HCEO WORKING PAPER SERIES

Monitoring Poverty in a Data Deprived Environment: The Case of Lebanon

Paul Makdissi
Walid Marrouch
Myra Yazbeck

Working Paper

2022-014

04/2022

HUMAN CAPITAL AND ECONOMIC OPPORTUNITY GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

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

Paul Makdissi[†]
University of Ottawa

Walid Marrouch[‡]
Lebanese American University

Myra Yazbeck[§]
University of Ottawa

April 2022

Abstract

This paper is motivated by the dearth of statistical capacity in the Middle East North Africa region and the unprecedented economic collapse in Lebanon. It proposes and illustrates a data augmentation approach to conduct poverty analysis in the absence of traditional sources of information on income distribution. Our approach shows that it is possible to exploit alternative data sources to conduct the much-needed poverty analysis. Building on available data augmentation techniques, we first recover the entire income distribution from the available interval data. Then we account for non-response and estimate the bounds of the set of admissible cumulative distributions of income. Finally, we analyze poverty dynamics using first-order dominance tests on the bounds of admissible cumulative distributions set. To illustrate the importance of the proposed approach, we apply this methodology to Lebanese data, provide a picture of poverty dynamics, and provide insights into the politico-economic dynamics preceding the economic collapse.

Keywords: Poverty dynamics, stochastic dominance, data deprivation, Lebanon
JEL Codes: I31, I32, O15

*The first author would like to thank Karim Emile Bitar for useful discussion on the Lebanese political system.

[†] 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; and School of Economics, University of Queensland, Australia; Email: myazbeck@uottawa.ca

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 essential. It is well-known that a country's statistical capacity is usually correlated with 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. This lack of statistical capacity prevents many countries in the region from developing much-needed evidence-based policies and is very 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. Building on this finding, Atamanov *et al.* (2020) advocate improving the statistical capacity in the MENA region. In addition, the authors suggest that researchers look for alternative information sources in parallel to those produced by countries' statistical agencies.[2] From this perspective, a good starting point would be to exploit all the data publicly available and propose new and original methods that allow researchers to overcome the issues related to data availability.

This paper aims to propose and illustrate an approach to conducting poverty analysis in the absence of (or lack of access to) household budget surveys. Our approach exploits publicly available data on income intervals and income sufficiency in the Arab Barometer as an alternative source of information on income distribution. More specifically, our approach is twofold. First, it uses the information on income intervals and constructs the underlying

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

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

continuous income distribution. Constructing the continuous income distribution allows us to conduct the much-needed poverty analysis, which allows us to depict the dynamics of poverty. Second, given that data on income categories may not always be available, we exploit ordinal information on income sufficiency as an alternative source of information on poverty. We thus provide an illustration that shows how to exploit information on income sufficiency. While the information on the income distribution is preferred, income sufficiency data capture a similar concern qualitatively and may be an excellent avenue to explore when income information is not collected.

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 illustration'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 Lebanese statistical agency 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 allow researchers to understand better poverty dynamics and the state of affairs starting from the period preceding the crisis (i.e., before the 2019 uprisings) to the present period.

To analyze the current poverty dynamics in Lebanon, we use the last three waves of the

Table 1: Statistical capacity in the MENA

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

Arab Barometer Survey that span from 2016 to 2021. The Arab Barometer Survey is the only widely available data that provides the necessary information to address our research questions. However, it presents two challenges. First, the income data is elicited in the form of intervals. Second, it suffers from a non-negligible number of non-response. To overcome the first challenge, we build on a data augmentation technique proposed by Groß *et al.* (2017), Walter and Weimer (2018), and Walter (2019) and construct the cumulative income distribution. In doing so, we exploit all the interval information on income available; that is, the information on income in the first two waves (in the third wave, this information is not available). To overcome 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) method to estimate the bounds of the sets of admissible cumulative distributions of income. To determine the dynamics of poverty, we test for first-order dominance on the bounds of the sets of admissible cumulative distributions in the spirit of Fakih *et al.* (2022).

Given that wellbeing is affected by income and is reflected in an individual's perception of deprivations, we complement our analyzes using the information on perceived income sufficiency available in all three survey waves. While using perceived income sufficiency allows expanding our analysis to investigate the poverty dynamic and include year 2021, the nature of the variable at hand is ordinal. We thus adapt the standard version of first-order stochastic dominance, and use the associated first-order dominance tests to address the latter issue (see Allison and Foster, 2004; Makdissi and Yazbeck, 2017). As in the case of the income intervals, we account for non-response by estimating the bounds for this ordinal variable's cumulative distribution and perform a first-order stochastic dominance test to determine the directional dynamics of poverty.

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. Our results also show that this decrease in the poverty levels was short-lived because the poverty levels spiked between 2019 and 2021, shortly after the elections. This increase in poverty levels counterbalanced the poverty reduction observed between 2016 and 2018, leading to poverty levels rising above those prevailing in 2016. Combined, these results are compatible with the hypothesis that politicians planned a debt-financed Ponzi scheme to ensure their re-election.[3] However, given that the high levels of expenditures it involved were unsustainable, the Ponzi scheme collapsed as soon as they were elected. As a result, this artificial poverty reduction vanished, inducing Lebanon into a huge financial disaster.

[3]In Lebanon, an implicit tripartite alliance between the government, central bank, and banks was created to provide a funding mechanism of 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 in the aftermath of 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/

While some of the effects obtained will be the result of a mix different contributing factors such as the financial collapse, the port explosion and the COVID19 pandemic, these results speak 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.[4] Indeed, these methods can be used in any survey where the information available is on income intervals or categorical (such as self-sufficiency). Moreover, these approaches may also be a good persuasion tool for the statistical agencies that are reluctant to share their data with academic researchers. The remainder of this paper is organized as follows. In Section 2, we describe the estimation and stochastic dominance testing strategy. Section 3 presents the political context and gives some information on the available data. In Section 4, we apply the estimation and testing approach to the Arab Barometer data for Lebanon to build a picture of the current dynamics of poverty in the country. 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 the absence continuous data usually available in living standard household survey data. 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 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

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

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 α 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. A solution was offered in Atkinson (1987) where the author shows that if one wishes to use a more robust approach, it is possible to identify poverty orderings that would remain valid for any choice of poverty index and any choice of poverty threshold if the following condition holds.

First order stochastic dominance condition. *A necessary and sufficient condition for no increase in poverty when moving from income distribution F_{Y_0} to income distribution F_{Y_1}, for any poverty line $z \in [0, z^+]$ and for any poverty index is achieved when:*

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

This stochastic dominance condition can be interpreted as follows. Suppose the cumulative distribution of interest is everywhere below the reference cumulative distribution for all income levels under z, then the proportion of the poor is lower in the distribution of interest. This result is valid for all potential poverty thresholds below z. Moreover, this dominance condition's result is not limited to the headcount index but also 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 this dominance condition using a small nationally representative survey that contains incomplete information on the income distributions and assess poverty dynamics in Lebanon. As mentioned earlier, the available data includes two types of information on income. The first type is in the form of a standard ratio scale variable on income intervals variable. The second type of information is ordinal, and contains information on income sufficiency. In both cases, we will use an adapted version of the first-order stochastic dominance condition proposed by Atkinson (1987) and assess the directional change in poverty.

2.1 Estimating the cumulative distribution's bound using income interval data

This section uses information on the households' total income for 2016 and 2018 and estimates the sets of admissible cumulative distribution functions of income. We use Walter and Weimer's (2018) estimation algorithm to account for the interval nature of the survey's income data available. We also build on their contribution by accounting for the potential non-random missingness in the data. We then test for stochastic dominance on the bounds of the two sets of admissible cumulative distributions. For continuous income information, there exists a simple estimator for the empirical cumulative distribution function (EDF):

$$\widehat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(y_i \leq y), \tag{2}$$

where $\mathbb{1}(\cdot)$ is an indicator function, n is the number of observations, and y_i is the income of observation i.[5] When the information on income takes the form of intervals, the direct use

[5]For non-random surveys, one can use 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.

of the estimator in equation (2) produces a step function CDF, 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). Another alternative 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 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 distribution that precisely match the proportion of observations below these income levels.[6]

We build on Walter and Weimer (2018) to estimate the bounds on the set of admissible CDF in the presence of non-response. This extension consists of a numerical integration of the estimated density function. 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 in the spirit of Horowitz and Manski (1995), 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.

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

[6] 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 $\int_0^{x_k} f(y) \mathrm{d}y = \widehat{F}_Y(x_k)$.

distribution

$$f_{Y|K}(y|k) = \begin{cases} \frac{f_Y(y)}{\Pr(k)} & \text{if } x_{k-1} \leq y \leq x_k \\ 0 & \text{otherwise} \end{cases} . \tag{3}$$

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, we draw 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 we force 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 $\int_{x_{k-1}}^{x_k} \widehat{f}_Y(y)\mathrm{d}y = \widehat{\Pr}(k)$.

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 draw pseudo-samples of y_j using the conditional distributions $f^S_{Y|K}(y|k)$ as sampling weights within each intervals. At each step of the algorithm, an estimated value of the $CDF^S(y)$ is computed using numerical integration of $\widehat{f}^S_Y(y)$:

$$\widehat{F}^S_Y(y) = \int_0^y \widehat{f}^S_Y(u)\mathrm{d}u, \quad S \in \{L, U\}. \tag{4}$$

By discarding the first B burned-in iterations and averaging the M following iterations, we estimate the expected value of the $CDF^S(y)$.[7] Since the algorithm is an adaptation based

[7]Since the estimation is based on a Markov Chain property of the process, we assume 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.

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

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

on three papers (Walter and Weimer, 2018, Walter, 2019; Groß, *et al.*, 2017), the detailed algorithm is provided in the appendix. This algorithm allows us to compute the expected $CDF^S(y)$ that is associated with the actual sample.

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. In the context of continuous income distributions, the test consists of testing

$$H_0 \ : \ F^L_{Y_0}(y) - F^U_{Y_1}(y) \geq 0 \quad \forall y \in [0, z^+]$$

$$H_1 \ : \ F^L_{Y_0}(y) - F^U_{Y_1}(y) < 0, \text{ for some } y \in [0, z^+]$$

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) and establish non-dominance (instead of strict dominance). 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. $H_0 : F^L_{Y_0}(y) - F^U_{Y_1}(y) \geq 0\ \forall y \in [0, z^+]$ and $H'_0 : F^U_{Y_1}(y) - F^L_{Y_0}(y) \geq 0\ \forall y \in [0, z^+]$. 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 statistics for the above tests: $\tau = \sup_p (F^U_{Y_1}(y) - F^L_{Y_0}(y))$. It is straightforward to construct a non parametric estimator of τ as follows:

$$\widehat{\tau} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sup_p \left(\widehat{F}^U_{Y_1}(y) - \widehat{F}^L_{Y_0}(y) \right) \tag{5}$$

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 an ordinal variable's cumulative distribution bounds using income sufficiency data

Estimating the bounds for the income sufficiency cumulative distribution (CDF) can be done directly by allocating non-responses to the lowest (highest) ordinal category for the upper (lower) bound. Each bound is estimated using a simple estimator for the discrete empirical cumulative distribution F_S. Again, we perform the stochastic dominance test on the bounds of two distributions. For this discrete ordinal distribution, the test consists of testing

$$\begin{aligned} H_0 &: F^L_{S_0}(s) - F^U_{S_1}(s) \geq 0 \quad \forall s \in [0, z^+] \\ H_1 &: F^L_{S_0}(s) - F^U_{S_1}(s) < 0, \text{ for some } s \in [0, z^+] \end{aligned}$$

The decision rules in Table 2 once again provide a guideline for the interpretation of these tests. At this point, one may wonder if such dominance tests may apply to ordinal data because analysts usually perform these tests on ratio-scale variables only. Nevertheless, despite the fact that such variables have unknown numerical scale, the underlying monotonicity of these ordinal categories allows us to apply a first-order dominance test and get a meaningful interpretation of the poverty dynamics (see Allison and Foster, 2004; Makdissi and Yazbeck, 2017).

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.[8] Soon after the IMF's memo, Banque du Liban (BDL) and Lebanese authorities accelerated their "financial engineering" operations.[9] In retrospect, these operations were nothing more than a regulated nationwide Ponzi scheme that was mislabelled as "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 LPB/USD).[10] 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. The public sector employs 14% of Lebanese workers (see CAS and ILO, 2019), and this non-negligible proportion of public servants is

[8] 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.

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

[10] 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

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,[11] 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 1990 ended the 1975-1990 Lebanese Civil War. Indeed, instead of addressing these distortions, the Taif Agreements enshrined a *redistributive kleptocracy.*[12] 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 then 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 this 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 multi-dimensional deprivation (waste management, water, electricity, sanitation, heath and education). In October 2019, a proposed tax (on gasoline, tobacco, and VoIP calls) affecting the disadvantaged population, combined with the publicly displayed opulent political leaders' lifestyle, triggered

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

[12] 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

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 Lebanese Central Administration of Statistics (CAS) is still reluctant to share data with academic researchers.

3.2 Data availabilty and challenges

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 accessible to academic researchers.[13] To overcome this hurdle, we use an alternative source of information, namely the Waves IV, V, and VI of the Arab Barometer surveys.[14] The interviews for Lebanon were conducted in 2016 for Wave IV, in September and October of 2018 for Wave V, and between September 29, 2020, and April 3, 2021, for Wave VI. 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 intervals and/or sufficiency. Despite the limited information on income and the relatively small sample size of these surveys, this paper shows that it is possible

[13] 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.

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

to make robust claims on changes in poverty in Lebanon between 2016 and 2021. 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

In this paper, we use the Arab Barometer (Lebanon) survey waves IV, V and VI. Usually these surveys are representative of the country's population. More specifically, Lebanon relies on the CAS' Public Housing and Population Census of 2011 and stratify the survey by governorate and sect to construct their sampling frame. Thus, in normal times, the survey team uses a stratified area probability sample and conducts face-to-face interviews. Nevertheless, because of the COVID-19 pandemic the survey team used mobile phone interviews for wave VI. The mobile phone numbers were called randomly using a list of 350,000 phone numbers stratified by governorate and sect based on the same sampling frame as waves IV and V.

The Arab Barometer Surveys contain two sets of questions that are useful in addressing our research question. The first set of questions focuses directly on income (see Tables 3 and 4), and the second on income sufficiency (see Table 5). While the first set of questions are available in the 2016-2017 and the 2018-2019 wave of the survey, they are not available in the last wave of the Arab Barometer, perhaps due to new constraints imposed by the COVID-19 pandemic. However, this income question for the 2020-2021 wave would not have been very useful in the case of Lebanon due to the presence of extreme inflation rates (131.05% in 2020 and 144.12% in 2021) and high exchange rate volatility.[15] The high monetary instability would have made it difficult to identify the real value of income without knowing

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

Table 3: Questions on income for 2016

Question	Number of observations per category
What is the total monthly income	70 non-responses
for all household members? Is it	
· Less than 500 USD	324
· 500 USD or more	1,106
You said your total household monthly income	2 additional non-responses
is less than 500 USD, is it	
· 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	16 additional non-responses
is 500 USD or more, is it	
· 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: Question on income sufficiency

Which of these statements comes closest to describing your net household income	2016	2018	2020/21
· 1: Our net household income does not cover our expenses; we face significant difficulties.	285	325	580
· 2: Our net household income does not cover our expenses; we face some difficulties.	467	936	1,420
· 3: Our net household income covers our expenses without notable difficulties.	581	903	770
· 4: Our net household income covers our expenses and we are able to save.	153	215	201
· Non-responses	14	21	29
Number of observations	$1,200$	$2,400$	$3,000$

the exact day of the interview. In addition, reporting income intervals would have been very challenging due to the multiplicity of exchange rates and the creation of a new currency, the "lollar" or "local dollar", with a value below that of the "parallel/actual" USD-LPB exchange rate market. Therefore, to gauge the extent to which poverty increased in 2020-2021, we leverage the information available in all three waves of the survey: the income sufficiency questions. We believe that in a highly volatile context, the income insufficiency provides the appropriate information for monitoring poverty and thus can be useful approach for other contexts similar to Lebanon's.

4 Results

4.1 Dominance Tests using Interval Income Data

This section will use the information available on income intervals to test for stochastic dominance of 2018 income's CDFs over 2016 income's CDFs and analyze the change in the income distribution between 2016 and 2018. Tables 3 and 4 display these survey questions and the distribution of the associated responses in the surveys. One interesting aspect of non-responses' distribution is that individuals are less (more) willing to answer more (less) precise questions on income. The number of non-responses increases when we move from a question with two income categories to questions with a finer grid of categories. To account for the non-response, we exploit this information structure 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" ("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

For 2016, we first converted the income categories bounds in 2018 LBP using the fixed exchange rate of 1,507.50 LBP/USD and the variation in the consumer price index between 2016 and 2018[16] and followed a similar approach using the corresponding income categories.

We first apply the estimation strategy of Section 2.1 to the data with non-responses

[16] The LBP/USD exchange rate was stable in these years it is only starting 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 the setting of a nationwide Ponzi scheme during the later years of that peg regime.

Figure 1: Bounds on the 2016 and 2018 cumulative income distributions

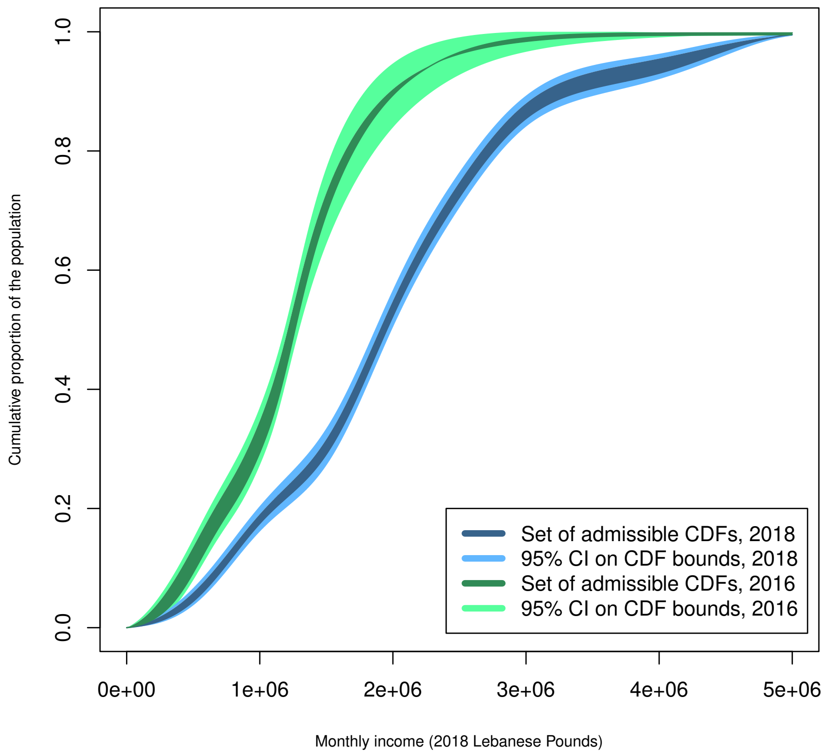


Table 6: Stochastic dominance tests on the distributions of income

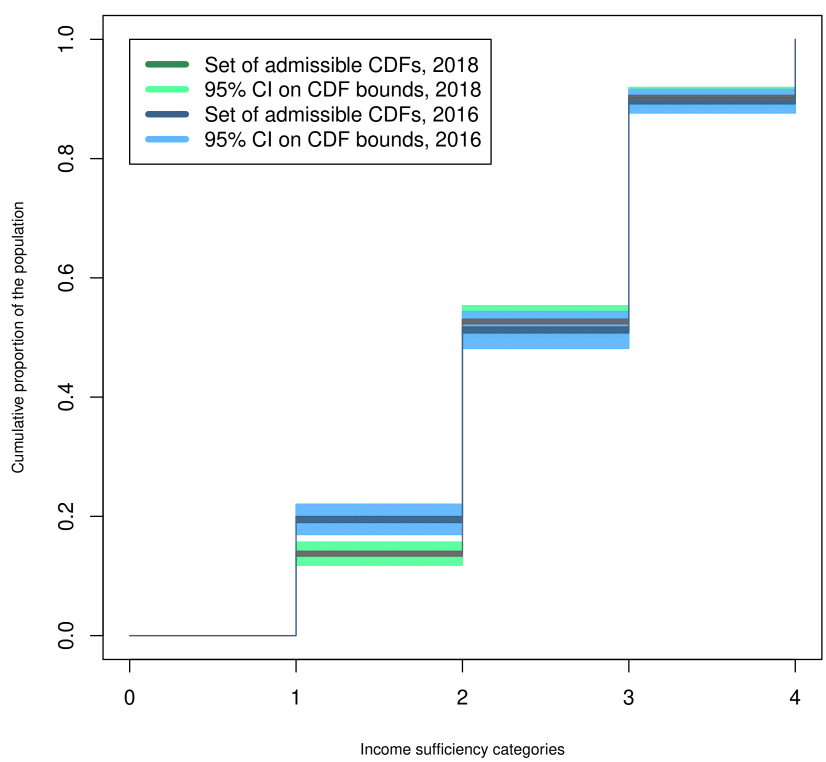
Test	p-value
$H_0 : F^L_{Y_{2016}}(y) - F^U_{Y_{2018}}(y) \geq 0 \ \forall y \in [0, 5000000]$ $H_1 : F^L_{Y_{2016}}(y) - F^U_{Y_{2018}}(y) < 0, \text{ for some } y \in [0, 5000000]$	0.9892
$H'_0 : F^U_{Y_{2018}}(y) - F^L_{Y_{2016}}(y) \geq 0 \ \forall y \in [0, 5000000]$ $H'_1 : F^U_{Y_{2018}}(y) - F^L_{Y_{2016}}(y) < 0, \text{ for some } y \in [0, 5000000]$	0.0000

allocated to income categories as explained earlier. We then 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 1 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 bootstrap test described in Section 2.1 with 999 replications. This stochastic dominance test confirms the result, the associated p-values in Table 6 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 any social welfare index increases between 2016 and 2018. 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.

4.2 Dominance Tests using Ordinal Income Sufficiency Data

To cross-check the results obtained from the analysis based on income intervals are reflected in the population's perception of their financial capacity, we focus on income sufficiency. Table 5 displays the survey question, its respective ordinal categories, and the distribution of responses and non-responses.[17] As in the case of income interval, we assume that non-random missingness for non-responses and compute lower and upper bounds. We allocate all non-responses to the category "Our net household income covers our expenses, and we are able to save" ("Our net household income does not cover our expenses; we face significant difficulties."), to produce the lower (upper) bound for the set of admissible CDFs.

Figure 2: Bounds on the 2016 and 2018 cumulative income sufficiency distributions



[17]To produce an increasing ordinal variable, the numbers associated with the different categories are inverted compared to the Arab Barometer.

Table 7: Stochastic dominance tests on the distributions of sufficiency of income

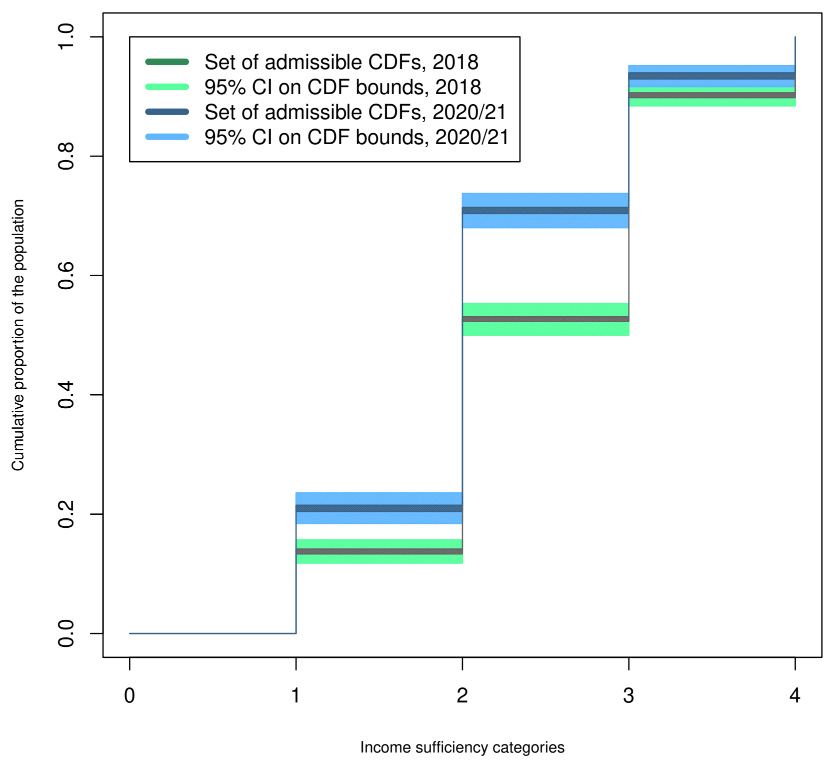
Test	p-value
$H_0 : F^L_{S_{2016}}(y) - F^U_{S_{2018}}(y) \geq 0 \ \forall s \in [0,4]$ $H_1 : F^L_{S_{2016}}(y) - F^U_{S_{2018}}(y) < 0, \text{ for some } s \in [0,4]$	0.1051
$H'_0 : F^U_{S_{2018}}(y) - F^L_{S_{2016}}(y) \geq 0 \ \forall s \in [0,4]$ $H'_1 : F^U_{S_{2018}}(y) - F^L_{S_{2016}}(y) < 0, \text{ for some } s \in [0,4]$	0.0040
$H_0 : F^L_{S_{2020/21}}(y) - F^U_{S_{2018}}(y) \geq 0 \ \forall s \in [0,4]$ $H_1 : F^L_{S_{2020/21}}(y) - F^U_{S_{2018}}(y) < 0, \text{ for some } s \in [0,4]$	1.0000
$H'_0 : F^U_{S_{2018}}(y) - F^L_{S_{2020/21}}(y) \geq 0 \ \forall s \in [0,4]$ $H'_1 : F^U_{S_{2018}}(y) - F^L_{S_{2020/21}}(y) < 0, \text{ for some } s \in [0,4]$	0.0000
$H_0 : F^L_{S_{2020/21}}(y) - F^U_{S_{2016}}(y) \geq 0 \ \forall s \in [0,4]$ $H_1 : F^L_{S_{2020/21}}(y) - F^U_{S_{2016}}(y) < 0, \text{ for some } s \in [0,4]$	0.9069
$H'_0 : F^U_{S_{2016}}(y) - F^L_{S_{2020/21}}(y) \geq 0 \ \forall s \in [0,4]$ $H'_1 : F^U_{S_{2016}}(y) - F^L_{S_{2020/21}}(y) < 0, \text{ for some } s \in [0,4]$	0.0000

Figure 2 displays these sets of admissible CDFs on income sufficiency for 2016 and 2018. A visual inspection of this figure indicates a reduction in the proportion of the population declaring to face significant difficulties to cover their expenses. Table 7 displays the p-values of the associated stochastic dominance test with 999 replications. 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 income sufficiency's CDF for 2018 first-order stochastically dominates the real income sufficiency's CDF for 2016. This result also indicates that any poverty or social welfare measure based on perceived income sufficiency would indicate a decrease in poverty and an increase in social welfare between 2016 and 2018. While this result is based on the perception of income sufficiency, it echoes the result based on the real distribution of income. It is also compatible with the change in confidence in public institutions that seems to have stabilized between 2016 and 2018 after a sharp decline between 2013 and 2016 (see Fakih *et al.*, 2022). All these suggestive evidences may be considered important elements that could explain how the acceleration of the Ponzi scheme between 2016 and 2018 has probably met its intended objective of ensuring the reelection of the same traditional political class in 2018.

Given that the Arab Barometer did not collect any information on income intervals between 2018 and 2020/2021, the only information available for us to assess the dynamics of poverty in Lebanon for this last period is income sufficiency. We thus rely on the evidence that both income sufficiency and interval income information qualitatively convey the same picture and analyze income sufficiency between 2018 and 2020/2021.

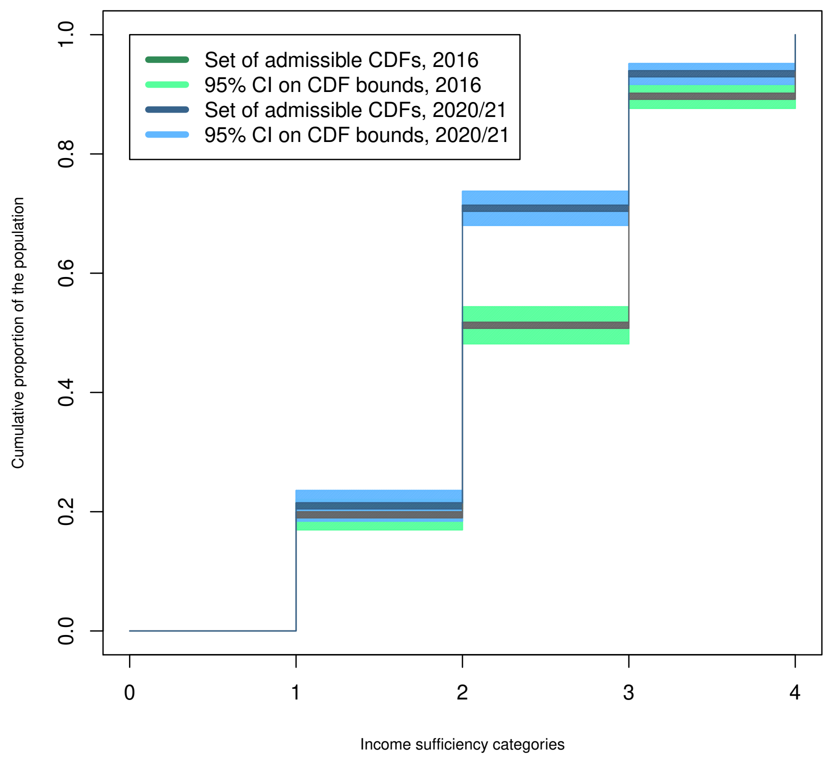
Figure 3 displays the sets of admissible CDFs for 2018 and 2020/21. A visual inspection of this figure indicates an increase in the proportion of people falling below each income sufficiency category. Table 7 displays the p-values of the associated stochastic dominance test. These p-values indicate that we cannot reject H_0 and that we can reject H_0'. Using

Figure 3: Bounds on the 2018 and 2021 cumulative income sufficiency distributions



the decision rules in Table 2, we can say that the real CDF of income sufficiency for 2018 first-order stochastically dominates the real CDF of income sufficiency for 2021. 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 2020/21. 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).

Figure 4: Bounds on the 2016 and 2021 cumulative income sufficiency distributions



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 2020/21. The graphical representation of this comparison presented in Figure 4 displays the sets of admissible CDFs for 2016 and 2020/21. A visual inspection of this figure indicates that the CDF of income sufficiency for 2016 first-order stochastically dominates the real CDF of income sufficiency for 2021. 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 real income sufficiency's CDF for 2016 first-order stochastically dominates the real income sufficiency's CDF for 2021. 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 2020/21. Thus there is a negative net effect of the pre-elections Ponzi scheme combined with the COVID-19 pandemic and the impact of the Beirut port blast. Thus, welfare is lower in 2020/21 compared to 2018, and it is 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) increasing thus inequalities.[18]

5 Conclusion

This paper is motivated by the 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 information for poverty analysis, our empirical illustration'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 case that illustrates the detrimental effect of the lack of data infrastructure necessary for

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

poverty monitoring. To overcome this issue, we adapt Walter and Weimer's (2018) and Walter's (2019) estimation method and derive bounds on admissible cumulative income distributions set while accounting for survey non-response and interval data on income. This approach allows us to analyze poverty dynamics using stochastic dominance tests readily available in the literature in a context where data on the income distribution 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. 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 that is 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 elections and collapsed shortly after the elections due to the uprising of the Lebanese population, who was struggling to make ends meet. 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 not available.

Finally, our empirical application exploits the additional information available 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 information on income and monitor the poverty dynamics in the MENA region.

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

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

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

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

[19] 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.

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

[21] 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.

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

[23] 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.

[24] 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 $\{g_1, g_2, \dots, g_J\}$, 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 $j \in \{1, 2, \dots, n\}$ 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.