a province listed in the 1882 census has an arbitrarily significant number of surviving registers in the 1848 or 1868 census, I included its registers in the sampling frame, and oversampled from them. Whether the sampling frame *represents* the

¹³ Logistic problems at the NAE did not allow me to collect the enumeration counts for the 1868 census. I instead estimated the actually- enumerated population in each province in 1868 using the total number of pages for each province (adjusted for blank pages and tabulations), and the average number of individuals per page in the sample taken from that province.

¹⁴ Besides the entirely missing provinces, I namely excluded Damietta in 1848, and Al-Minya and Girga in 1868. Damietta (by 1882 boundaries) had only one surviving district in 1848 that was actually enumerated in that year as part of another neighboring province, Al-Daqahliya, and I thus sampled it as part of the latter province. So Damietta was not sampled as an independent province. I decided to sample Al-Sharqiya in 1868 although its coverage rate is pretty low (1% at the village-level and 6% at the individual-level) because it has 20 census registers, which is (arguably) a relatively large number.

target population in this province depends on the validity of the assumption that the missing registers are randomly distributed in that given province. In the post-sampling phase, I addressed this point by adjusting the sampling weight (Subsection III.B).

2. Multiple Enumerations: The enumerations for the 1848 census started as early as in 1846, but were repeated in most of the geographic units in 1847 and 1848 perhaps in order to raise the accuracy of enumeration and to mitigate under-reporting of individuals. Similarly for the 1868 census, there were early enumerations in 1865-67, and enumeration operations seem to have been extended to 1869 in some provinces. In order to keep the consistency between the target population and the sampling frame, and to avoid double counting of individuals, only the register(s) from the latest year of each geographic unit was included in the sampling frame. For instance, if a village was enumerated in 1846, 1847, and 1848, I included only the register(s) from 1848 in the sampling frame. This assumes that (i) the register(s) from the earlier year(s) enumerate exactly the same population of the unit, and, (ii) if there are more than one register from the latest year that they enumerate *different* segments of the population and thus that they should be added up. In specific provinces in the 1848 census, and due to logistic problems, I discovered the existence of duplicate registers for the same unit *after* sampling started, and hence very few units had higher chance of selection. To correct for these duplicates, I adjusted their sampling weights in the post-sampling phase (Section III.B). Table (3) shows the number of units with multiple enumerations (in different years) in each province for each census. As is evident from the table, the problem is prevalent in very few provinces, and is more prevalent in the 1848 census.

[Insert Table (3)]

B. Sampling Mechanism

1. Explicit Stratification by Province: I started sampling in each census by explicitly stratifying the country into the provinces that are included in the sampling frame. In order to obtain an estimate of the population size of each stratum, I used the relative share of each province in 1882 census and the a priori information on the total population size of the country in 1848 and 1868. This calculation is based on two assumptions: (i) that the administrative boundaries between provinces did not change over the period, and (ii) that the distribution of the population in 1848 and 1868 across provinces is the same as the distribution in 1882. As for the first assumption, it suffices to notice that only 2.5% of the geographic units of 1848 (less than 1% in 1868) belonged to a different neighboring province in 1882. The second assumption may be justified for the distribution *across* the rural provinces, but might be questionable for the urban-rural distribution with presumably an increasing share of the urban provinces over time due to urbanization and foreigners' migration to Cairo and Alexandria in the second half of the 19th century as well as the growth of the Suez Canal cities. However, table (2) showed that the actual enumeration counts in the urban provinces are not much different from the estimated sizes of strata.

Setting the sampling rate at 5% in Cairo and Alexandria and at 1% in all the other provinces, I calculated the targeted sample size in each stratum (province). Two notes are in order: First, I chose to *explicitly* stratify the sample by province. By explicit stratification I mean that I calculated the targeted sample size in each province, and I sampled independently from within each province to achieve that sample size. The other option that I did not choose, namely implicit stratification, was to apply the fixed sampling rate of 5% on Cairo and Alexandria as one stratum, and 1% on all the other provinces as another stratum, without the need for setting a

targeted sample size in each province¹⁵. In other words, under implicit stratification I would have taken one page every 20 pages in Cairo and Alexandria and one page every 100 pages in all the other provinces. My choice of explicit stratification by province is actually motivated by the desire to make the geographic distribution of the sample approximately reflect the *true* geographic distribution of the target population at the province- level. In particular, given the missing registers problem, implicit stratification would have resulted in a geographic distribution of the sample that reflects that of the surviving registers, which is not of significance *per se*.

Second, I calculated the provinces' sizes using the 1882 census shares because I could not find reliable *a priori* information on the 1848 or 1868 total population in each province. The closest approximation is to use the sum of the enumeration counts from all the available geographic units in each province. This, however, will result in the true stratum size only if there is no missing registers problem or no under- (or over-) enumeration. Since this is not always the case, it was perhaps safer to use the 1882 shares in order to calculate the sizes of the strata. Table (4) shows the geographic distribution of the two samples along with the targeted and actual sampling rates.

[Insert Table (4)]

2. Sampling Rules: I applied systematic sampling by page within each stratum. Given the targeted sample size, knowing the total number of pages of the registers of each province, and by assuming an average number of individual records per page, I calculated the total number of pages required in each stratum to reach the targeted sample size and hence the *interval* of the systematic sample. I selected a page randomly in the beginning of each stratum, and then I took

¹⁵ This is not to be confused with explicit stratification between Cairo and Alexandria on the one hand, and the rural provinces on the other hand, which was essential under both scenarios discussed in the text since I applied different sampling rates between these two groups.

the successive pages according to the calculated interval. A selected page is accepted in the sample if at least 75% of it has individual-level records. If a page is rejected, the closest preceding or succeeding page, which satisfies the 75% rule, is taken on an alternating basis. If a page is to be taken in the sample, I recorded all the households that *start* in the page along with information on the dwelling they are residing in (including the total number of households and individuals in the dwelling)¹⁶. I also recorded the information on the register code and page number to facilitate referring to the original register if need arises in the future. In applying this rule, if a household starts in a previous page and continues on the sample page it is not included in the sample. Similarly, if a household starts on the sample page and continues in the following page I entered it in its entirety. This ensures that all households have an equal chance of appearing in the sample regardless of their size. Overall, systematic sampling was less costly than a pure random sample and was much easier to apply by data entry assistants. It also ensures the geographic spread of the sample within each stratum (implicit sub-stratification within the province).

A few notes on the sampling rules are in order. First, group quarters are dwellings where individuals live together with no household relationships between them (e.g. churches, monasteries, and jail). The individuals in these group quarters are recorded in the registers as one single household. If the sample page has a group quarter *starting* in the page, I entered the group quarter as a single household in its entirety. Otherwise, the page is rejected (since it won't satisfy the 75% rule) and a preceding or a successive page is taken¹⁷. Second, fragments, which are the

¹⁶ I defined the starting point of the household as the line including the information of the first member in the household.

¹⁷ The IPUMS 1850 U.S. census sample applies individual-level sampling to individuals residing in the group quarters, instead of entering the group quarter in its entirety. The drawback of my approach is the resulting increase in the standard error because the observations in the group quarter are more likely to be correlated. Nevertheless, the

individuals enumerated separately from their place of residence or their households for some reason, are taken into the sample if they happen to be on the sample page. I entered the fragments as separate households in the same way as they are recorded in the registers, with the note that they are fragments. In 1848 Cairo, fragments are usually confined to the end of the urban district where the census takers record individuals who were away at the time of the census, but came back later on and were thus enumerated afterwards. In 1868 Cairo, however, fragments are found within the pages of the regular enumeration. In other provinces from both years, fragments are usually mentioned at the end of the geographic unit.

Finally, in a few registers in Alexandria in 1848, the households' breaks within the dwelling were not marked. In such dwellings, individuals were classified by nationality, place of origin, and gender, *regardless of* the household they belong to. Household relationships were sometimes mentioned (for males), but were mostly ignored for females. For such cases, and since it was often impossible to identify the households inside the dwelling, I decided to enter the dwelling in its entirety. In particular, if a page is selected into the sample, all the *dwellings* that start on the page are taken into the sample. Nevertheless, since most of the male household members do have common nationality and place of origin (defined in section II.B), it was mostly possible to identify the male members of each household.

Most of the variables were coded either *ex ante* or during the data entry operations. Through the data entry software, the data enterer was able to add new codes to the variables as new categories are discovered in the registers. These coded variables are: type of dwelling, legal status of dwelling, relationship to household head, title (both of dwelling's property right holder and of individual; e.g. pasha), legal status of the individual, nationality (both of dwelling's property

very small number of group quarters that I found in the registers and the fact that they are limited to Cairo and Alexandria mitigate such concerns.

right holder and of individual), ethnicity (both of dwelling's property right holder and of individual), religion, place of origin (which I coded according to the 1882 census geographic coding), and the presence of infirmities. On the other hand, numerical variables included number of households, and number of individuals in the dwelling. Finally, the variables that were entered as text are: register code, page number, street name, dwelling number, name of the dwelling's property right holder, age¹⁸, occupational title (both of dwelling's property right holder and of individual), individual's name, and census taker's (or data enterer's) notes on both the dwelling and the individual. After the data entry was complete, I coded the occupational titles according to the Historical International Standard Classification of Occupations (HISCO), which is the same coding scheme used in NAPP samples, up to the most detailed five- digits classification. This will hence make the integration of the Egyptian occupational titles to the NAPP samples easier. Examples of the individual records in the digitized samples are shown in Appendix B.

3. Creating Post-sampling Weights: In the post-sampling phase, sampling weights are required when calculating population means, totals, and proportions in order to adjust for (i) different *actual* sampling rates across provinces (table 4), (ii) non-coverage of some geographic units within the rural provinces (missing registers), and (iii) duplicate registers for the same geographic unit (multiple enumerations) which resulted in a higher chance of selection for some units. The final weight for each geographic unit is the product of these three weights. In what follows I will discuss the weight adjustment needed to correct for each of these three concerns.

(i) Adjusting for Different Sampling Rates: Defining the probability of selection as the total number of individuals in the sample divided by the total number of individuals in the population, individuals in Cairo and Alexandria have a higher probability of selection than the individuals in

¹⁸ Age is entered as text because it is categorical for females in Cairo and Alexandria. Also, for infants, the age is recorded in days or months, so I entered it as text as mentioned in the register.

the rest of the provinces, and sampling weights must adjust for that. Moreover, actual sampling operations may have resulted in different *actual* sampling rates across provinces. I use the estimated population figures, based on the 1882 relative population shares, as the denominator for the probability of selection in each province. The sampling weight (W_A) is defined as the reciprocal of the probability of selection.

(ii) Adjusting for Non-Coverage: Missing registers within urban provinces is negligible. I thus focus here on non-coverage within the rural provinces. I estimate the probability of coverage of a geographic unit at the village-level in the rural provinces using a Probit model¹⁹. The dependent variable of interest is a dummy for having a surviving register in the relevant census. The regressors of interest are the characteristics of the geographic units that are known for both missing and surviving units. These are: location of the unit (I try three specifications with region, province, and district fixed effects), type of geographic unit (village, hamlet or Arab tribe, and provincial town), and population size. I estimated the regression for each year separately. The results of these regressions are shown in table (5), and the sampling weight (W_B) is defined as the reciprocal of the estimated probability of coverage. From the table, one can see that larger units have a higher chance of survival, perhaps because they have bigger registers that are less likely to be lost. Paradoxically, provincial towns have lower chance of survival than villages, and hamlets and Arab tribes have a higher chance of survival than villages. Finally, units in Upper Egypt are more likely to have a surviving register than units in Lower Egypt.

[Insert Table (5)]

(iii) Adjusting for Duplicates: For the units that have multiple surviving registers in different years, only the register(s) from the latest enumeration was included in the sampling frame, and

¹⁹ I am grateful to Ragui Assaad, University of Minnesota, for this advice.

hence no weight adjustment is needed for these units. However, in very few cases, I discovered the existence of multiple enumerations only *after* sampling started, and hence such units had higher probability of selection. The sampling weight for these units has to be adjusted by multiplying the original sampling weight ($W_A * W_B$) by the reciprocal of the number of duplicate enumerations that were included in the sampling frame.

C. Comparison to the IPUMS 1850 U.S. Census Sample and Alleaume and Fargues 1848 Egyptian Census Sample

My sampling strategy is similar to the methods used for the 1850 U.S. census sample constructed by IPUMS (Ruggles and Menard 1995). The systematic sampling mechanism was used for both samples. In the IPUMS sample, as in my sample, sampling is done every given interval of pages. However, there are two main differences: First, while my sampling unit is the page and the households within the page, the U.S. 1850 census sample is a sample of dwellings. In particular, while the U.S. sample selects a random line from the page and a dwelling is accepted if the line falls on the first individual in the dwelling; in the Egyptian sample, and due to the irregular content of pages and the non-numbered lines, the randomization is at the page- level. Once a page is selected, *all* the households in the page are taken into the sample. Hence, clustering in the Egyptian sample is at the *page-* level while in the U.S. sample is at the *dwelling-* level. Second, stratification in the U.S. sample is implicit, since the 1% fixed sampling rate is applied to all census records so as to choose 1 individual every 100 individuals. In the Egyptian sample, the missing registers problem led to the choice of explicit stratification.

Alleaume and Fargues collected an unreleased and uncompleted sample of the 1848 Egyptian census, which is described in their 1998 article (henceforth, AF sample) and Fargues (2003). The AF sample applies a two-stage stratified sampling strategy. Initially, the country is stratified into

29 strata (Cairo's 10 districts, Alexandria's 5 districts, Damietta, other urban governorates, and the 13 rural provinces²⁰). In addition, rural provinces were further stratified into three substrata: main town, rural localities (villages and hamlets), and Arab tribes. In the first stage, a number of geographic units were selected within each stratum based on an a priori sampling rate f_1 . If the stratum consists of exactly one unit (such as a district of Cairo) it is selected with sampling rate 100%. Also, Arab tribes are treated differently and each tribe is taken with probability 100%. In the second stage, a number of households are selected within the selected units with a sampling rate f_2 . If a household is chosen, it is entered in its entirety.

My sampling strategy is different from the AF sample in the following aspects. First, in deciding the size of each stratum, the AF sample used cited population figures available for the urban provinces in 1848, and calculated the population figure for the rural provinces (as one stratum) as the residual. I instead used the 1882 population shares to estimate the population figure for each province. Second, by using systematic sampling by page within each stratum (province), I applied implicit geographic sub-stratification where the probability of selection of any geographic unit into the sample is directly proportional to the number of pages of its register (in other words, its population). The AF sample, on the other hand, applies explicit stratification within each rural province, and oversamples from the central towns as compared to the rural localities. Third, the AF sample treats the non-coverage problem by choosing close substitutes (in terms of size) to replace the missing geographic units. The AF sample also chooses close substitutes for the households if non-coverage is within the geographic unit (e.g. missing pages). I instead use sampling weights to adjust for the non-coverage problem at the geographic unit-level within a rural province. If non-coverage occurs within the geographic unit, I assumed that

²⁰ Qena and Isna were a single province in 1848 (Ramzi 1994). Hence, there are 14 rural provinces in the 1882 census but only 13 in the 1848 census.

the missing records are random and oversampled from the surviving records in order to facilitate the sampling methodology.

IV. Application: Exploring 19th Century Egyptian Modernization Program:

In this section, I first provide selected descriptive statistics from the 1848 and 1868 samples to give a general idea of the information contained in this data source. Second, I discuss how this data source can provide useful insights as to evaluating the social impact of the ambitious modernization programs carried out by Muhammad Ali Pasha, as an example for the questions that can be answered using the digitized samples.

A. Descriptive Statistics

The highest frequencies of selected variables from the 1848 and 1868 samples are shown in Table (6). I show the statistics for Cairo and Alexandria separately from all the other provinces to emphasize the interesting contrast between urban and rural provinces. Regarding the dwelling type, in both 1848 and 1868 about 73% of the individuals in Cairo and Alexandria live in houses, unspecified, or unknown dwelling types, versus 100% in the other provinces. What is more interesting, however, is the composition of the dwelling types in which the remaining 27% of Cairo and Alexandria population resided in. About 16% of the population lived in low-status dwellings. These mainly include ruined huts, courtyards, and single rooms. Moreover, about 7% lived in multiple- household dwellings such as tenement houses. Finally, 3% lived in production sites, i.e. they resided in the place where they worked (e.g. coffee shops and bathhouses).

[Insert Table (6)]

Dwelling legal status is almost only found in Cairo and Alexandria. Private or public ownership is the most dominant legal status as about 64% in 1848 and 71% in 1868 resided in privately- or

publicly- owned dwellings. This is followed by *Waqf* dwellings; i.e. dwellings owned by the pious or religious endowments of individuals or entities (whether Muslim, Christian, or Jewish).

As for the individual- level variables, gender distribution seems quite balanced and does not immediately indicate a gender bias in the sample. Muslims form the vast majority in urban and rural provinces in both the 1848 and 1868 samples. Christians are the largest religious minority followed by Jews. With respect to nationality, about 7% of the individuals are outside government's control or foreigners in the urban provinces, as compared to 1% in the rural provinces. Also, the percentage of slaves is higher in urban provinces than in rural provinces, since the majority of the slave population in Egypt (and in the Middle East at large) was involved in urban domestic service. In 1868, however, there was a slight increase in the percentage of slaves in rural provinces, perhaps due to the cotton boom that resulted from the American civil war (1861-65), and that led to a significant improvement in the living conditions of the Egyptian peasants and their financial ability to hold slaves. Finally, regarding ethnicity, the vast majority of the individuals are locals. Turks represent the second largest ethnic group in 1848, followed by blacks. The latter group, however, became the largest ethnic minority in 1868.

B. Application:

Historians have long debated various aspects of the intriguing manufacturing program carried out by Muhammad Ali Pasha, who is widely regarded as the founder of modern Egypt. Apart from the debate on the *economic* evaluation of the program with respect to efficiency in production, which requires gathering data on the costs of production and market prices, the census records can provide useful insight into the *social* impact of the manufacturing experiment on the labor force. This social face of the experiment is a central and an often ignored aspect of the manufacturing experiment. As K. Fahmy (2009, 126) points out, it is important when evaluating

Muhammad Ali's development experiment to take into account "*the millions of Egyptian people whose lot, as a direct result of what Mehmed Ali actually did achieve, was that of hardship and suffering*" (italics mine). In this subsection, I will discuss the type of information that one can extract from the census records to shed light on this social side. I will focus on two questions: First, to what extent did the manufacturing experiment by Muhammad Ali lead to the disappearance of the traditional guilds? Second, did the factories established under this experiment provide routes for upward mobility for the labor force? In discussing these two questions, however, I want to emphasize that I am not trying to *answer* the questions in this limited space. I am only giving examples to the sort of information that the census samples can provide with respect to these questions.

The first question triggered a long debate among historians. M. Fahmy (1954) argued that the manufacturing program led to the abolition of restrictions on labor that were present under the guild system, and thus to the complete liberation of the labor market. By contrast, Baer (1964) criticizes this argument and refers to the strong qualitative evidence that shows that the guild system survived until the early 20th century, and that Muhammad Ali's factories did not give the final blow to the medieval guilds, although it might have indeed hurt specific guilds. In support of Baer's thesis, Owen (2002) points out that even the textile workers who were hurt most by the factories did indeed manage to survive in the aftermath of the closure of many of the factories. In the same direction, Marsot (1984) states that the guild system remained largely unaffected by the experiment except in the occupations that were in direct competition with the factories, although she also notices that the factory system opened up possibilities for children to work and to be promoted "by training rather than by heredity" in contrast to the guild rules (Marsot 1984, 181-83). Nonetheless, a slightly different viewpoint is provided by Ghazaleh (1999), who criticizes

Baer's thesis on the grounds that although the guilds did indeed survive until early 20th century they were largely weakened by the manufacturing experiment and gradually came under state control. In response to Baer's claim that the artisanal guilds, which were essentially the most affected by the manufacturing experiment, did not constitute more than one third of the total number of guild members, Ghazaleh points to the difficulty of estimating the total number of guild members and their breakdown.

The Egyptian censuses samples can provide useful insights into answering this question. First, by having the occupational title recorded for every male, along with the name of the work establishment if the person is employed by the governmental/public sector, one is able to estimate the size of labor force in *each of* the manufacturing projects (henceforth, modern sector) in 1848 under Muhammad Ali and twenty years later in 1868 after many of Muhammad Ali's factories closed down. Second, equally important, one is able to estimate the size and breakdown of the traditional guilds outside the factories (henceforth, traditional sector). An important caveat here is that the census records include information on *occupations* rather than guild membership. Although the two variables are not the same, one may argue that the occupational title is a reasonable proxy for guild membership, given the fact that there was a high degree of specialization within the guild system, with "the extreme splitting of occupations into guilds of specialized branches" (Baer 1964, 25). At any rate, holding this caveat in mind, the census records allow us to estimate the extent to which the occupations/guilds actually disappeared due to the emergence of the modern projects, and also, to determine the specific occupations/guilds that were mostly hurt. Table (7) shows the descriptive statistics from 1848 and 1868 on the size of the labor force (adult males) that was employed in the modern sector. From the table one can see that the modern projects employed in 1848 about 11% of Cairo's adult male population (at

least 15 years old), but the percentage went down to 3% in 1868. In Alexandria, the second largest city, the percentage of workers in the modern sector was about 2% in 1848 and went up slightly to 4% in 1868. The other provinces had very low share of the population working in the modern sector in both years but the share was higher in 1848. Overall, it seems that the modern projects in 1848 were not large enough to actually lead to the complete disappearance of the traditional occupations/guilds. Moreover, the percentage of workers employed in these projects fell sharply in 1868 as compared to 1848, with the corresponding increase in the relative size of the traditional occupations/guilds. I have to emphasize, however, that Ghazaleh's argument that the guilds were *weakened* by state control even if they survived cannot be answered using the census records.

Furthermore, one can extract more detailed information that can actually help determine which guilds were hurt most by modernization. Table (8) shows the breakdown (or occupational distribution) of workers in the modern and traditional sectors in both years into ten aggregate occupational categories (same as HISCO one- digit categories). This aggregate breakdown is for presentation purposes only, and the detailed occupational categories given by the five- digit HISCO coding are available in the digitized samples. From the table, one can see that 26% of the workers in the traditional sector are working in the category "construction workers, carpenters, building painters, transportation workers, and laborers". About 20% of the workers in the traditional sector are in the service occupations "cooks, servants, slaves, policemen, military personnel, and assistants". The occupational distribution in 1868 is quite similar. Going through the detailed occupational distribution in both years, one can decide the exact occupations/guilds that were affected most by modernization in both years.

The second question has to do with the routes of occupational mobility that were provided by the modern sector in 1848 and 1868. While most of the historians admitted with varying degrees that the workers in the factories suffered from conscription into the factories and the bad working conditions they faced there (M. Fahmy 1954; Marsot 1984; Owen 2002), little is known about the occupational distribution of the workers in these factories, and how it compares to that in the traditional sector. This knowledge is crucial to evaluate the degree to which jobs in the modern sector compare (whether favorably or not) to those in the traditional sector, and hence, whether they provided routes for upward or downward occupational mobility. To the best of my knowledge, only Ghazaleh (1999) provides a list of workers broken down by occupation, their working days, and their salaries in *Khurunfish* factory, the oldest textile factory in Egypt that was constructed in 1816. While the censuses samples do not have information on the wages, they do provide detailed information on the occupation and the work establishment (factory) of each male worker in the modern sector. Table (8) shows the occupational distribution of workers in the modern sector in 1848 and 1868 in Cairo and Alexandria (the two cities with the vast majority of modern projects), and how it compares to that in the traditional sector. The vast majority of the adult male workers in the modern sector in 1848, about 69%, are in the occupational category “construction workers, carpenters, building painters, transportation workers, and laborers”. About 9% are working as specialized artisans “wood treaters, gunpowder makers, textile workers, millers, and food processors”. The white-collar workers (scribes, stores clerks, etc) constitute only 9% of the workers. In 1868, however, the distribution seemed to have shifted more towards white-collar workers (17%) and away from “construction workers, carpenters, building painters, transportation workers, and laborers” (57%). This may have been the case because of the nature of the modern transportation projects in 1868 such as railways,

telegraph, and steam navigation, which required more administrative/clerical jobs than Muhammad Ali's factories. This may suggest that the modern projects were actually a route for downward rather than upward mobility for Egyptians in 1848, but might have provided a route for upward mobility to white-collar jobs in 1868.

[Insert Table (7) and Table (8)]

V. Conclusion:

The 1848 and 1868 newly digitized census samples provide a rich source of information on the Egyptian population in mid- 19th century, which can benefit researchers in various disciplines. First, historical demographers can use the Egyptian census samples to examine patterns of fertility, mortality, marriage, and immigration in this population. Multigenerational households can also be studied using this source. Moreover, given the availability of contemporary historical census records from other countries in the NAPP database, international demographic comparisons will be feasible. Second, cliometricians and historians of Egypt and the Middle East can benefit from these digitized samples in studying slavery, modernization, and *Waqfs*, among other topics. Quantitative statements about the magnitude of each of these phenomena will be feasible. Third, genealogists can make use of these census records in studying history of families in Egypt. Fourth, the spatial data on location and addresses in the cities can benefit researchers in spatial sciences and urban history of Middle Eastern cities in reconstructing a detailed historical map of 19th century Cairo and Alexandria. Overall, the 1848 and 1868 census records open up entirely new possibilities for quantitative research in studying the social and economic history of Egypt and the Middle East.

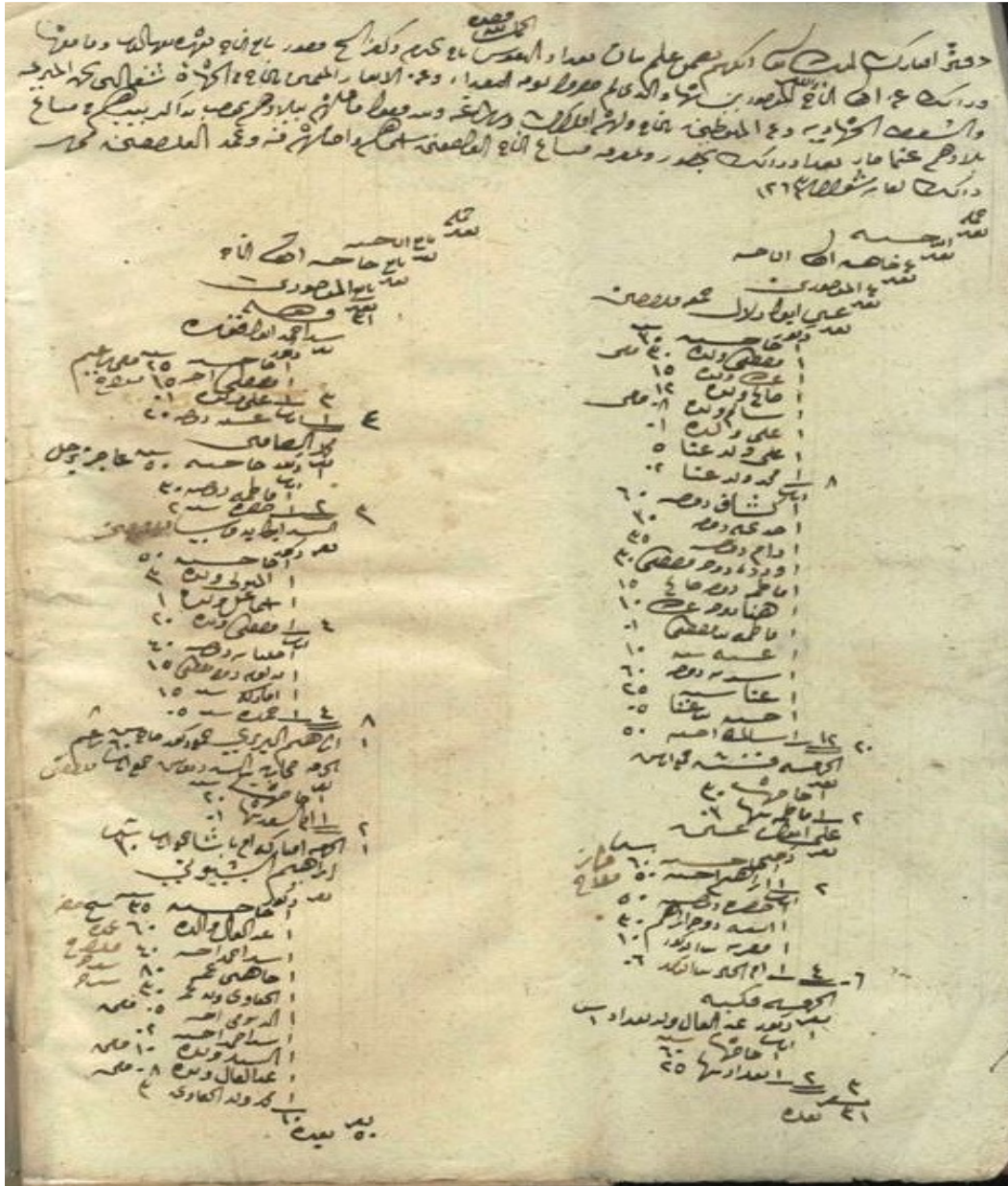
References:

- Alleaume, G. and P. Fargues. 1998. La naissance d'une statistique d'état. *Histoire & Mesure* 13: 147-193.
- Baer, G. 1964. *Egyptian Guilds in Modern Times*. Jerusalem: The Israel Oriental Society.
- Barkan, Ö. 1957. Essai sur les données statistiques des registres de recensement dans l'empire Ottoman aux XVe et XVIe siècles. *Journal of the Economic and Social History of the Orient* 1: 9-36.
- Cosgel, M. M. 2004. Ottoman tax registers (*Tahrir Defterleri*). *Historical Methods* 37: 87-100.
- Cuno, K. M. 1992. *The Pasha's peasants: Land, society, and economy in Lower Egypt, 1740-1858*. Cambridge: Cambridge University Press.
- Cuno, K.M. and M. J. Reimer. 1997. The census registers of nineteenth-century Egypt: A new source for social historians. *British Journal of Middle Eastern Studies* 24: 193-216.
- El-Gritli, A. 1952. *Tarikh al-sina'a fi Misr fil nisf al-awwal min al-qarn al tasi' 'ashar* (History of industry in Egypt in the first half of the nineteenth century). Cairo: Dar El-Ma'aref.
- El-Hitta, A. A. 1967. *Tarikh Misr al-iqtisadi fil qarn al tasi' 'ashar* (Economic history of Egypt in the nineteenth century). Cairo: Al-Masry Publishing.
- Fahmy, K. 1998. *All the Pasha's men: Mehmed Ali, his army, and the making of modern Egypt*. Cambridge: Cambridge University Press.
- Fahmy, K. 2009. *Mehmed Ali: From Ottoman governor to ruler of Egypt*. Oxford: One World.
- Fahmy, M. 1954. *La revolution de l'industrie en Egypte et ses consequences sociales au 19^e siecle (1800-1850)*. Leiden: E. J. Brill.
- Fargues, P. 2003. Family and household in mid-nineteenth-century Cairo. In *Family history in the Middle East*, edited by B. Doumani, New York: State University of New York Press.
- Ghazaleh, P. 1999. *Masters of the trade: Crafts and crafts people in Cairo, 1750-1850*. Cairo: The American University in Cairo Press.
- Hütteroth, W. and K. Abdulfattah. 1977. *Historical Geography of Palestine, Transjordan and Southern Syria in the Late 16th Century*. Erlangen: Selbstverlag der Fränkischen Geographischen Gesellschaft.
- Karpat, K. H. 1978. Ottoman population records and the census of 1881/82- 1893. *International Journal of Middle East Studies* 9: 237-274.
- Marsot, A. L. A. 1984. *Egypt in the reign of Muhammad Ali*. Cambridge: Cambridge University Press.
- Owen, R. 2002. *The Middle East in the world economy 1800-1914*. UK: I. B. Tauris & Co Ltd.
- Ramzi, M. 1994. *Al-qamus al-gughrafi lil bilad al-misriya min 'ahd qudama' al-misriyeen ila sanat 1945* (The geographic dictionary of the Egyptian lands from the time of ancient Egyptians to 1945). Cairo: General Egyptian Book Organization.

Ruggles, S. and R. R. Menard. 1995. *Public use microdata sample of the 1850 United States census of population: User's guide and technical documentation*, 7-10. Minneapolis: Social History Research Laboratory.

Sami, A. 1928. *Taqwim Al-Neel (Chronicles of the Nile)*. Cairo: Dar Al-Kutub.

2- A Page from a Rural Province (Al-Gharbiya)



First page of register of Village "Bigirim wa Kafr al-Sheikh Mansour", Al-Gharbiya Province, 1847, Preserved at the NAE

Appendix (B): Actual Examples of the Digitized Census Records

1. (Al-Mu' allem) Antonios Luqa: Male, Free, Able-bodied, HH Head, 40 years, Inside the government's control, Coptic Christian, From Abu-Tig (in Abu-Tig, Asyut), Scribe at the Customs Department in Bulaq.

Address: 15 Harat Al-Izba (From Darb Al-Geneina), Shyakhat (Quarter of) Youssef Alam, (District of) Al-Azbakiya, (Province of) Cairo.

Type of Dwelling: House Waqf (Religious Endowment) of Copts.

2. Salman Abdel-Rehim: Male, Free, Able-Bodied, HH Head, 45 years, Inside the government's control, Muslim, From Tukh (in Qena, Qena), Farmer.

Address: House of Salman Abdel-Rehim, Hissat Naser Eleiwa, (Village of) Tukh, (District of) Qena, (Province of) Qena.

3. Mardukh Youssef: Male, Free, Able-Bodied, HH Head, 25 years, Inside the government's control, Jew, From Cairo, Goldsmith.

Address: 35 Harat Al-Yahud Al-Qarra'een (Karaites), Shyakhat (Quarter of) Khidr Ibrahim, (District of) Bab El-Shiriyya, (Province of) Cairo.

Type of Dwelling: House owned by the heirs of Ibrahim Khidr.

Table (1): Evaluating Coverage in the Rural Provinces in 1848 and 1868

1882 Province	Number of Districts in 1882	% Covered 1848	% Covered 1868	Number of Units in 1882	% Covered 1848	% Covered 1868
<u>Lower Egypt:</u>						
Al-Daqahliya	6	100	100	443	74	32
Al-Sharqiya	6	100	33	433	74	1
Al-Qalyubiya	3	100	0	163	84	0
Al-Gharbiya	10	80	0	547	7	0
Al-Minufiya	5	100	0	331	62	0
Al-Buhayra	6	100	0	304	48	0
<u>Upper Egypt:</u>						
Al-Giza	4	100	100	167	93	90
Bani Suwayf	3	100	100	168	83	31
Al-Fayyum	4	100	50	91	97	91
Al-Minya	4	100	25	267	89	1
Asyut	10	100	100	322	55	57
Girga	5	100	20	190	67	1
Qina	4	100	0	109	62	0
Isna	6	33	0	107	14	0

Provinces are defined according to the 1882 census administrative borders and division. The percentages of the 1882 districts and villages for which there exists at least one surviving register in 1848/1868 censuses are recorded in columns 3, 4, 6, and 7.

Table (2): Estimating Individual-Level Coverage Rate in Each Province in 1848 and 1868

Province	1848			1868		
	Estimated Population	Actual Population Count	Coverage Rate (%)	Estimated Population	Actual Population Count	Coverage Rate (%)
Cairo	255,978	237,000	93	320,360	258,898	81
Alexandria	158,021	119,788	76	197,766	154,876	78
Rosetta	13,233	18,187	137	16,562	2,257	14
Al-Arish	2,005	2,311	115	2,509	1,820	73
Al-Qusayr	1,659	3,024	182	-	-	-
Al-Daqahliya	397,720	336,402	85	497,752	144,489	29
Al-Sharqiya	259,766	279,918	108	325,100	20,384	6
Al-Qalyubiya	173,804	178,232	103	-	-	-
Al-Gharbiya	627,764	42,287	7	-	-	-
Al-Minufiya	439,745	126,287	29	-	-	-
Al-Buhayra	270,526	144,583	53	-	-	-
Al-Giza	187,672	217,513	116	234,874	257,038	109
Bani Suwayf	133,099	190,176	143	166,576	44,424	27
Al-Fayyum	145,444	173,931	120	182,025	121,448	67
Al-Minya	201,507	238,457	118	-	-	-
Asyut	390,438	247,104	63	488,638	424,512	87
Girga	350,928	192,465	55	-	-	-
Qena and Isna	413,234	219,320	53	-	-	-

Provinces in this table are defined according to: (i) the 1882 census borders and division for the geographic units that were found in the 1882 list, and (ii) the 1848/68 censuses borders for the units that were not found in the 1882 list. Estimated population of each province (which is an estimate of the true population size) is computed based on the province's share in 1882 population census. Actual population of each province is the total enumeration count from the census registers. In 1868, I estimated the actual counts based on the total number of pages of the registers of each province and the average number of individuals per page. Coverage rate is the ratio of actual population to estimated population.

Table (3): Duplicates: Geographic Units that Were Enumerated More than Once by Province

Province	Duplicates in 1848	Duplicates in 1868
Cairo	0	0
Alexandria	0	0
Rosetta	0	0
Al-Arish	0	0
Al-Qusayr	0	-
Al-Daqahliya	185	0
Al-Sharqiya	118	0
Al-Qalyubiya	0	-
Al-Gharbiya	15	-
Al-Minufiya	7	-
Al-Buhayra	2	-
Al-Giza	1	2
Bani Suwayf	1	9
Al-Fayyum	1	0
Al-Minya	9	-
Asyut	0	21
Girga	0	-
Qena and Isna	0	-

Provinces in this table are defined according to: (i) the 1882 census borders and division for the geographic units that were found in the 1882 list, and (ii) the 1848/68 censuses borders for the units that were not found in the 1882 list. A duplicate is defined as the geographic unit that has more than one register belonging to *different* years over the periods 1846-48 (for the 1848 census) and 1865-69 (for the 1868 census). Under this definition, a unit that has more than one register that all belong to the same year is *not* a duplicate.

Table (4): Sampling in 1848 and 1868: Geographic Distribution by Province, and Targeted and Actual Sampling Rates

1848/1868 Province	1848				1868			
	Target Population	Actual Sample Size	Target Sampling Rate	Actual Sampling Rate	Target Population	Actual Sample Size	Target Sampling Rate	Actual Sampling Rate
Cairo	255,978	20,635	0.05	0.08	320,360	33,285	0.05	0.10
Alexandria	158,021	16,061	0.05	0.10	197,766	23,617	0.05	0.12
Rosetta	13,233	448	0.01	0.03	16,562	513	0.01	0.03
Al-Arish	2,005	51	0.01	0.03	2,509	26	0.01	0.01
Al-Qusayr	1,659	175	0.01	0.11	-	-	-	-
Al-Daqahliya	397,720	6,374	0.01	0.02	497,752	5,039	0.01	0.01
Al-Sharqiya	259,766	3,012	0.01	0.01	325,100	3,257	0.01	0.01
Al-Qalyubiya	173,804	3,908	0.01	0.02	-	-	-	-
Al-Gharbiya	627,764	7,369	0.01	0.01	-	-	-	-
Al-Minufiya	439,745	5,661	0.01	0.01	-	-	-	-
Al-Buhayra	270,526	3,135	0.01	0.01	-	-	-	-
Al-Giza	187,672	3,509	0.01	0.02	234,874	2,590	0.01	0.01
Bani Suwayf	133,099	1,456	0.01	0.01	166,576	1,568	0.01	0.01
Al-Fayyum	145,444	1,489	0.01	0.01	182,025	2,403	0.01	0.01
Al-Minya	201,507	2,264	0.01	0.01	-	-	-	-
Asyut	390,438	4,309	0.01	0.01	488,638	6,117	0.01	0.01
Girga	350,928	3,540	0.01	0.01	-	-	-	-
Qena and Isna	413,234	4,212	0.01	0.01	-	-	-	-

Provinces are defined according to the 1848/68 censuses administrative borders and division. Target population is calculated based on the province's population share in the 1882 census. Actual sample size is the number of individuals that were actually included in the sample in each province according to the province's definition given above. Target sampling rate is the *a priori* planned sampling rate. Actual sampling rate is the ratio of the actual sample size to the target population in each province.

Table (5): Probit Model for Probability of Coverage in Rural Provinces in the 1848 and 1868 Censuses

	1848			1868		
	Dependent Variable: Indicator for having at least one surviving register in the 1848/1868 census					
log(population)	0.138*** (0.021)	0.314*** (0.025)	0.353*** (0.029)	0.046* (0.027)	0.144*** (0.040)	0.166*** (0.049)
Provincial Town	-0.132 (0.165)	-0.160 (0.195)	-0.222 (0.219)	-0.184 (0.196)	-0.029 (0.289)	0.115 (0.347)
Hamlet or Arab Tribe	1.261*** (0.143)	1.652*** (0.172)	-	1.319*** (0.201)	0.974*** (0.219)	-
Lower Egypt	-0.400*** (0.044)	-	-	-1.037*** (0.052)	-	-
Province Fixed Effects	No	Yes	No	No	Yes	No
District Fixed Effects	No	No	Yes	No	No	Yes
N	3,959	3,959	3,336	3,671	2,115	1,440
Pseudo R ²	0.046	0.297	0.283	0.139	0.424	0.408
Log Likelihood	-2,475.95	-1,823.27	-1,547.17	-1,516.80	-765.83	-575.57
LR Chi- Squared	236.36 (<i>p</i> =0.000)	1541.74 (<i>p</i> =0.000)	1218.45 (<i>p</i> =0.000)	487.94 (<i>p</i> =0.000)	1126.29 (<i>p</i> =0.000)	791.89 (<i>p</i> =0.000)

Standard errors are in parentheses. The observations are the combined list of geographical units of the rural provinces in the 1848 (or 1868) census and the 1882 census. Dependent variable takes one if the geographic unit has at least one surviving register in 1848/1868 census and zero otherwise. Regressors include: (i) log(population) measured (a) using the 1882 census population shares for the units that existed in the 1882 census, and (b) using the census registers actual enumeration counts for the units of 1848/1868 censuses that were not matched in the 1882 census. (ii) Type of the geographic unit with "village" being the base category. The other two types are provincial town and hamlet or Arab tribe. The type is measured using the recorded type in the 1882 list for the units that existed in this list, or using the 1848/68 census types for the units that did not exist in the 1882 list. (iii) Lower Egypt region, province, or district fixed effects. The regression is estimated for each year separately.

Table (6): Selected Descriptive Statistics from the 1848 and 1868 Samples

	1848		1868	
	Cairo and Alexandria	Other Provinces	Cairo and Alexandria	Other Provinces
<u>Dwelling Type</u>				
<i>House/unspecified/unknown</i>	73%	100%	73%	100%
<i>Low-Status dwellings</i>	16%	0%	17%	0%
<i>Multiple-Household dwellings</i>	7%	0%	6%	0%
<i>Production sites</i>	3%	0%	3%	0%
<u>Dwelling Legal Status</u>				
<i>Unspecified</i>	25%	99%	15%	100%
<i>Ownership (public or private)</i>	64%	1%	71%	0%
<i>Waqf (religious endowment)</i>	11%	0%	14%	0%
<u>Gender</u>				
<i>Male</i>	49%	50%	51%	50%
<u>Religion</u>				
<i>Muslim</i>	89%	92%	86%	91%
<i>Christian</i>	6%	6%	4%	7%
<i>Jew</i>	1%	0%	1%	0%
<i>Unspecified</i>	4%	2%	9%	2%
<u>Nationality</u>				
<i>Outside government control</i>	7%	1%	7%	1%
<u>Legal Status</u>				
<i>Slave</i>	2%	0%	3%	1%
<u>Ethnicity</u>				
<i>Local</i>	87%	97%	87%	97%
<i>Turkish</i>	3%	1%	2%	0%
<i>Black</i>	1%	0%	5%	2%
<i>European</i>	1%	0%	1%	0%
<i>Levantine</i>	1%	0%	1%	0%
<i>Nubian</i>	1%	0%	1%	0%
Sample Size	36,509	43,519	56,902	21,513

Table (7): Sectoral Distribution of Adult Male Workers in 1848 and 1868 Egypt

	1848			1868		
	Cairo	Alexandria	Other Provinces	Cairo	Alexandria	Other Provinces
% Adult males in the modern sector	9.92	2.2	1.64	2.79	3.61	0.32
N	6,086	4,735	9,034	9,986	6,446	4,018

Sample is restricted to males who are at least 15 years old with non-missing occupational title.

Table (8): Occupational Distribution of Workers in the Modern Sector by Year

	1848		1868	
	Modern	Traditional	Modern	Traditional
Engineers, physicians, pharmacists, ships' masters	1%	1%	4%	1%
Judges, agents, teachers, religious workers, artists	0%	5%	0%	6%
Administrative and managerial workers	2%	0%	2%	0%
Scribes, financiers, stores clerks, customs clerks, post clerks	11%	5%	20%	5%
Merchants, street sellers, auctioneers, water porters, slaves traders	1%	8%	0%	13%
Cooks, servants, slaves, policemen, military, assistants	2%	15%	8%	18%
Farmers, animal husbandry workers, fishermen	0%	35%	0%	14%
Wood treaters, gunpowder makers, textile workers, millers, food processors	10%	11%	5%	14%
Shoe makers, blacksmiths, goldsmiths, silversmiths, potters, stone cutters	5%	4%	6%	6%
Construction workers, building painters, transportation workers, laborers	68%	17%	55%	22%
N (Males at least 15 years old with non-missing occupational title)	856	18,999	525	19,925