CAHIER DE RECHERCHE #2302E
Département de science économique
Faculté des sciences sociales
Université d'Ottawa

WORKING PAPER #2302E
Department of Economics
Faculty of Social Sciences
University of Ottawa

Monitoring poverty in a data deprived environment: The case of Lebanon[*]

Paul Makdissi[†], Walid Marrouch[‡], and Myra Yazbeck[§]

March 2023

[*] The first author would like to thank Karim Emile Bitar for useful discussion on the Lebanese political system. This is a substantially modified version of HCEO Working Paper 2022-014.
[†] Department of Economics, University of Ottawa, Canada; Email: paul.makdissi@uottawa.ca
[‡] Department of Economics, Lebanese American University, Lebanon; Email: walid.marrouch@lau.edu.lb
[§] Department of Economics, University of Ottawa, Canada; Email: myazbeck@uottawa.ca

Abstract

This paper is motivated by the dearth of statistical capacity in the Middle East and North Africa region and the unprecedented economic collapse in Lebanon. We expand and apply a novel data augmentation technique to conduct poverty analysis when the usual data sources on income distribution are limited or unavailable. Building on available data augmentation techniques, we recover the continuous income distribution from the available data when the income variable takes the form of intervals. We expand existing techniques to derive dominance conditions for interval income data accounting for non-response. This extension allows us to run robustness checks of our empirical results by estimating the bounds of the set of admissible cumulative distribution functions. Our empirical application then analyzes poverty dynamics using first-order dominance tests on the bounds of admissible cumulative distribution functions sets and shows the importance of the proposed approach using Lebanese data. The empirical application provides a picture of poverty dynamics and insights into the politico-economic dynamics preceding and following the economic collapse. More generally, we show that the development analyst can exploit alternative data sources to conduct the much-needed poverty analysis.

Key words: *Poverty dynamics, stochastic dominance, data deprivation, Lebanon.*

JEL Classification: I31, I32, O15, O53.

1 Introduction

Household budget surveys are essential for monitoring poverty and developing poverty reducing-policies (Ferreira *et al.*, 2016). However, such surveys may not always be available to researchers in regions where poverty-reducing interventions are needed the most. It is well-known that there is a correlation between a country's statistical capacity and its level of economic development.[1] However, many countries in the Middle East and North Africa (MENA) region have a statistical capacity that is substantially below their level of development, which presents clear cases of outliers.[2] This lack of statistical capacity prevents many countries in the region from developing much-needed evidence-based policies and is likely slowing their economic development. Recently, Arezki *et al.* (2020) assessed the impact of the lack of data transparency on economic growth in the MENA region. Their findings suggest that since 2005 this lack of data transparency imposed a yearly average loss of between 7% and 14% GDP per capita. This lack of transparency is a logical consequence of what El Haddad (2020) coins as the *unsocial contract* characterizing the economies of the Arab region in which the state uses economic policies to favor a group of crony capitalists. In such a context, it is essential, from a development planning perspective, to find alternative ways of documenting the evolution of poverty and inequality in these economies.

Building on these findings, Atamanov *et al.* (2020) advocate improving the statistical capacity in the MENA region. Unfortunately, because of the reality of the *unsocial contract*, international institutions and researchers aiming at improving this statistical capacity may face the resistance of policymakers. For this reason, Atamanov *et al.* (2020) suggest that researchers look for alternative information sources in parallel to those produced by countries' statistical agencies.[3] In other regional contexts, researchers have used remote sensing

[1] See https://databank.worldbank.org/reports.aspx?source=Statistical-capacity-indicators.

[2] See for example: https://www.arabdevelopmentportal.com/whatwedo

[3] One of the suggestions they made was to use non-traditional mobile phone surveys as in Hoogeveen *et al.* (2014).

data and satellite imagery to estimate poverty in the absence of surveys on income and expenditures (see van der Weide *et al.*, 2022; Jean *et al.*, 2016). In the Arab context, Harb and Rouhana (2020) and Harb (2022) collect their primary data by interviewing workers in the Beirut region to study gender and sectarian inequalities in the distribution of earnings. However, in the context of Arab countries, researchers did not explore enough the use of all the publicly available data. Extending existing methodologies may allow researchers to use these available, albeit non-obvious, data sources to extract the maximum information on poverty and inequality.

This paper proposes extending an existing data augmentation methodology to allow poverty analysis without (or lack of access to) household budget surveys using alternative data sources. It also provides an empirical application that substantiates the relevance and importance of the proposed approach. In the empirical application, we exploit publicly available data in the Arab Barometer surveys as an alternative source of information on income distribution in a context where access to continuous income data is restricted. More specifically, we focus on the publicly available data, namely income intervals for 2016 and 2018 and income in 2021-22, to present a clear picture of the evolution of poverty and social welfare in Lebanon during that period. To that effect, we first adopt and extend a data augmentation method for the analysis of interval data proposed by Groß *et al.* (2017), Walter and Weimer (2018), and Walter (2019). The original method consists of reconstructing the density function underlying the interval observations in the survey. Our extension aims at reconstructing the underlying cumulative distribution function. Unlike the density function, the bounds on the cumulative distribution function in the presence of non-response are well-defined. Therefore, estimating the underlying cumulative distribution allows us to adapt the methodology to estimate bounds on the set of admissible cumulative distribution functions to account for survey non-response. This extension to the methodology allows us to conduct the much-needed poverty analysis and thus depict the dynamics of poverty.

Our paper's motivation is rooted in the widespread issue of data poverty in the MENA region in general and the recent economic history of Lebanon in particular. Specifically, the empirical application's choice is driven by the lack of evidence-based policy in a country going through an unprecedented crisis, Lebanon. The case of Lebanon is sad, but it constitutes an interesting example from an empirical perspective. Among the thirteen countries of the MENA region, Lebanon ranks ninth in terms of its statistical capacity (see the Statistical Capacity Indicator (SCI) ranking in Table 1). In addition, since October 2019, Lebanon has been experiencing one of the most severe economic crises since the mid-nineteenth century (World Bank, 2021). While the Central Administration of Statistics (CAS) of Lebanon has some household budget surveys, these surveys are not widely available to academic researchers despite the urgent need for evidence-based poverty-reducing interventions. Given this context and the significant value and urgency of documenting the poverty situation in Lebanon, it is imperative to develop tools that allow researchers to exploit alternative sources of statistical information. Furthermore, these tools will enable researchers to better understand poverty dynamics and the state of affairs from the period preceding the crisis (i.e., before the 2019 uprisings) to the present.

To analyze the current poverty dynamics in Lebanon, we use the 2016, 2018, and 2021-22 waves of the Arab Barometer Survey. The Arab Barometer Survey is the only widely available data source that provides the necessary information to address our research questions. However, it presents two critical challenges. First, in the 2016 and 2018 waves, the income data is elicited as intervals. Second, there is a non-negligible number of non-response: 5.9% of the 1,500 observations in 2016, 3.0% of the 2,400 observations in 2018, and 15.1% of the 2,399 observations in 2021-22. To overcome the first challenge, we use a data augmentation technique proposed by Groß *et al.* (2017), Walter and Weimer (2018), and Walter (2019) to construct the cumulative income distribution. In doing so, we exploit all the available interval information on income, that is, the information on income in the first two waves (in the third wave, the income variable is continuous). To address the second issue arising from non-response in the survey, we extend Groß *et al.*'s (2017), Walter and Weimer's (2018), and Walter's (2019) methods to estimate the bounds of the sets of admissible cumulative distribution functions of income. To determine poverty dynamics, we test for first-order dominance on the bounds of the sets of admissible cumulative distribution functions in the spirit of Fakih *et al.* (2022).

Our empirical results show a noticeable reduction in poverty levels between 2016 and 2018. This reduction is compatible with the hypothesis of a political attempt to please the electoral base and ensure the re-election of the incumbent politicians in the 2018 general election.[4] Our results also show that this decrease in poverty levels was short-lived because the poverty levels spiked between 2019 and 2021-22, shortly after the elections. This increase in poverty counterbalanced the poverty reduction observed between 2016 and 2018, leading to higher poverty levels than those prevailing in 2016. These results, when combined, are compatible with the hypothesis that politicians may have been involved in a planned debt-financed Ponzi scheme to ensure their re-election.[5] However, given that the high expenditures involved were unsustainable, the Ponzi scheme collapsed as soon as they were elected. As a result, this artificial poverty reduction vanished, pushing Lebanon towards

[4]The same class of politicians was also reelected in the May 2022 election. However, for the first time, a small number of newcomers on the Lebanese political scene: https://foreignpolicy.com/2022/05/18/lebanon-election-results-parliament-hezbollah-opposition-candidates-reform/

[5]In Lebanon, an implicit tripartite alliance between the government, central bank, and banks was created to provide a funding mechanism for public debt whereby banks, and later the central bank, would buy government bonds using new deposits. These new deposits were attracted using high-interest payments financed largely by newer ones and government interest payments. The structure is precisely a Ponzi scheme. The sustainability of this scheme first came into question in 2016 when the central bank resorted to "financial engineering" to save the banking sector from collapse. Eventually, the scheme unraveled by the end of 2019 after the October uprisings and the subsequent nationwide bank run. The interested reader can refer to:

https://www.foreignaffairs.com/articles/lebanon/2022-04-18/ponzi-scheme-broke-lebanon

https://www.reuters.com/world/middle-east/lebanons-central-bank-denies-swiss-report-about-2016-imf-paper-2021-10-08/

a deep financial crisis. Unfortunately, we cannot exclusively attribute the drastic rise in poverty in Lebanon to the financial crisis because three coexisting factors (i.e., the economic collapse, the August 2020 Beirut port explosion, and the COVID-19 pandemic) may have also influenced the poverty dynamics. Nevertheless, the drastic increase in poverty levels in the last few years is an alarming result that speaks to the urgency of closely investigating the poverty situation in Lebanon.

The approach proposed in this paper is widely applicable and is valuable beyond our empirical application's specific case.[6] Indeed, these methods can be used in any survey where the income intervals information is available. Moreover, producing results using this approach may be a good persuasion tool for statistical agencies reluctant to share their data with academic researchers. The remainder of this paper runs as follows. First, section 2 describes the estimation and stochastic dominance testing strategy. Next, section 3 presents the political context and provides information on the available data. Section 4, applies the estimation and testing approach to the Arab Barometer data for Lebanon. Finally, Section 5 presents a brief conclusion.

2 Measurement framework

Our objective is to monitor the evolution of poverty in a highly fragile governance environment, and a context where continuous income data is not always available. In general, when income is continuous, one can compute additive poverty indices, $P(F_Y;z)$:

$$P(F_Y;z) = \int_0^z p(y;z)\mathrm{d}F_Y(y), \tag{1}$$

where y represents income, z, a poverty threshold, and F_Y, the cumulative distribution (CDF) of income. For a household with income y and for a given poverty threshold z, the function $p(y;z)$ denotes the household's contribution to total poverty. This function should

[6] We provided an empirical application on Lebanon because of the much-needed empirical evidence given its current economic and political context.

be such that $p(y;z) = 0$ if $y \geq z$ and $p(y;z) \geq 0$ if $y < z$. In addition, one needs to assume that an increase in an individual's income cannot increase total poverty, i.e. $\partial p(y;z)/\partial y \leq 0$. A widely used class of indices for measuring monetary poverty is the FGT class (Foster, Greer, and Thorbecke, 1984). The FGT is an additive poverty index for which a household's contribution to total poverty $p(y;z)$ is defined as follows: $p(y;z) = [(z-y)/z]^{\alpha}$. The parameter $\alpha \geq 0$ reflects aversion to poverty with higher values for α reflecting a higher aversion to poverty.

Comparing poverty levels between two income distributions (e.g., F_{Y_0} and F_{Y_1}) using the FGT class of indices is a common practice. However, the conclusion derived when estimating the FGT indices will depend on its mathematical form and the specified poverty threshold. Thus, using another index and/or another threshold may lead to different conclusions. Atkinson (1987) offered a solution to this issue by showing that if one wishes to use a more robust approach, it is possible to check if there is a reduction of poverty when moving from F_{Y_0} to F_{Y_1}, and if this reduction is valid for any poverty index and threshold $z \in [0, z^+]$. In order to identify these robust reductions in poverty, Atkinson (1987) shows that one needs to test if

$$F_{Y_0}(y) - F_{Y_1}(y) \geq 0 \quad \forall y \in [0, z^+]. \tag{2}$$

The stochastic dominance condition in equation 2 can be interpreted as follows. Suppose the CDF of interest is everywhere below the reference CDF for all income levels under z^+. This means that the proportion of the poor is lower in the distribution of interest for all potential poverty thresholds below z^+. In addition, this dominance condition's result is not limited to the headcount index, it extends to all possible indices (i.e., it provides robust orderings). In addition to allowing for the identification of robust poverty rankings, this stochastic dominance condition allows for a robust ranking of social welfare if one selects the maximum income level as z^+ (see Foster and Shorrocks, 1988; Duclos and Makdissi, 2004).

In this paper, we test the dominance condition in equation (2) using a nationally representative survey containing incomplete income distribution information and assess poverty dynamics in Lebanon. As mentioned earlier, the available data includes two types of income information. The first type is in the form of a standard continuous income variable. The second type of information takes the form of income interval variables. In both cases, we use the first-order stochastic dominance condition proposed by Atkinson (1987) and assess the directional change in poverty. We first test for stochastic dominance assuming missing value at random. We then perform a robustness check of our results and conclusions by adapting the data augmentation methodology to estimate bounds on the sets of admissible CDFs. These bounds on the set of CDFs allow checking for the robustness of the conclusion for all possible assumptions on the non-response generating process. We also complement our results by presenting estimated values of the headcount ratio and identifying upper and lower bounds for these headcount ratios for two poverty lines proposed by the World Bank: the lower-middle-income country, and the upper-middle-income country poverty lines (see Jolliffe, *et al.*, 2022)

2.1 Estimating the cumulative distributions

We first estimate the CDFs assuming missing at random for each wave of the survey. For continuous income information we have in 2021-22, we use the natural estimator for the empirical distribution function (EDF):

$$\widehat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(y_i \leq y), \tag{3}$$

where $\mathbb{1}(\cdot)$ is an indicator function, n is the number of observations, and y_i is the income of observation i.[7]

For the 2016 and 2018 survey waves, the information on income takes the form of intervals. In such cases, the direct use of the estimator in equation (3) produces a step function EDF, which is not suitable for poverty analysis. To analyze poverty or inequality, it is important to recover the underlying continuous CDF associated with the observed income distribution. One solution would be to estimate a parametric model of the CDF (see Cowell and Flachaire, 2015). An alternative approach proposed by Walter and Weimer (2018) adapts the estimation algorithm developed by Groß *et al.* (2017) to surveys with interval income data. Walter and Weimer's (2018) method relies on pseudo-samples of the y_i to estimate a density function, $f_Y(y)$. The main advantage of using this method over a parametric estimation is that, for income levels equal to an interval's bound, the algorithm produces (by construction) values of the empirical cumulative distribution function that precisely match the proportion of observations below these income levels. This means that it intersects the "step function" estimator at each jump of this step function.[8]

The main idea underlying their estimation method is relatively simple to understand.[9] Assume that we have K income intervals, denoted by $k \in \{1, 2, \ldots, K\}$. Each interval k is delimitated by income thresholds $(x_{k-1}, x_k]$. The income support is thus partitioned

[7] We account for survey design by using the following Hájek weighted estimator:

$$\widehat{F}_Y(y) = \frac{1}{\sum_{i=1}^n \omega_i} \sum_{i=1}^n \omega_i \mathbb{1}(y_i \leq 0y),$$

where ω_i is the survey weight of observation i.

[8] Assume that we have K income intervals, denoted by $k \in \{1, 2, \ldots, K\}$. Each interval k is delimitated by income thresholds $(x_{k-1}, x_k]$. If we keep the data as is, we can estimate the empirical distribution function at the bounds x_k of the intervals using:

$$\widehat{F}_Y(x_k) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(y_i \leq x_k).$$

The advantage of Walter and Weimer's (2018) approach is that their method produces a $f(y)$ that is such that if you $\int_0^{x_k} f(y)\mathrm{d}y = \widehat{F}_Y(x_k)$.

[9] A complete algorithm is in the appendix

as $\{[0, x_1], (x_1, x_2], \ldots, (x_{K-1}, x_K]\}$. From Bayes' theorem we know that the conditional distribution

$$f_{Y|K}(y|k) = \begin{cases} \frac{f_Y(y)}{\Pr(k)} & \text{if } x_{k-1} \leq y \leq x_k \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where $f_{Y|K}(y|k)$ is the density of income conditional of being in interval k and $\Pr(k)$ is the proportion of observations that falls into interval k. Walter and Weimer's (2018) algorithm builds on a Markov chain result that maintains $\Pr(k)$ equal to its value in the original sample. They propose to build a grid of equally spaced points on the overall distribution. For each interval, the algorithm consists in first allocating all observations to the mid point of the interval and to evaluate the conditional density at each point of the grid that falls within this interval. Then, for each interval, using the value of the conditional kernel density estimates, $\widehat{f}_{Y|K}(y|k)$ as sampling weights, the algorithm draws with replacement a pseudo-sample that has the same size as the the original number of observations within this interval. Since this process has a Markov chain property, repeating this exercise makes the conditional density converge to stationary values. Given that at each iteration the algorithm forces the number of pseudo-observations in each interval to be equal to the number of observations originally in that interval, the $\widehat{\Pr}(k) = N^{-1}\sum_{i=1}^{n} \mathbb{1}(x_{k-1} \leq y_i \leq x_k)$ of the original sample is maintained. Maintaining the original probabilities allows the overall density estimation $\widehat{f}_Y(y)$ to be such that $\int_{x_{k-1}}^{x_k} \widehat{f}_Y(y)\mathrm{d}y = \widehat{\Pr}(k)$. At each step of the algorithm, an estimated value of the $CDF(y)$ is computed using numerical integration of $\widehat{f}_Y(y) = \sum_{k=1}^{K} \widehat{\Pr}(k)\widehat{f}_{Y|K}(y|k)$. By discarding the first B burned-in iterations and averaging the M following iterations, the algorithm estimates the expected value of the $CDF(y)$.[10] The continuous income variable in the 2021-22 wave of the survey allows us to test for goodness of fit of this estimation approach in the first part of the empirical application.

[10]Since the estimation is based on a Markov Chain property of the process, it is assumed that after the first B iterations, the process has reached a stationary distribution. Under this assumption, the average of the next M iterations produces an estimate of the CDFs.

Establishing strict dominance requires strong evidence against the null. In addition, this strong evidence is impossible to obtain over the entire $[0, z^+]$ when the variable of interest is continuous (Davidson and Duclos, 2013). Consequently, we follow the usual practice in the empirical literature and perform the above test in both directions, i.e. $H_0 : F_{Y_0}(y) - F_{Y_1}(y) \geq 0 \; \forall y \in [0, z^+]$ and $H_0' : F_{Y_1}(y) - F_{Y_0}(y) \geq 0 \; \forall y \in [0, z^+]$. We then interpret the results according to the decision rules in Table 2.

To perform the aforementioned test, Barrett and Donald (2003) suggested to use a directional version of the Kolmogorov-Smirnov statistic for the above tests: $\tau = \sup_p (F_{Y_1}(y) - F_{Y_0}(y))$. A natural non parametric estimator of τ can be written as follows:

$$\widehat{\tau} = \sqrt{\frac{n_0 n_1}{n_0 + n_1}} \sup_p \left(\widehat{F}_{Y_1}(y) - \widehat{F}_{Y_0}(y) \right) \tag{5}$$

To perform this test, we follow Linton, *et al.* (2005) and perform a standard full-sample bootstrap applied to the re-centred version of the test statistic.

2.2 Estimating the cumulative distribution's bounds

Given the presence of a substantial number of non-responses in the survey, we need to check for the robustness of our empirical conclusions on the dynamics of poverty if we relax the assumption of missing at random. For this, we need to identify bounds on the sets of admissible CDFs to include all possible CDFs that would be produced by making any possible assumption on the missing values generating process. The bounds of the set of admissible CDFs for 2021-22 is relatively easy to identify. Following the spirit of Horowitz and Manski (1995), a lower bound for this set is identified by allocating all missing values to the highest income level in the data set. Similarly, the upper bound is identified by allocating all missing values to the lowest income level.

In this paper, we expand the methodology of Walter and Weimer (2018) to build bounds on the set of admissible of admissible CDFs to account for the potential non-random missingness in the data when the income variable takes the form of intervals. Although the bounds on the density function are not well defined, the bounds on the CDFs associated with the intervals are well defined. Thus, we allocate non-responses to the lowest (highest) interval for the upper (lower) bounds, and estimate the densities associated with these two different distributions of observations within interval incomes and then integrate these densities to recover the lower and upper bounds of the set of admissible CDFs, F_Y^L and F_Y^U.

In our estimation, we account for non-response (or partial non-response) and assign them to produce bounds on the CDF. This produces two different samples, L and U. For each $S \in \{L, U\}$, we apply the methodology described in the previous section to estimate the bounds F_Y^L and F_Y^U (the details of the extended algorithm can be found in the appendix).

When testing for dominance on sets of admissible CDFs, we follow Fakih *et al.* (2022) and compare the upper bound of the set of CDFs of one distribution with the lower bound of the set of CDFs of the other distribution. This approach ensures that the result would hold for any combination of alternative CDF belonging to the two sets. In the context of continuous income distributions, the dominance test consists of testing the following:

$$
\begin{aligned}
H_0 &: F_{Y_0}^L(y) - F_{Y_1}^U(y) \geq 0 \quad \forall y \in [0, z^+] \\
H_1 &: F_{Y_0}^L(y) - F_{Y_1}^U(y) < 0, \text{ for some } y \in [0, z^+]
\end{aligned}
$$

At this point, it is important to note that while we follow Fakih *et al.* (2022) when we compare the bounds, our approach remains different from their analysis because we test for dominance (instead of non-dominance) to find strong evidence in favor of dominance (instead of strict dominance). To perform this test, we follow a similar procedure and decision rule as the one described in Section 2.1

3 Political Context and Data Availability

3.1 Political context

By the end of 2015, Banque du Liban (Lebanon's central bank) already had a 4.8 billion USD deficit in its net reserves. In April 2016, this worrying issue was flagged by the International Monetary Fund (IMF) in a memo to the Lebanese authorities. However, this memo was never publicly disclosed.[11] Soon after the IMF's memo, Banque du Liban (BDL) and Lebanese authorities accelerated their "financial engineering" operations.[12] In retrospect, these operations were nothing more than a regulated nationwide Ponzi scheme that was mislabelled as a "financial engineering".

When the Lebanese authorities received the IMF's warning, the Lebanese parliament was in its unconstitutional seventh year of mandate. The last elections, at that time, were held in 2009. In 2013, at the end of the official mandate, the parliament implemented an unconstitutional extension of its four-year mandate. Given that the parliament was due for reelection in 2018 (after lawmakers had already extended their mandate twice), the Lebanese authorities successfully convinced the IMF to edit the information out of their January 2017 official country report. In the same year, the lawmakers voted for new public salary scales with massive nominal increases. For instance, the magnitude of the salary at the entry-level (category one) increased from $1,800 to $3,000 per month (the Lebanese pound had a fixed peg with the US dollar at 1,507.50 LBP/USD).[13] Anecdotal evidence suggests that the objective of this increase in the salary scale was to buy the public opinion's approval and prepare the ground for their 2018 reelection. According to CAS and the International Labour Organization (ILO), the public sector employs 14% of Lebanese workers (see CAS

[11] The interested reader can consult this article of October 28, 2021 from Reuters: https://www.reuters.com/world/middle-east/before-lebanons-current-financial-crisis-central-bank-faced-47-billion-hole-2021-10-28.

[12] Financial engineering is the term used by the BDL and the Lebanese authorities.

[13] The interested reader can consult this article of July 19, 2017: http://www.businessnews.com.lb/cms/Story/StoryDetails/6162/Salary-scale-ratified-by-Parliament

and ILO, 2019), and this non-negligible proportion of public servants is an important part of the political parties' electoral base via an elaborate system of political patronage. Even if labor markets in the Arab region are typically characterized by *wasta* systems,[14] the nature of the clientelist system in place in Lebanon is more extreme than in the rest of the Arab region. This severe clientelism is the consequence of the tight militia control over government since the 1980s. This militia control did not improve, even after the Taif Agreement of 1989 ended the 1975-1990 Lebanese Civil War. Indeed, instead of addressing these distortions, the Taif Agreements enshrined a *redistributive kleptocracy*.[15] In this system, militia leaders would share the maximum rent they could extract from the Lebanese state and redistribute part of the proceeds to their political base. The rent extraction operations started during the civil war and peaked between 2016 and 2018, when banks granted the highest interest rates of the 2010s decade to finance a national Ponzi scheme. This coordination between banks and the ruling class was made possible because of the strong connections between the banking sector and the political class (see Chaaban, 2019). Even after the banks ceased buying government debt directly in 2015, the central bank continued to use bank funds at BDL to purchase government debt. This Ponzi scheme positively impacted the income of households who were interest income earners, which in turn boosted their satisfaction and allowed for a broader electoral base.

With these substantial and distracting nominal raises of salary scales and the fictitious rates of return on deposits, lawmakers started increasing indirect taxation. At the same time, they maintained citizens in acute and increasing deprivation in non-income dimensions (waste management, water, electricity, sanitation, heath and education). In October 2019, a proposed excise tax (on gasoline, tobacco, and VoIP calls) affecting the disadvan-

[14] *Wasta* is an Arabic word that loosely translates into nepotism or patronage.

[15] The term *redistributive kleptocracy* was coined by Ghassan Salame, Minister of Culture of Lebanon from 2000 to 2003. See: https://www.lorientlejour.com/article/1296329/ghassan-salame-le-liban-est-arrive-a-un-point-ou-un-regime-radicalement-different-doit-etre-envisage.html

taged population, combined with the publicly displayed opulent political leaders' lifestyle, triggered Lebanon's most significant national protests. These protests led to the resignation of the second Saad Hariri cabinet and a political deadlock that impeded the implementation of the necessary reforms. All these events resulted in the unavoidable collapse of the Ponzi scheme with the default on US dollar denominated public debt in March 2020, creating a substantial economic contraction. As a result, Lebanon has been suffering since 2019 from currency, banking, and economic crises, which created conditions for economic depression and galloping inflation. In addition, the ongoing COVID-19 pandemic and the August 2020 Beirut Port explosion exacerbated this economic contraction. Thus, there is a clear and urgent need for empirical evidence to document the rising poverty and guide the country out of its economic crisis yet, in face of such a crisis, the CAS is still reluctant to share any data they have with academic researchers.

3.2 Data availability and challenges

As mentioned earlier, the objective of this paper is to monitor the change in poverty during the aforementioned period. Although Lebanon's CAS has some surveys, these surveys are not readily accessible to academic researchers.[16] To overcome this hurdle, we use an alternative source of information, namely the Waves IV, V, and VII of the Arab Barometer surveys.[17] The interviews for Lebanon were conducted between July and August 2016 for Wave IV, between September and October 2018 for Wave V, and between December 2021, and February 2022, for Wave VII. It is important to note that the Arab Barometer surveys are not designed for income distribution analysis. The purpose of the Arab Barometer is to establish a baseline socioeconomic profile to gauge public opinion in the Arab region. Nevertheless, some survey waves contain valuable information on income or income inter-

[16] The CAS are only open to share aggregate data and do not allow access for their micro data. Such aggregate information is useless for poverty analysis.

[17] The data is available for researchers: https://www.arabbarometer.org/

vals. Despite the limited information on income and the relatively small sample size of these surveys, this paper shows that it is possible to make robust claims on changes in poverty in Lebanon between 2016 and 2022. The approach used and the results provided in this paper illustrate a promising potential of using and collecting high-frequency data using mobile phone surveys, as pointed out in Hoogeven *et al.* (2014), and the importance of such surveys in the Arab region suggested in Atamanov *et al.* (2020).

3.3 Data

We use the Arab Barometer (Lebanon) survey waves IV, V and VII. Usually these surveys are representative of the country's population. More specifically, for the case of Lebanon, the Arab Brometer survey team relies on the CAS' Public Housing and Population Census of 2011 and stratifies the survey by governorate and sect to construct the sampling frame. The survey team uses a stratified area probability sample and conducts face-to-face interviews.

The Arab Barometer Surveys contain two types of questions that help address our research question. The first type consists of a set of questions focusing directly on income intervals (see Table 3 for 2016 and Table 4 for 2018). The second type of questions consists of the usual continuous information on income (see Table 5 for 2021-22). An important observation is worth mentioning: the proportion of non-responses seems positively correlated with the precision of the income question. This observation suggests an arbitrage between having a precise question on income and using income intervals in these surveys.

It is important to mention that after 2019, Lebanon experienced extreme inflation rates (131.05% in 2020 and 144.12% in 2021) and high exchange rate volatility.[18] The high monetary instability adds an additional challenge in identifying the real value of income for the 2021-22 wave. In order to make income comparable over all the period, we use the August 2016 consumer price index (CPI) for the 2016 wave and the October 2018 CPI for

[18] See https://blog.blominvestbank.com/42028/lebanons-inflation-rate-reached-144-12-by-september-2021

the 2018 wave displayed in Table 6. Since income variability is more pronounced in 2021-22, we use the CPI corresponding to the exact month of the interview for each observation in the 2021-22 wave. For comparability purposes, we adjust all incomes and express them in LBP of October 2018 using the local CPI.

4 Results

4.1 Goodness of fit of the data augmentation approach

In order to empirically verify that the methodology we are proposing offers a reasonable approximation of the continuous income distribution, we use the continuous income observations to generate an artificial interval variable for 2021-22 using the same interval values used in 2018 (see Table 4). This artificial interval income variable is created by allocating each deflated continuous income observation to its corresponding interval.

In order to illustrate how the simulated CDF would compare to the estimated CDF in case of no missing observations, we assume missing at random and discard the 363 observations with a missing response and estimate the simulated CDF with interval data and the real CDF. Figure 1 depicts the two CDFs with their 95% confidence bands. The simulated CDF with income intervals in Figure 1 is very close to the CDF estimated from continuous data. In order to test formally for the goodness of fit of the data augmentation model, we compute a Kolmogorov-Smirnoff goodness of fit test. The p-value of this test is 0.3859. Therefore, we cannot reject that the simulated CDF with income interval data equals the estimated empirical distribution function. This result is a solid empirical argument favoring this estimation approach.

4.2 A dominance approach to the dynamics of poverty in Lebanon from 2016 to 2022

This section uses the information available on income intervals in 2016 and 2018 and income in 2021-22 to test for stochastic dominance and build a narrative on poverty and social welfare dynamics during this recent period. Tables 3, 4, and 5 display these survey questions and the distribution of the associated responses in the surveys. In this section, we assume missing at random, which is the usual assumption made in applied work on inequality and poverty. This assumption allows us to discard non-responses from the sample and to estimate the CDFs based on the remaining observations.

For 2016, we first convert the income categories' bounds in LBP of October 2018 using the fixed exchange rate of 1,507.50 LBP/USD and the variation in the CPI between August 2016 and October 2018.[19] For 2021-22, for each observation, we use the CPI of the month of the interview relative to the October 2018 CPI to transform the observed household income into its equivalent in LBP of October 2018.

We start by comparing the 2018 income's CDFs over 2016 income's CDFs and analyze the change in the income distribution between 2016 and 2018. We apply the estimation strategy of Section 2.1 to our dataset assuming missing at random. Figure 2 displays the CDFs of 2016 and 2018 with their 95% confidence bands. A visual inspection suggests that the CDF of income for 2018 lies below the CDF of income for 2016. To confirm that the results are statistically meaningful, we perform the statistical test described in Section 2.1 using a bootstrap procedure with 999 replications. This stochastic dominance test confirms the result since the associated p-values in Table 7 indicate that we cannot reject H_0 and that we can reject H_0'. Using the decision rules in Table 2 we can say that the CDF of income of 2018 first-order stochastically dominates the CDF of income of 2016. This result means that for any poverty index and any poverty line, poverty decreased between 2016 and 2018. In addition, since there is no maximum value chosen for potential poverty lines, this also implies that for any social welfare index, welfare increases between 2016 and 2018.

[19] The LBP/USD exchange rate was fixed and stable in these years. It was only starting in October 2019 that this exchange rate became unstable. The period between 1997 and 2019 is characterized by a "hard" peg, which might have been a facilitating factor for setting a nationwide Ponzi scheme during the later years of that peg regime.

Therefore this result allows for a conclusion that is compatible with the hypothesis that the acceleration of the nationwide Ponzi scheme after 2016 succeeded in temporarily reducing poverty and increasing welfare in Lebanon.

We now turn to the dynamics between 2018 and 2021-22. Figure 3 displays the CDFs for 2018 and 2021-22. A visual inspection of this figure indicates an increase in the proportion of people falling below each income level. Table 7 displays the p-values of the associated stochastic dominance test. These p-values indicate that we cannot reject H_0 and that we can reject H_0'. Using the decision rules in Table 2, we can say that the CDF of income for 2018 first-order stochastically dominates the CDF of income for 2021-22. This result implies that any poverty and social welfare index would indicate an increase in poverty and a reduction in social welfare between 2018 and 2021-22. This result is compatible with the general perception and modeled change in the poverty rate for Lebanon (see Abu-Ismail and Hlásny, 2020, and ESCWA, 2020 and 2021).

Finally, it is interesting to check if the reduction of poverty and increase in social welfare generated by the Ponzi scheme persisted after 2018 or if there was a side effect to the Ponzi scheme. To answer this question, we must compare the distributions of 2016 with 2021-22. The graphical representation of this comparison presented in Figure 4, which displays the CDFs for 2016 and 2021-22. A visual inspection of this figure indicates that the CDF of income for 2016 first-order stochastically dominates the CDF of income for 2021-22. Table 7 displays the p-values of the associated stochastic dominance test. These p-values indicate that we cannot reject H_0 and that we can reject H_0'. Using the decision rules in Table 2, we can say that the income CDF for 2016 first-order stochastically dominates the income CDF for 2021-22. This last result implies that any poverty and social welfare index would indicate an increase in poverty and a reduction of social welfare between 2016 and 2021-22. Thus there is a negative net effect of the pre-elections Ponzi scheme that is combined with the COVID-19 pandemic and the impact of the August 2020 Beirut port blast. Thus, not only welfare is lower in 2021-22 compared to 2018, it is also lower than what it was in 2016. In addition to the increase in poverty, it is important to highlight that in the context of Lebanon, and the prevalent political connectedness mentality or "wasta", this kind of financial mismanagement can add an excess burden by reallocating income inefficiently between those who have a political connection (usually at the top of the income distribution) and those who do not benefit from such a network (usually at the bottom of the income distribution) thus increasing inequalities.[20]

4.3 Robustness check: relaxing the missing at random assumption

Of 2,399 observations for the 2021-22 wave in Table 5, we have 363 missing values for the income question (i.e., "don't know" or "refused"). These missing values represent 15.1% of the observations, which is substantially higher than what we had for the 2016 and 2018 waves with respectively 5.9% and 3.0% of missing values (see Tables 3 and 4). One interesting aspect of this distribution of non-responses is that individuals are less (more) willing to answer more (less) precise questions on income. If we refer to Tables 3 and 4, it also seems that the number of non-responses increases when we move from a question with two income categories to questions with a finer grid of categories. In both waves, the relative increase in non-responses is slightly higher for the individuals in the highest categories for the first question for both waves. This variation in non-response rates may indicate that missing values are not random.

Since the non-response in 2021-22 is substantially larger than in 2018 and 2016, one could wonder if the missing-at-random assumption is not artificially inducing the poverty increase described in the preceding section. One should address this issue if it is reasonable to believe that there is a positive correlation between non-responses and income levels. In

[20] Lebanon has an economically exclusive political system is that sense. See Acemoglu and Robinson (2012).

such a context, it is essential to check if the empirical results on the dynamics of poverty presented in the preceding section are robust to a change in the assumption of the missing values-generating process. In order to perform this robustness check, we use the estimation approach described in Section 2.2 to estimate bounds on the sets of admissible CDFs, allowing for any potential assumption on the missing values generating process.

In proportion, we have five times more missing values in the data of 2021-22 than in 2018. By construction, this will automatically imply that any dominance test aiming at establishing that 2018 is everywhere below 2021-22 is doomed to failure. This is due to the fact that 15.1% of missing values in 2021-22 are allocated at the upper limit of the observations to produce the lower bound compared to 3.0% for 2018. However, despite this issue, it may be possible to test for poverty dominance of a set of poverty lines $z \in [0, z^+]$. To perform these tests we need to set an upper limit to the admissible poverty lines. We chose $z^+ = 1,500,000$ LBP of 2018/month, which represents more than 10 times the World Bank per capita upper-middle-income country poverty line for Lebanon in 2018.[21]

To account for non-response, we exploit this information structure of the 3 questions on income intervals in the 2016 and 2018 survey and produce the bounds of the CDFs of income. Typically, for a standard question on income categories, the worst-case lower (upper) bound is produced by allocating all non-responses to the highest (lowest) category. In this paper, the context is more complex because we have more refined information than a typical income category question. This is why we exploit the informational structure of the survey and allocate the non-responses for the lower (upper) bound in the following way for 2018:

- the 40 non-responses to the first question are allocated to the "4,000,001 LBP or more"

[21] The information on household size being only available in wave VII, we use household income to compare between years. This value would represent more than twice an adjusted poverty line at the household level if we use the average household size of 4.8 in the wave VII survey.

("450,000 LBP or less") category,

- the 4 non-responses to the second question are allocated to the "751,000-1,000,000 LBP" ("450,000 LBP or less") category, and
- the 16 non-responses to the third question are allocated to the "4,000,001 LBP or more" ("1,000,000-1,500,000 LBP") category

We follow a similar procedure for 2016. To compute the bounds for the 2021-22 observations, we allocate the missing values of 200,000 (120,000,000) for the continuous data to produce the upper (lower) bound.

We apply the estimation strategy of Section 2.2 to estimate the bounds of the sets of admissible CDFs of income for 2016 and 2018 and test for stochastic dominance. The darker areas of Figure 5 represent these sets of admissible CDFs of income. A visual inspection suggests that the entire set of admissible CDFs of income for 2018 lies below the admissible CDFs of income set for 2016. To confirm that the results are statistically meaningful, we perform the statistical test described in Section 2.2 with 999 bootstrap replications. This stochastic dominance test confirms the result, the associated p-values in Table 8 indicate that we cannot reject H_0 and that we can reject H_0'. Using the decision rules in Table 2 we can say that the real CDF of income of 2018 first-order stochastically dominates the real CDF of income of 2016. This result means that for any poverty index and any poverty line, poverty decreased between 2016 and 2018. In addition, since there is no maximum value chosen for potential poverty lines, this also implies that for any social welfare index, welfare increases between 2016 and 2018. This robustness test confirms the narrative of the preceding section on the dynamics of poverty and social welfare for the 2016 to 2018 period.

Figure 6 displays the sets of admissible CDFs for 2018 and 2021-22. A first interesting observation is that the bounds on the set of admissible CDF for 2021-22 are substantially larger than for 2018. This is due to the huge increase of non-response in presence of a precise question on income. A visual inspection of this figure also indicates an increase in the proportion of people falling below each income level for a wide range of income levels. Table 8 displays the p-values of the associated stochastic dominance test on the restricted domain $z \in [0, 1500000]$. These p-values indicate that we cannot reject H_0 and that we can reject H_0'. Using the decision rules in Table 2, we can say that the real CDF of income for 2018 first-order stochastically dominates the real CDF of income for 2021-22 at least up to 1,500,000 LBP of 2018/month.[22] This result implies that for any poverty line below 1,500,000 LBP of 2018/month any poverty index would indicate an increase in poverty between 2018 and 2021-22. This robustness test confirms the narrative of the preceding section on the dynamics of poverty for the 2018 to 2021-22 period.

Figure 7 displays the sets of admissible CDFs for 2016 and 2021-22. A visual inspection of this figure also suggest that dominance of 2016 over 2021-22 in terms of poverty. Table 8 displays the p-values of the associated stochastic dominance test on the restricted domain $z \in [0, 1500000]$. These p-values indicate that we cannot reject H_0 and that we can reject H_0'. Using the decision rules in Table 2, we can say that the real CDF of income for 2016 first-order stochastically dominates the real CDF of income for 2021-22 at least up to 1,500,000 LBP of 2018/month. This result implies that for any poverty line below 1,500,000 LBP of 2018/month any poverty index would indicate an increase in poverty between 2016 and 2021-22. This robustness test confirms again the narrative of the preceding section on the dynamics of poverty for the 2016 to 2021-22 period.

4.4 Poverty headcount estimation using interval income data

Although our dominance tests already offer a complete picture of the evolution of poverty during the 2016-2022 period in Lebanon, it is still interesting to estimate headcount indices

[22]Visual inspection of the figure suggests that the result would also hold for higher values of the poverty line, but we wanted to choose a similar upper limit for all comparisons.

for diffusion purposes. The motivation to produce these poverty headcount figures is that they are usually easier to explain to policymakers and the general public.

The first step in estimating these headcount indices consists in selecting poverty lines. For this selection, we rely on the World Bank poverty lines. After ranking Lebanon as an upper-middle-income country for 25 years, the World Bank changed its ranking of Lebanon in 2022 to a lower-middle-income country.[23] For this reason, we use two World Bank poverty lines for Lebanon: the upper-middle-income country poverty line, and the lower-middle-income country poverty line. These poverty lines correspond respectively to per-capita monthly incomes of 79,619.59 and 149,422.93 in 2018 LBP.[24] Since only wave VII of the survey has the complete information on household size, we transform these poverty lines into total household income poverty lines using the average household size of 4.8 in the 2021-22 survey.

Figure 8 presents the bounds for the headcount index for 2016, 2018, and 2022 for the upper-middle-income country poverty line. The thick line inside the box for each year represents the estimated headcount index assuming missing at random. These values are respectively 18.2%, 10.0%, and 71.0% for 2016, 2018, and 2022. The box for each year represents the set of all admissible values for the headcount if one relaxes the missing-at-random assumption, and the error bar is the 95% confidence interval on these bounds. For example, in 2016, the bounds on the values of the headcount index varied between 17.2% and 21.5% (15.4% and 23.3% if we account for the 95% confidence interval). In 2018, the bounds on the values of the headcount index varied between 9.8% and 11.4% (8.8% and 12.4% if we account for the 95% confidence interval). Within a window of less than three and a half years, the bounds on the value of the headcount index in 2021-22 lies between

[23] See https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2022-2023

[24] We use the per-capita household income poverty lines produced for 2020 by Jolliffe *et al.* (2022) and adjust to 2018 values using the implicit deflator to the CPI produced by the Central Administration of Statistics of Lebanon.

59.0% and 75.9% (57.0% and 77.6% if we account for the 95% confidence interval). These results are consistent with the modeling approach offered by Abu-Ismail and Hlásny (2020) and ESCWA (2021), as the values offered in these two papers fall within the set of the headcount index values for 2021-22. These results also indicate that if an analyst uses the upper-middle-income country poverty line, the incidence of poverty in 2021-22 is between 5.2 to 7.7 times higher compared to 2018 for any assumption this analyst makes on the missing-value generating process.

Figure 9 presents the bounds for the headcount index for 2016, 2018, and 2022 for the lower-middle-income country poverty line. The estimated values of the headcount index are respectively 5.1%, 2.6%, and 38.1% for 2016, 2018, and 2022. If we relax the missing-at-random assumption for 2016, the bounds on the values of the headcount index varied between 4.8% and 8.5% (3.9% and 9.7% if we account for the 95% confidence interval). In 2018, the bounds on the values of the headcount index varied between 2.5% and 3.7% (2.1% and 4.2% if we account for the 95% confidence interval). The bounds on the value of the headcount index in 2021-22 lie between 31.6% and 48.5% (29.8.0% and 50.5% if we account for the 95% confidence interval). These results indicate that if an analyst uses the lower-middle-income country poverty line, the incidence of poverty in 2021-22 is between 8.5 to 19.4 times higher compared to 2018 for any assumption this analyst makes on the missing-value generating process. The proportional increase is much higher when one uses the lower-middle-income country poverty line indicating that the burden of what the World Bank (2020) coined as a *deliberate depression* more than proportionally impacts those at the extreme bottom of the income distribution.

5 Conclusion

This paper is motivated by the perverse issue of data poverty in the MENA region in general and the recent economic history of Lebanon in particular. While the proposed method finds its inspiration in the lack of necessary data for poverty analysis, our empirical application's choice is driven by the lack of evidence-based policy in a country going through an unprecedented crisis, Lebanon. The case of Lebanon is despairing but interestingly provides empirical evidence that illustrates the detrimental effect of the lack of data infrastructure necessary for poverty monitoring. To overcome this issue, we extend Walter and Weimer's (2018) and Walter's (2019) estimation methods and derive bounds on admissible cumulative income distributions sets while accounting for survey non-response and interval data on income. This approach allows us to analyze poverty dynamics using stochastic dominance even if access to data is limited. This same method can be applied to estimate bounds for poverty and inequality indices if one wishes to produce more complete rankings of the distributions. In the paper, we also offer an application with the canonical headcount index. Thus, the framework proposed in this paper is applicable in many contexts, can be adapted for studying other types of data deprivations, and may be a valuable tool for policymakers and international organizations (see Serajuddin *et al.*, 2015).

We illustrate the proposed approach using recent Lebanese data and show that small surveys with limited information can be used to produce meaningful information in terms of poverty dynamics. The results from our empirical illustration allow for a conclusion compatible with an artificial decrease in poverty in Lebanon between 2016 and 2018, financed through a Ponzi scheme. This artificial decrease happened just before the general elections and collapsed shortly after due to the uprising of the Lebanese population, who was struggling to make ends meet by 2019. The observed pattern suggests that the expensive and unsustainable vote-buying strategy that led to the most significant financial collapse in the country's history has contributed to the spike in poverty rates to higher levels than those prevailing in 2016. Moreover, our results support Hoogeveen *et al.* (2014)'s idea regarding the importance of using short and quick mobile-phone surveys as an alternative source of information between surveys or when surveys are unavailable.

Finally, our empirical application exploits the additional information from the funnel-design questionnaire. More specifically, it uses the additional information provided by larger income interval and are missing in narrower intervals. It thus highlights the value of having surveys on income with a similar design. In the future, it would be valuable to use similar phone-based surveys to collect income information and monitor the MENA region's poverty dynamics.

References

[1] Abu-Ismail, K., and V. Hlásny (2020). Wealth distribution and poverty impact of COVID-19 in Lebanon. ESCWA Technical Paper 8, E/ESCWA/CL3.SEP/2020/TP.8.

[2] Acemoglu, D. and J.A. Robinson (2012), *Why nations fail: The origins of power, prosperity and poverty*, Profile, London.

[3] Atkinson, A.B. (1987), On the measurement of poverty, *Econometrica*, 55, 759-764.

[4] Allison, R.A. and J.E. Foster (2004), Measuring health inequality using qualitative data, *Journal of Health Economics*, 23, 505-524.

[5] Arezki, R., D. Lederman, A. Abou Harb, N. El-Mallakh, R.Y. Fan, A. Islam, H. Nguyen, and M. Zouaidi (2020), Middle East and North Africa economic update, April 2020 : How Transparency Can Help the Middle East and North Africa, World Bank.

[6] Atamanov, A., S.A. Tandon, G.C. Lopez-Acevedo, and M.A. Vergara Bahena (2020), Measuring monetary poverty in the Middle East and North Africa (MENA) region :

Data gaps and different options to address them, Policy Research Working Paper Series 9259, The World Bank.

[7] Barrett, G.F. and S.G. Donald (2003), Consistent tests for stochastic dominance, *Econometrica*, 71, 71-104.

[8] CAS and ILO (2019), *Labour Force and Household Living Conditions Survey (LFHLCS), 2018–2019, Lebanon*, online report. https://www.ilo.org/beirut/publications/WCMS_732567.

[9] Chaaban, J. (2019), I've got the power. Mapping connections between Lebanon's banking sector and the ruling class, in I. Diwan, A. Malik, and I. Atiyas (eds.), *Crony Capitalism in the Middle East*, Oxford University Press, 330-343.

[10] Duclos, J.-Y. and P. Makdissi (2004), Restricted et unrestricted dominance for welfare, inequality and poverty orderings, *Journal of Public Economic Theory*, 6, 145-164.

[11] El-Haddad, A. (2020), Redefining the social contract in the wake of the Arab Spring: The experiences of Egypt, Morocco and Tunisia, *World Development*, 127, 104774.

[12] ESCWA (2020), Poverty in Lebanon: Solidarity is vital to address the impact of multiple overlapping shocks, Policy Brief 20-00268.

[13] ESCWA (2021), Multidimensional poverty in Lebanon (2019-2021). Painful reality and uncertain prospects, Policy Brief 21-00634.

[14] Fakih, A., P. Makdissi, W. Marrouch, R. Tabri and M. Yazbeck (2022), A stochastic dominance test under survey nonresponse with an application to comparing trust levels in Lebanese public institutions, *Journal of Econometrics*, 228, 342-358.

[15] Ferreira, F., Chen, Sh., Dabalen, A., Dikhanov, Y., Hamadeh, N., Jolliffe, D., Narayan, A., Prydz, E.B., Revenga, A., Sangraula, P., Serajuddin, U. and Yoshida (2016), A

global count of the extreme poor in 2012: data issues, methodology and initial results, *Journal of Economic Inequality*, 14, 141-172.

[16] Foster, J.E. and A.F. Shorrocks (1988), Poverty orderings and welfare dominance, *Social Choice and Welfare*, 5, 179-198.

[17] Groß,M., U. Rendtel, T. Schmid, S. Schmon, and N. Tzavidis (2017), Estimating the density of ethnic minorities and aged people in Berlin: multivariate kernel density estimation applied to sensitive georeferenced administrative data protected via measurement error, *Journal of the Royal Statistical Society: Series A*, 180, 161-183.

[18] Harb, N. (2022), Earnings function in Lebanon: does religion matter?, *Applied Economics*, 54, 821–840.

[19] Harb, N. and T. Rouhana (2020), Earnings and gender wage gap in Lebanon: the role of the human and social capital, *Applied Economics*, 52, 4834-4849.

[20] Hoogeveen, J., K. Croke, A. Dabalen, G. Demombynes, and M. Giugale (2014), Collecting high frequency panel data in Africa using mobile phone interviews, *Canadian Journal of Development Studies / Revue canadienne d'études du développement*, 35, 186-207.

[21] Horowitz, J. L. and C. F. Manski (1995), Identification and robustness with contaminated and corrupted data, *Econometrica* 63, 281-302.

[22] Jean, N., M. Burke, M. Xie, W.M. Davis, D.B. Lobell, and S. Ermon (2016), Combining satellite imagery and machine learning to predict poverty. *Science*, 353, 790-794.

[23] Jolliffe, D., D. Gerszon Mahler, C. Lakner, A. Atamanov, and S. Kofi Tetteh-Baah (2022), Assessing the Impact of the 2017 PPPs on the International Poverty Line and Global Poverty, Policy Research Working Paper 9941, World Bank.

[24] Linton, O., E. Maasoumi, and Y-J Whang (2005), Consistent testing for stochastic dominance under general sampling schemes, *Review of Economic Studies*, 72, 735-765.

[25] Makdissi, P. and M. Yazbeck (2017), Robust rankings of socioeconomic health inequality using a categorical variable, *Health Economics*, 26, 1132-1145.

[26] Serajuddin, U., H. Uematsu, C. Wieser, N. Yoshida, and A. Dabalen (2015), Data deprivation: another deprivation to end, World Bank Policy Research Working Paper, 7252.

[27] van der Weide, R, B. Blankespoor, C. Elbers, and P. Lanjouw (2022), How Accurate Is a Poverty Map Based on Remote Sensing Data? An Application to Malawi, Policy Research Working Paper 10171, World Bank.

[28] Walter, P. (2019), A selection of statistical methods for interval-censored data with applications to the German microcensus. PhD thesis, Freie Universität Berlin.

[29] Walter, P., and K. Weimer (2018), Estimating poverty and inequality indicators using interval censored income data from the German microcensus, Freie Universität Berlin, School of Business & Economics Discussion Paper 2018/10.

[30] World Bank (2020), Lebanon Economic Monitor, Fall 2020 : The Deliberate Depression. World Bank, Washington, DC.

[31] World Bank (2021), *Lebanon sinking (to the top 3)*, Lebanon Economic Monitor, Spring 2021.

A Algorithm for the estimation of the bounds on the set of admissible CDFs.

1. Use the midpoints of the intervals as pseudo $\hat{y}_i$ for the unknown y_i. Estimate pilots of $\widehat{f}_Y^S$, $S \in \{L, U\}$ using kernel density estimation.

2. Evaluate $\widehat{f}_{Y|K}^S(y|k)$, $S \in \{L, U\}$ on an equal-spacing grid $\{g_1, g_2, \ldots, g_J\}$, where

$$g_j = j * \frac{x_K}{J}$$

Note that in the empirical application we set k_K as twice the value of y in the highest interval category "income higher than y".

3. Draw with replacement from $f_{Y|K}^S(y|k)$ by drawing randomly from $G_\ell = \{g_j | g_j \in (x_{k-1}, x_k]\}$ with sampling weights $\widehat{f}_{Y|K}^S(g_j|k)$. The number of observation to draw for each interval is given by the number of observations within each interval. Obtain two series of $\hat{y}_j$ for $j \in \{1, 2, \ldots, n\}$ one for each $S \in \{L, U\}$.

4. Recompute the densities $\widehat{f}_{Y|K}^S(y|k)$ and then $\widehat{f}_Y^S(y) = \widehat{f}_{Y|K}^S(y|k)/\widehat{\Pr}[k]$, $S \in \{L, U\}$ and numerically integrate these functions to obtain the bounds on the CDF^S

$$\widehat{F}_Y^S(y) = \int_0^y \widehat{f}_Y^S(u)\mathrm{d}u, \quad S \in \{L, U\}.$$

5. Repeat steps 2 to 5 with B burn-in and M additional iterations.

6. Discard the B burn-in iterations and estimate the average of the bounds on the set of admissible CDFs using the M estimates.

Figure 1: Cumulative income distribution in 2021-22 (missing at random)

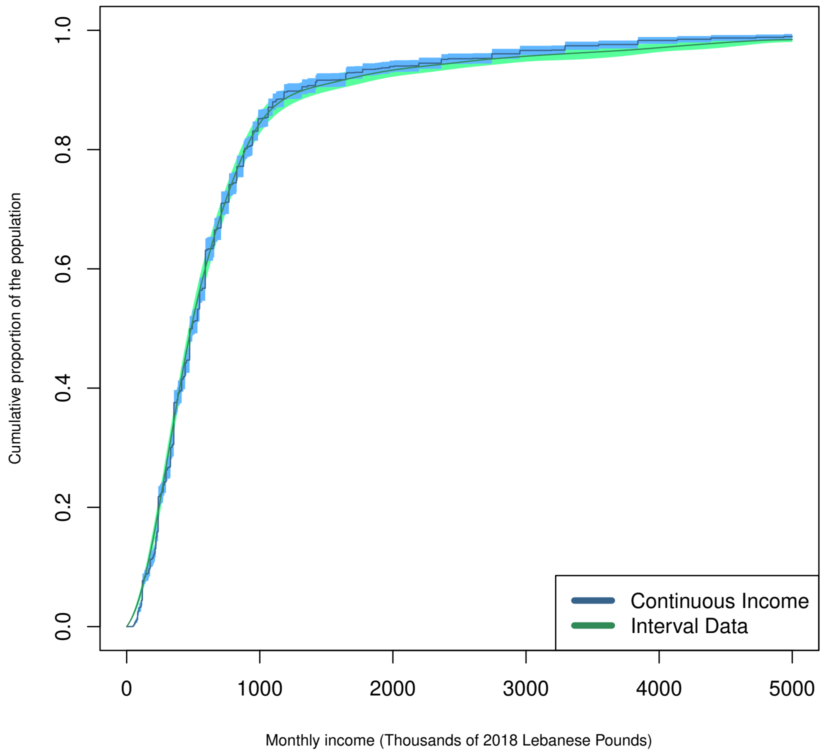

Figure 2: Cumulative income distributions for 2016 and 2018 (missing at random)

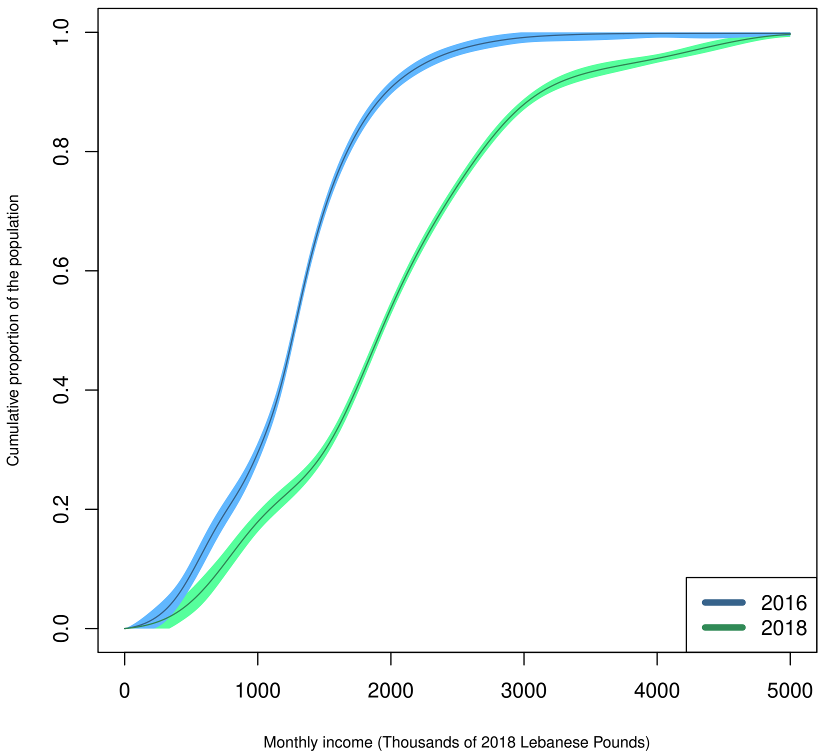

Figure 3: Cumulative income distributions for 2018 and 2022 (missing at random)

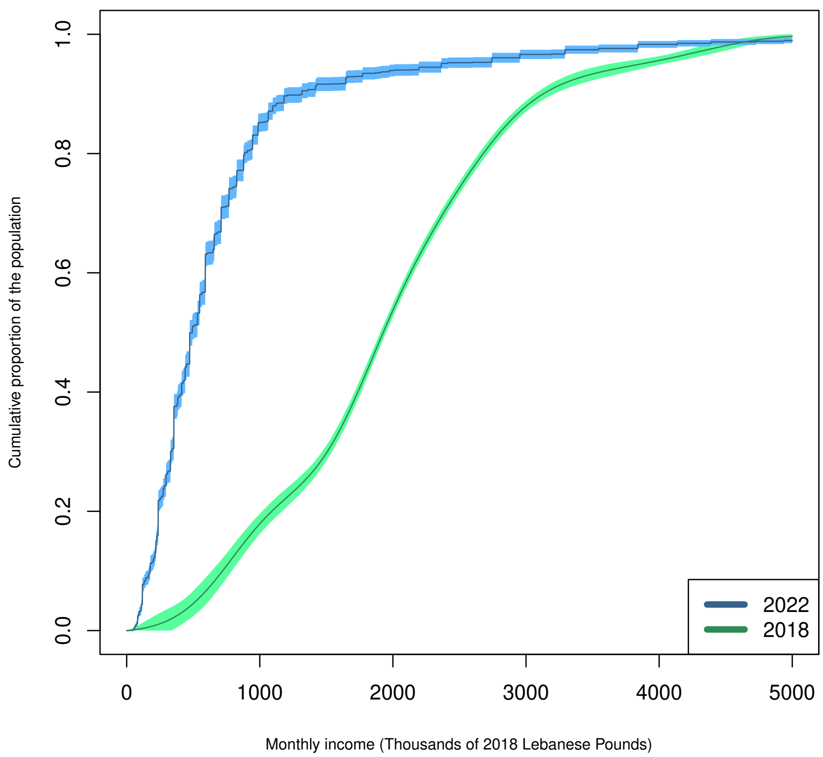

Figure 4: Cumulative income distributions for 2016 and 2022 (missing at random)

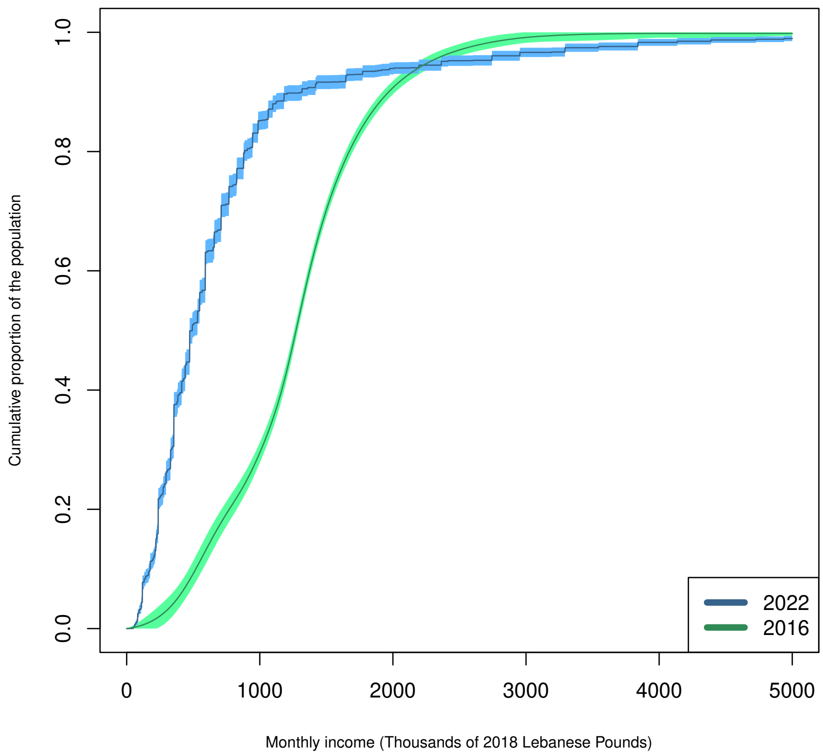

Figure 5: Bounds on the 2016 and 2018 cumulative income distributions

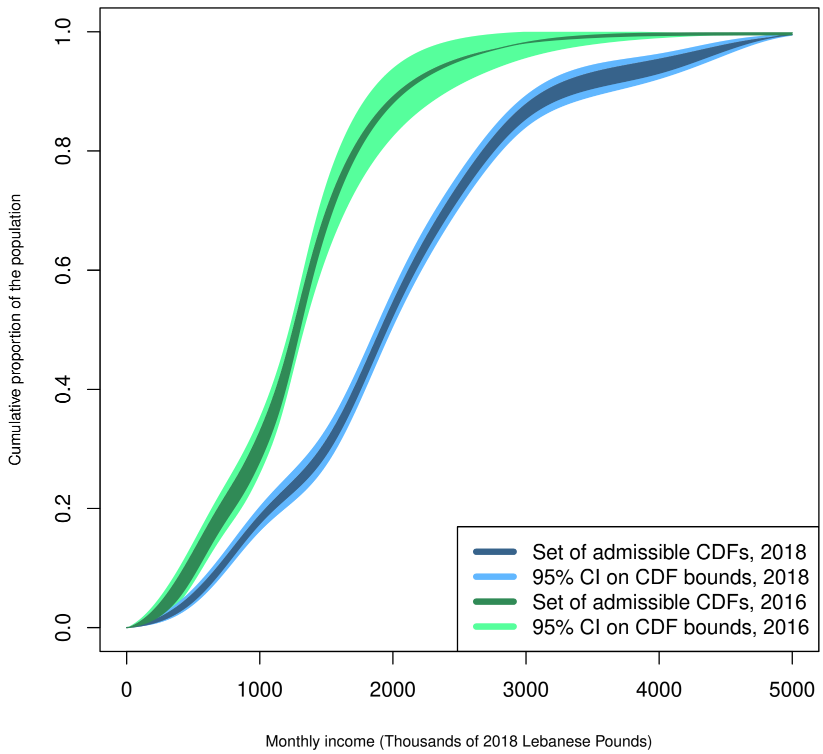

Figure 6: Bounds on the 2018 and 2022 cumulative income sufficiency distributions

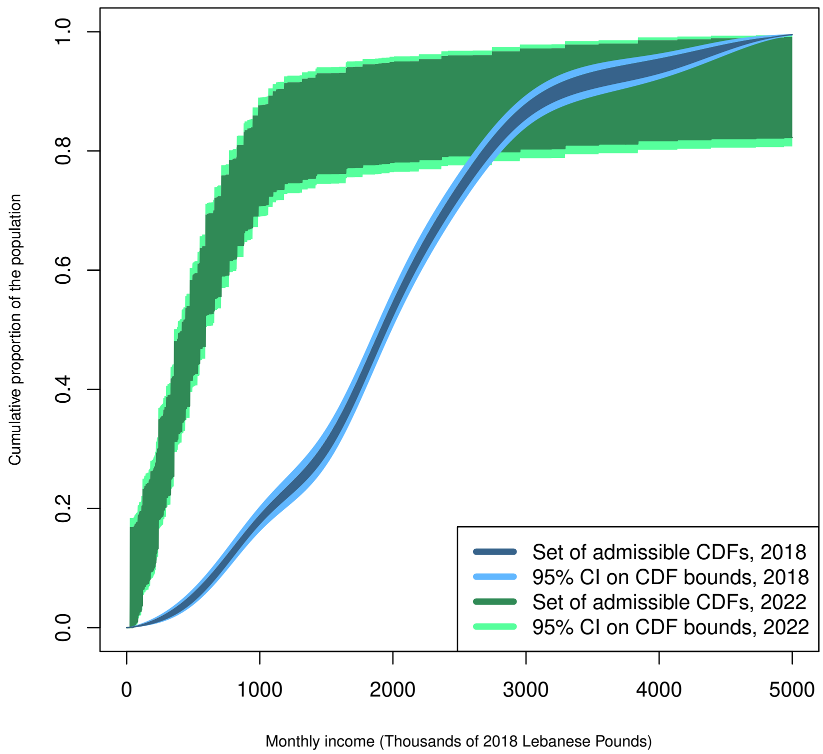

Figure 7: Bounds on the 2016 and 2022 cumulative income sufficiency distributions

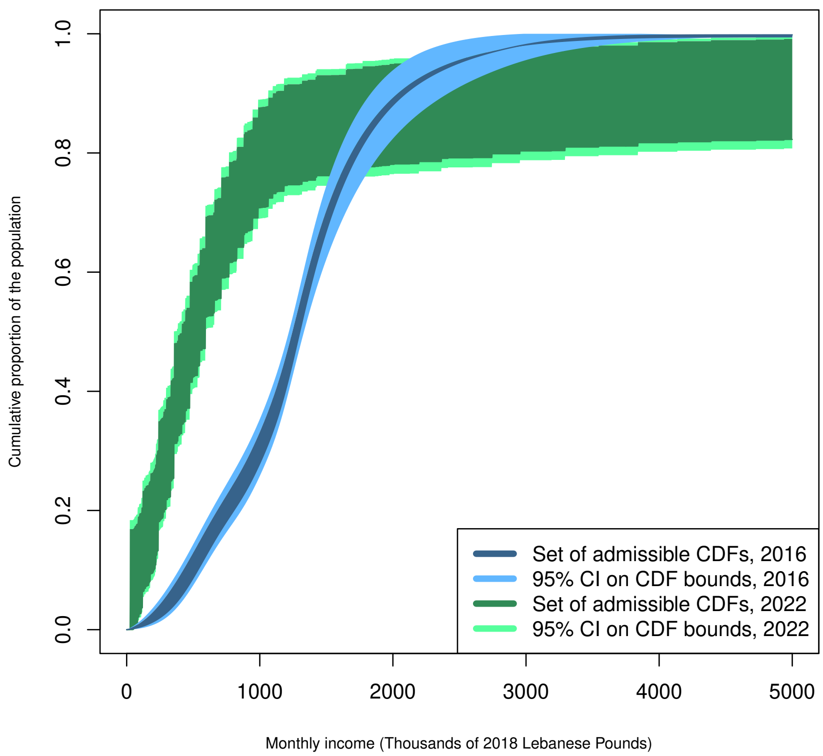

Figure 8: Bounds on headcounts (Upper-Middle-Income Country poverty line)

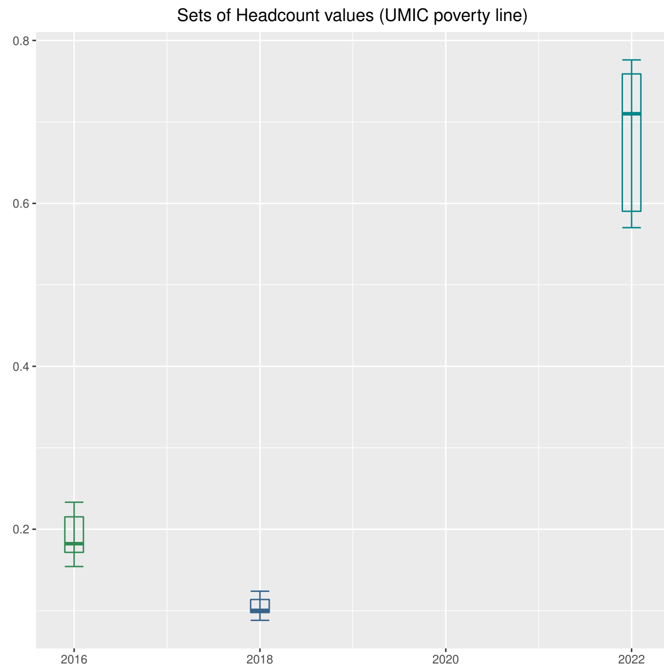

Figure 9: Bounds on headcounts (Lower-Middle-Income Country poverty line)

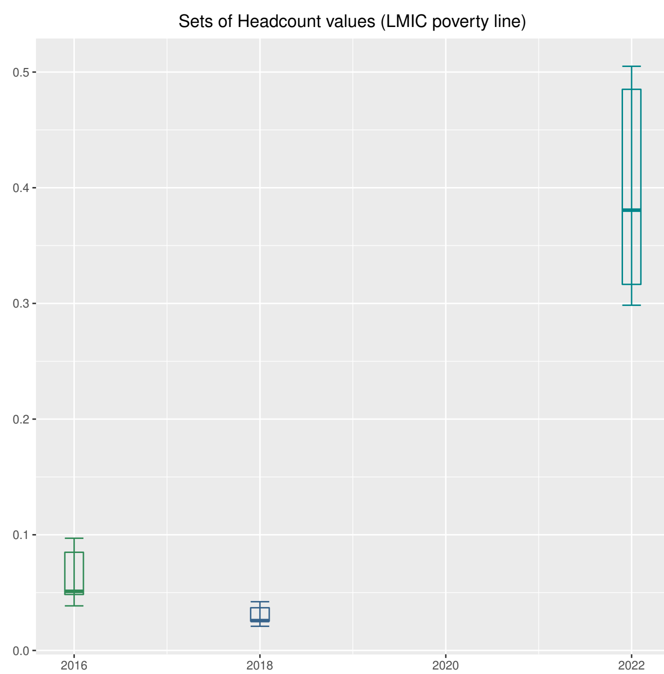

Table 1: Statistical capacity in the MENA as of 2021

Rank	Country	Statistical Capacity Indicator (SCI)
1	Egypt	82.22
2	Jordan	77.78
2	West Bank and Gaza	77.78
4	Iran	75.56
5	Moroccco	66.67
6	Tunisia	58.89
7	Djibouti	57.78
8	Algeria	50.00
9	Lebanon	44.44
10	Iraq	36.67
11	Yemen	27.78
12	Libya	25.56
13	Syria	22.22

Source: World Bank, Data on Statistical Capacity

Table 2: Interpretation of dominance tests for a 0.05 level of significance

p-values	Interpretation
$p \geq 0.05$ and $p' \geq 0.05$	$F^{U}_{Y_0}(y) = F^{L}_{Y_1}(y)$
$p < 0.05$ and $p' \geq 0.05$	$F^{U}_{Y_0}(y) \leq F^{L}_{Y_1}(y)\ \forall y \in [0, z^{+}]$
$p \geq 0.05$ and $p' < 0.05$	$F^{U}_{Y_0}(y) \geq F^{L}_{Y_1}(y)\ \forall y \in [0, z^{+}]$
$p < 0.05$ and $p' < 0.05$	$F^{U}_{Y_0}(y)$ and $F^{L}_{Y_1}(y)$ intersect

Table 3: Questions on income for 2016

Question	Number of observations per category
What is the total monthly income for all household members? Is it	70 non-responses
· Less than 500 USD	324
· 500 USD or more	1,106
You said your total household monthly income is less than 500 USD, is it	2 additional non-responses
· Less than 250 USD	84
· 250-300 USD	56
· 301-350 USD	51
· 351-400 USD	58
· 401-450 USD	41
· 451-500 USD	32
You said your total household monthly income is 500 USD or more, is it	16 additional non-responses
· 550 or less	40
· 551-600 USD	89
· 601-650 USD	44
· 651-700 USD	74
· 701-750 USD	179
· 751 USD or more	664
$n = 1,500$	

Table 4: Questions on income for 2018

Question	Number of observations per category
What is the total monthly income for all household members? Is it:	40 non-responses
Less than 1,000,000 LBP	448
1,000,000 LBP or more	1,912
You said your total household monthly income is less than 1,000,000 LBP, is it:	4 additional non-responses
450,000 LBP or less	67
451,000-700,000 LBP	183
751-000-1,000,000 LBP	194
You said your total household monthly income is 1,000,000 LBP or more, is it:	27 additional non-responses
1,000,000-1,500,000 LBP	216
1,500,001-2,000,000 LBP	614
2,000,001-2,500,000 LBP	464
2,500,001-3,000,000 LBP	330
3,000,001-4,000,000 LBP	162
4,000,001 LBP or more	99
$n = 2,400$	
1 USD=1,507.50 LBP	

Table 5: Questions on income for 2021-22

Question	Number of observations
What is your monthly household income in LBP?	
Lowest observation=200,000	1
Income $\in (200,000; 120,000,000)$	2,033
Highest observation = 120,000,000	2
Don't know	222
Refused to answer	141
$n = 2,399$	

Table 6: Consumer Price Index (December 2013 =100)

Time of interview	CPI
August 2016	95.61
October 2018	108.89
December 2021	921.40
January 2022	992.24
February 2022	961.15

Source: Central Administration of Statistics, Lebanon: Consumer Price Index - CPI

Table 7: Stochastic dominance tests on the distributions of income (missing at random)

Test	p-value
$H_0 : F^L_{Y_{2016}}(y) - F^U_{Y_{2018}}(y) \geq 0 \;\; \forall y \in [0, 5000000]$ $H_1 : F^L_{Y_{2016}}(y) - F^U_{Y_{2018}}(y) < 0, \;$ for some $y \in [0, 5000000]$	0.9915
$H'_0 : F^U_{Y_{2018}}(y) - F^L_{Y_{2016}}(y) \geq 0 \;\; \forall y \in [0, 5000000]$ $H'_1 : F^U_{Y_{2018}}(y) - F^L_{Y_{2016}}(y) < 0, \;$ for some $y \in [0, 5000000]$	0.0000
$H_0 : F^L_{Y_{2022}}(y) - F^U_{Y_{2018}}(y) \geq 0 \;\; \forall y \in [0, 5000000]$ $H_1 : F^L_{Y_{2022}}(y) - F^U_{Y_{2018}}(y) < 0, \;$ for some $y \in [0, 5000000]$	0.6294
$H'_0 : F^U_{Y_{2018}}(y) - F^L_{Y_{2022}}(y) \geq 0 \;\; \forall y \in [0, 5000000]$ $H'_1 : F^U_{Y_{2018}}(y) - F^L_{Y_{2022}}(y) < 0, \;$ for some $y \in [0, 5000000]$	0.0000
$H_0 : F^L_{Y_{2022}}(y) - F^U_{Y_{2016}}(y) \geq 0 \;\; \forall y \in [0, 5000000]$ $H_1 : F^L_{Y_{2022}}(y) - F^U_{Y_{2016}}(y) < 0, \;$ for some $y \in [0, 5000000]$	0.2093
$H'_0 : F^U_{Y_{2016}}(y) - F^L_{Y_{2022}}(y) \geq 0 \;\; \forall y \in [0, 5000000]$ $H'_1 : F^U_{Y_{2016}}(y) - F^L_{Y_{2022}}(y) < 0, \;$ for some $y \in [0, 5000000]$	0.0000

Table 8: Stochastic dominance tests on the bounds of distributions of income

Test	p-value
$H_0 : F^L_{Y_{2016}}(y) - F^U_{Y_{2018}}(y) \geq 0 \ \forall y \in [0, 5000000]$ $H_1 : F^L_{Y_{2016}}(y) - F^U_{Y_{2018}}(y) < 0, \text{ for some } y \in [0, 5000000]$	0.8636
$H'_0 : F^U_{Y_{2018}}(y) - F^L_{Y_{2016}}(y) \geq 0 \ \forall y \in [0, 5000000]$ $H'_1 : F^U_{Y_{2018}}(y) - F^L_{Y_{2016}}(y) < 0, \text{ for some } y \in [0, 5000000]$	0.0000
$H_0 : F^L_{Y_{2022}}(y) - F^U_{Y_{2018}}(y) \geq 0 \ \forall y \in [0, 1500000]$ $H_1 : F^L_{Y_{2022}}(y) - F^U_{Y_{2018}}(y) < 0, \text{ for some } y \in [0, 1500000]]$	0.8820
$H'_0 : F^U_{Y_{2018}}(y) - F^L_{Y_{2022}}(y) \geq 0 \ \forall y \in [0, 1500000]]$ $H'_1 : F^U_{Y_{2018}}(y) - F^L_{Y_{2022}}(y) < 0, \text{ for some } y \in [0, 1500000]]$	0.0000
$H_0 : F^L_{Y_{2022}}(y) - F^U_{Y_{2016}}(y) \geq 0 \ \forall y \in [0, 1500000]]$ $H_1 : F^L_{Y_{2022}}(y) - F^U_{Y_{2016}}(y) < 0, \text{ for some } y \in [0, 1500000]]$	0.9041
$H'_0 : F^U_{Y_{2016}}(y) - F^L_{Y_{2022}}(y) \geq 0 \ \forall y \in [0, 1500000]]$ $H'_1 : F^U_{Y_{2016}}(y) - F^L_{Y_{2022}}(y) < 0, \text{ for some } y \in [0, 1500000]]$	0.0000