

University of Groningen

Religion, social desirability bias and financial inclusion

Ahmad, Syedah; Lensink, Robert; Mueller, Annika

Published in:
Journal of Behavioral and Experimental Finance

DOI:
[10.1016/j.jbef.2023.100795](https://doi.org/10.1016/j.jbef.2023.100795)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Ahmad, S., Lensink, R., & Mueller, A. (2023). Religion, social desirability bias and financial inclusion: Evidence from a list experiment on Islamic (micro-)finance. *Journal of Behavioral and Experimental Finance*, 38, Article 100795. <https://doi.org/10.1016/j.jbef.2023.100795>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Full length article

Religion, social desirability bias and financial inclusion: Evidence from a list experiment on Islamic (micro-)finance[☆]Syedah Ahmad, Robert Lensink^{*}, Annika Mueller

Department of Economics, Econometrics and Finance, Faculty of Economics and Business, University of Groningen, The Netherlands

ARTICLE INFO

Article history:

Received 27 January 2022
 Received in revised form 23 December 2022
 Accepted 16 February 2023
 Available online 19 February 2023

JEL classification:

C83
 C93
 G28
 G51
 P43
 Z12

Keywords:

List experiment
 Islamic (micro-) finance
 Social desirability bias
 Field experiment
 Pakistan

ABSTRACT

In recent years, academic and policy research has placed increasing focus on the study of the attitudes of the Muslim poor towards Islamic and non-Islamic (micro-)finance to inform financial inclusion strategies. Survey questions are a common way to measure these attitudes and have been included into large-scale surveys such as the Global Findex. However, survey-based measures that ask about non-Islamic finance in an Islamic context may be affected by social desirability bias. In this paper, we propose a possible solution to this issue. We conduct the first list experiment designed to measure attitudes towards the usage of non-Islamic financial products and services, with 2,145 poor Muslims from Multan, Pakistan. Our list experiment uncovers that 37 percent of our sample use non-Islamic finance, almost twice as many as respond affirmatively to a similar direct survey question. Using our rich survey data on demographics, socio-economic factors and religiosity, we are further able to document substantial heterogeneity in the magnitudes of underreporting this usage in the direct survey question and in the usage of non-Islamic finance. We conclude by discussing the significance of our results in terms of (policy) implications and for the measurement of the demand for Islamic finance.

© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This paper reports the results of a list experiment – a technique used for eliciting truthful responses to sensitive questions – to study the attitudes towards the usage of non-Islamic (micro-)finance¹ of the Muslim poor in Pakistan, a group that is largely excluded from traditional banking services. Financial inclusion is known to play a key role in realizing 7 of the United Nation's 17 Sustainable Development Goals (SDGs) (World Bank, 2018). Prominent among these are SDG 1 (eradicating poverty), SDG 5 (achieving gender equality and economic empowerment of women), and SDG 10 (reducing inequality). Yet, despite enormous progress in terms of account ownership in recent years, and more than 60 countries devising and instituting financial inclusion strategies over the last decade, close to a third of all adults remains currently unbanked (Demirgüç-Kunt et al., 2018).

[☆] Funding: The research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

^{*} Corresponding author.

E-mail address: b.w.lensink@rug.nl (R. Lensink).

¹ Microfinance generally refers to financial products, e.g., small loans, and services that are being provided to individuals, households and micro-businesses, which are excluded from receiving financial products and services by the traditional/formal banking sector.

Strikingly, nearly half of these unbanked adults reside in only seven countries. Thus, realizing the SDGs will critically depend on the progress achieved in these countries in particular. It is notable that five of these seven countries, including Pakistan, rank highest in terms of total Muslim population, and that globally financial exclusion disproportionately affects the Muslim poor (Demirgüç-Kunt et al., 2014). Thus, devising effective and cost-efficient financial inclusion strategies mandates an evidence-based understanding of the barriers to financial inclusion that poor Muslim populations are facing. As religious Muslims may be less willing to use conventional, i.e., non-Islamic, (micro-)finance (El-Gamal et al., 2014),² policy-makers, academics and practitioners have recently shown an increased interest in the demand for and impact of (micro-)finance compliant with Islamic financial principles (Demirgüç-Kunt et al., 2014; Karim et al., 2008; World Bank, 2017). Thus, survey measures to gauge the attitudes of poor Muslim populations towards Shari'a-compliant (micro-)finance have been incorporated into several large-scale surveys, e.g., into a survey conducted by the Consultative Group to Assist the Poor

² Islamic microfinance complies with the principles of Islam/is Shari'a-compliant, while conventional (micro-)finance is not. E.g., conventional financial products often involve the receipt or payment of interest, which is considered *haram* (not permissible) according to the Quran.

(CGAP) (Karim et al., 2008), a number of surveys commissioned by the International Finance Corporation (Makhlouf, 2017), and the Global Findex (El-Gamal et al., 2014). However, little is known about how meaningful these measures are, as false reporting is a prominent weakness of the survey approach (Tourangeau, 2018). In particular, as survey responses for sensitive topics are known to suffer from social desirability bias (Tourangeau and Yan, 2007), we might expect there to be underreporting of demand for non-Islamic finance in an Islamic context, because such demand might be seen as less socially desirable in a society in which religion is of very high and increasing importance (also see Section 2.2). Evidence for what has been termed the “demand conundrum” (El-Zoghbi, 2013) – which is, the actual take-up of Shari’a-compliant products falling short of demand estimated from survey results – is in line with the presence of such social desirability bias.

We conduct – to the best of our knowledge – the first ever list experiment, a method to elicit truthful responses on sensitive topics, to measure attitudes towards the usage of non-Islamic financial products among a poor Muslim population in Pakistan. List experiments are designed to provide the data necessary to estimate the proportion of respondents in the population who agree with a sensitive statement. They do so by presenting an identical set of non-sensitive statements to (randomly assigned) treatment and control groups of respondents, with the treatment group being shown an additional sensitive statement, which in our case concerned non-Islamic finance. The key principle behind a list experiment is that respondents need to report only the number of statements that they agree with, which guarantees a higher level of anonymity than direct survey questions and makes respondents more likely to answer truthfully.

Our list experiment yields four main findings. First, we find from our list experiment that in a sample of 2,145 individuals, who are for the most part excluded from the formal financial system, a sizeable fraction – about 37% – uses non-Islamic financial products and services. Second, we establish significant underreporting of the usage of non-Islamic financial products and services in our sample, i.e., compared to the proportion of individuals who answer affirmatively to a similar direct survey question almost twice as many affirm the sensitive statement in the list experiment. Third, there is notable heterogeneity in this underreporting. That is, direct survey questions are not able to accurately reflect the patterns of usage of non-Islamic finance across different sub-groups of the population, as the magnitudes of underreporting vary between them. Fourth, using our rich survey data on demographics, socio-economic factors and religiosity, we are able to document substantial heterogeneity in the usage of non-Islamic finance across different sub-groups of the population.

Our results imply that direct survey questions do not appear to be suitable to correctly inform the ongoing discussions in both research and policy domains about the demand for, and cross-sectional patterns of the demand for, Shari’a-compliant finance; whereas list experiments present one possibility to overcome the issue of social desirability bias using an effective and scalable³ approach. This is a crucial contribution that is relevant for public policy debates around the globe, such as those surrounding the proper design and implementation of costly financial inclusion strategies to achieve the SDGs.

The rest of the paper is organized as follows. We first review relevant literature and provide contextual background regarding

³ By scalable, we mean that list experiments can be administered as part of large-scale surveys without encountering a number of logistical hurdles associated with scaling up alternative survey- and non-survey-based methods, e.g., high monetary costs, difficulty of implementation in the field, etc. We discuss these issues further in the conclusion.

Pakistan. Next, we describe our data, outline our methodology, and report validity checks of the list experiment data. Subsequently, we present our empirical results, and finally provide a discussion and conclude.

2. Literature review and country context

This section first surveys the relevant literature. It then provides some relevant background information about Pakistan, with a particular focus on religion and financial inclusion, to establish why it is a suitable location for our study and to motivate some of the dimensions along which we conduct our later heterogeneity analysis.

2.1. Social desirability bias and list experiments

Obtaining truthful responses to sensitive questions is a long-standing and important issue in the survey methodology literature (Höglinger et al., 2016). A frequently encountered type of sensitive question is one that asks about the possible violation of a social norm (Tourangeau and Yan, 2007). While some respondents may decide to omit answering the sensitive question or to opt out of participation in the survey altogether, others may provide a biased response, thereby placing themselves in the socially desirable category, in their own eyes, i.e., by practicing self-deception, or in the eyes of others, i.e., by engaging in “impression management” (Paulhus, 1986; Tourangeau, 2018). The provision of biased responses, i.e., the presence of “social desirability bias”, is seen as the main concern when it comes to collecting data on sensitive topics (Tourangeau, 2018), and was described in the seminal work of Sudman and Bradburn (1974, pp. 9–10) as follows: “If the respondent has a socially undesirable attitude or if he has engaged in socially undesirable behavior, he may [...] desire to appear [...] to be in the socially desirable category. It is frequently assumed that most respondents resolve this conflict in favor of biasing their answer in the direction of social desirability”. Note that social desirability bias entails ex-ante unpredictable variation in the degree of underreporting, e.g., because different segments of the sample might have different perceptions of the degree of sensitivity (Höglinger et al., 2016). We address these issues in our empirical analysis.

Validation studies (in which answers to sensitive questions are compared to accurate information about the respondents that is known to the surveyor from other sources) show that the prevalence of social desirability bias can be substantial for a given sample. For instance two studies find that 42 percent (Preisdörfer and Wolter, 2014) and 75 percent (van der Heijden et al., 2000) of individuals surveyed displayed social desirability bias in face-to-face interviews.

Several qualitative (Blattman et al., 2016) and quantitative (Bullock et al., 2011) approaches have been developed to mitigate the issues discussed above; these comprise self-administered questions, open-ended questions, and a number of more recently developed methods, prominent among them being the list experiment technique. List experiments have been shown to provide more accurate responses to sensitive questions,⁴ compared to direct questioning (Blair et al., 2020b; Ehler et al., 2021; Li and van den Noortgate, 2022), and have been employed across various disciplines, to understand phenomena such as voters’ attitudes (Redlawsk et al., 2010), prejudice (Janus, 2010), illegal behavior (Biemer et al., 2005), and more recently, in economics, for studying topics such as how microfinance loans are used by borrowers (Eriksen and Lensink, 2015; Karlan and Zinman, 2012).

⁴ Glynn (2013, p. 159) provides the following overview over topics of an inherently sensitive nature: “Sex, drugs, crime, religion, race, and politics are all inherently sensitive subjects”.

In the context of religion, the existence of social desirability bias for sensitive questions has been investigated, e.g., in the context of reporting attendance of religious services (Presser and Stinson, 1998), as well as in the context of voter discrimination against candidates who belong to a religious minority (Kane et al., 2004; Benson et al., 2011). Our study is the first to uncover and discuss underreporting bias in the context of attitudes towards Islamic (micro-)finance using a list experiment.

A list experiment attempts to elicit truthful responses through indirect questioning, by asking respondents about how many binary questions on a list they answer in the affirmative. Thus, a list experiment provides an aggregate in the form of the proportion of respondents who agree or disagree with a sensitive statement. This anonymity rules out the possibility of correlating individual-level characteristics to agreement with the sensitive statement at the individual level. However, sub-group analyses can still be conducted (Ahart and Sackett, 2004): a list experiment is well-suited to investigate heterogeneity in the socially less desirable behavior across sub-groups; and, since the degree of underreporting compared to a similar direct question can also be determined, it is further suitable for exploring heterogeneity in underreporting across sub-groups.

2.2. Country background – Pakistan

Pakistan's population size has approximately doubled over the last three decades, despite population growth slowing since the 1980s. With a population of 216.5 million in 2019 (World Bank, 2020), Pakistan now is one of the most populous nations in the world.

Poverty rates in Pakistan are high: In 2015, a quarter of the population lived below the national poverty line (World Bank, 2020). While Pakistan adopted financial inclusion as a national priority well before many other low- and middle-income countries, progress has been significantly slower than expected (the so-called “Pakistan Enigma” of financial inclusion (Rasmussen, 2018)). In fact, according to the 2017 edition of the Global Findex, average account ownership in Pakistan is only 20%, which is the lowest among lower- and middle-income countries, compared to 63% across developing economies. This is the case even though Pakistan has a large microfinance sector that is, however, concentrated in the Sindh and Punjab provinces. The three kinds of microfinance institutions operating in Pakistan are Rural Support Programmes, non-governmental organizations, and microfinance banks, but only the latter category is supported by a favorable and robust regulatory regime under the auspices of the State Bank of Pakistan (Kustin, 2015). Interestingly, Islamic microfinance institutions – while expected to grow substantially in terms of market size, market share and market penetration over the course of the next years – currently comprise a niche segment (Pakistan Microfinance Network, 2020), despite 96.3% of Pakistanis being Muslim (Pakistan Bureau of Statistics, 2020), and Pakistan having the second largest Muslim population in the world, after Indonesia. For 2017, Ahmad et al. (2020) report that while 13 out of a total of 101 Islamic MFIs globally operated in Pakistan, market size in terms of total assets and market share in terms of financial revenues for Islamic MFIs were only approximately one-eighth of those of non-Islamic MFIs.⁵

Note that while Pakistan became an Islamic republic in 1956 and religion has always been important to its population, the

self-stated importance of religion has increased further over the last two decades: according to the World Values Survey (WVS), the percentage of the population reporting religion to be “very important” increased by 10 percentage points (from 80% to 90%) between the 1994–98 and 2017–2020 waves of the WVS. While there are some differences along gender lines (e.g., men being more likely than women to attend religious services weekly), men and women do not differ statistically along most religious dimensions (e.g., along their views on whether religion is very important in their lives, daily prayers, belief in heaven, belief in hell, etc. (Pew Research Center, 2016)). It is noteworthy that the younger generation generally finds religion no less important than older generations: according to data from the 2017–2020 wave of the WVS, the percentages of respondents reporting religion to be very important were roughly the same in all age brackets (World Values Survey, 2020). However, the importance of religion notably varies with income levels, monotonically decreasing from 95.2% for low-income respondents to 80.2% for high-income respondents (World Values Survey, 2020). We explore these and other possible sources of heterogeneity in the use of non-Islamic finance, and its underreporting.

3. Methodology and data

This section provides: (i) an overview over the methodology employed, in particular, it describes the list experiment as well as the survey that we use for our analysis; and (ii) a discussion of the data, including the results of several data validity checks.

3.1. Methodology

We conducted a list experiment at the end of a larger survey among a sample of microloan borrowers of an NGO called Akhuwat in Pakistan (further details are provided below in the “Data” sub-section). We will first discuss the details of the list experiment, before outlining the scope of the survey.

3.1.1. List experiment

In a list experiment, participants are randomized into a treatment and a control group. The control group receives a list comprising exclusively of non-sensitive statements, whereas the treatment group receives the same list with an additional sensitive statement, which in our case concerned non-Islamic financial products. By letting respondents report only the number of statements that they agree with, the answer to the sensitive item is not revealed (provided that not all items are answered negatively or in the affirmative). The idea behind using the list experiment is that this higher level of anonymity and respondent protection makes respondents answer more truthfully. Given a large enough sample size, under certain conditions that are discussed in the “Data” sub-section, the difference in the mean number of supported statements between the two groups provides an estimate of the proportion of respondents in the population who agree with the sensitive statement.

The details of our list experiment are as follows. One of several enumerators employed for the study, none of whom had any relation to Akhuwat (a fact that was made clear to respondents during the introduction of our project to them and the consenting), explained the procedure of the list experiment to a participant of the control (treatment) group in a one-on-one setting that guaranteed privacy, by reading the following statement:

“Now I will read four (five) statements that apply to some people but not to others. While I am reading these statements to you, please count how many of them are true for you. Do not count loudly or count on your fingers. After I have read all four (five), just tell me

⁵ The market concentration of Islamic MFIs in other Organization of Islamic Cooperation (OIC) member states was approximately one-sixth in comparison to non-Islamic MFIs. In non-OIC member states, the market share for Islamic MFIs is reported to be 115 million USD in comparison to 5,581 million USD in 2017 (Ahmad et al., 2020).

HOW MANY of these apply to you – none, 1, 2, 3, or 4 (none, 1, 2, 3, 4, or 5). I do not want to know which ones, just HOW MANY.”

Then, after ensuring that the respondent understood these instructions, the following four non-sensitive statements were read to each member of the control group individually, in the following order:

- “A salaried job would suit me more than running my own business as it would guarantee a regular, predictable pay and provide me with a sense of security.”
- “I enjoy running my own business as it grants me a flexible lifestyle and I am the one in control: I have choices and get to make decisions.”
- “I care for eating home-cooked meals, as they are very nutritious and healthy.”
- “I care for eating meals prepared by roadside restaurants, as doing so is convenient and the food is tasty.”

The same four statements were read to each member of the treatment group individually, with a fifth, sensitive, statement added in the third position:

- “I use formal or informal non-Islamic financial products or services from time to time, such as bringing jewelry or a vehicle to a pawn shop and retrieving it later by paying interest on the loan”.

All five statements were selected on the basis of local context specific to the subjects (e.g., that pawn shops are a widely known type of non-Islamic financial institution in the area). More specifically, the first two non-sensitive statements were paraphrased from statements made by participants of focus group discussions (FGDs) conducted by us. These FGDs (15 in total, with 5–10 participants in each group, each group discussion lasting 45–60 min) involved a total of 130 Akhuwat “Family Enterprise Loan” holders in Multan, who were recruited using convenience sampling of participants in five loan disbursement meetings of Akhuwat branches. All five statements were finalized after consultation with the regional manager of Akhuwat before being piloted.⁶ The pilot was conducted in September 2017, at one randomly chosen branch out of the 25 Akhuwat branches in the Multan area.⁷

3.1.2. Survey

The questionnaire of the survey focused on demographic and socio-economic variables as well as measures of religiosity. A full list of the variables used in our analysis, along with a brief description of each, is presented in Table 1. The survey further contained a question closely related to the sensitive question of the list experiment:

- “Would you become or are you a customer of a financial institution that is non-Islamic?”.

Note that it was elucidated to the participants that the term ‘financial institution’ comprises both formal and informal financial institutions. Since both the sensitive statement and the direct question were part of the same survey, the latter was deliberately not phrased in the same manner as the sensitive statement in the list experiment, to avoid concerns about one response leading to the other, and also to avoid revealing the intent behind the list experiment. It should be emphasized that due to their wording, the difference in responses to the two questions can be seen as identifying a lower bound of underreporting: If responses to

⁶ Note that the non-sensitive items used in the study do not well align with several general best practice principles of selecting such statements, e.g., they are comparatively long and include terms that are open to interpretation (e.g., “sense of security”).

⁷ Both the FGDs as well as the pilot not only included elements concerning the list experiment, but also other ones, which were utilized for a larger project on financial inclusion and microfinance (e.g., we also piloted a baseline questionnaire for this larger project when we piloted the list experiment). Further details about this project can be found in Ahmad et al. (2020).

the direct question include actions and potential actions, while responses to the sensitive question only include actions, the difference will underestimate underreporting.

3.2. Data

Data collection took place between September 16 and October 31, 2017, in Multan, Pakistan, which is located in the Punjab province. Our sample consisted of all 2,220 current loan applicants to Akhuwat in the Family Enterprise Loan category in the area, whose loan had already been approved at the time of data collection. Akhuwat is one of the two largest Shari’a-compliant-microfinance providers in Pakistan.⁸ The loan disbursements were scheduled to take place through the 25 branches of Akhuwat in Multan. Out of these, we randomly chose one branch for piloting, and hence our final sample consisted of borrowers from the 24 remaining branches of Akhuwat in the area. We randomized subjects into a treatment and a control group, stratified by branch. Some further details about how the individuals were contacted are as follows. Starting with the list of 2,220 approved loan applicants to Akhuwat referred to above, we contacted the persons on the list by telephone and by loudspeaker announcements in the vicinity of locations where consenting would take place (see below), inviting them to take part in a research study whose objective was described as exploring human behavior from the point of view of culture. The introduction and consenting took place at locations that were also used as loan disbursement locations, but prior to the disbursements. The survey analyzed in this paper, and the list experiment, were conducted after the disbursements, one-on-one, either at the location of consent, or at a mutually agreed upon location and time (i.e., an appointment was set-up with the participants at their homes or workplace) that guaranteed privacy. During the introduction and seeking of consent for the study, we clearly communicated to the participants that they were taking part in a European research study unrelated to Akhuwat and that the staff we hired to help us conduct the study were also unrelated to Akhuwat.

Three key assumptions underpin the validity of a list experiment (Imai, 2011). First, that the randomization procedure was effective, which for our case can be seen from Table 2 showing that the means of key variables used in our analysis are not statistically significantly different between treatment and control groups.

Second, that there are no “design effects”, i.e., answers to control statements are not affected by the inclusion of the sensitive statement. We test this statistically using the appropriate likelihood ratio test (Blair and Imai, 2012) and cannot reject the null hypothesis of “no design effects” (p -value = 0.77).

⁸ Given that, a few points may be noted about the characteristics of the sample from the summary statistics of Table 2. While the sample of participants are by and large quite religious – approximately 50% attend religious gatherings at least once a week (see summary statistics in Table 2) – according to data from the World Values Survey (2020) the corresponding percentage for Pakistan overall is 47%, for countries that belong to the Organization of Islamic Cooperation (OIC) is 44%, and for Muslims in non-OIC countries is 52%. Thus, in terms of this measure of religiosity (attending religious gatherings at least once a week), individuals in our sample are quite comparable to the average Pakistani, as well as to populations in other Muslim-majority countries and Muslim populations in non-Muslim-majority countries. Also note that our sample consists of 87% of females. This, however, does not necessarily constitute a bias as, at the time of the study, approximately 80% of microfinance customers worldwide were female, with even higher shares in the years just prior (authors’ own calculation using the Microfinance Information Exchange, or MIX Market, database). Thus, as our study concerns financial inclusion strategies (particularly as they relate to the Muslim poor) and microfinance, which does have a strong focus on female customers, this sample is very relevant in the context of our research topic both from an academic and policy perspective.

Table 1
Description of variables used in the analysis.

Variable (abbreviation)	Description
Dependent variables	
Client of non-Islamic financial institution (Direct survey question)	1 = respondent reports that they are or may become a customer of non-Islamic financial institutions, 0 otherwise
List experiment	Number of positive responses from the list experiment
Main explanatory variable	
List treatment	1 = respondent is in the treatment group and received the sensitive statement, 0 otherwise
Variables for heterogeneity analysis	
Female	1 = female, 0 = male
Education above median (Educ. abv. Median)	1 = respondent falls above the median (grade 3) in terms of the highest grade of schooling completed, 0 otherwise
Formal education (Formal educ.)	1 = respondent's education is larger than 0 in terms of the highest grade of schooling completed, 0 otherwise
Read Urdu	1 = respondent is able to read Urdu, 0 otherwise
Richest household (Richest hh)	1 = respondent's overall monthly household income lies above 30,000 Pakistani Rupee (PKR), 0 otherwise
Major income household (Maj. inc. hh)	1 = respondent's overall monthly household income is at least 20,000 PKR, 0 otherwise
Household owns land (HH own land)	1 = respondent or the respondent's household owns land, 0 otherwise
Recite Quran daily (Quran daily)	1 = respondent is reading the Quran daily, 0 otherwise
Recite Quran daily or weekly (Quran daily/weekly)	1 = respondent is reading the Quran daily or weekly, 0 otherwise
Recite Quran daily or weekly or monthly (Quran daily/weekly/monthly)	1 = respondent is reading the Quran daily, weekly or on a monthly basis, 0 otherwise
Attend gathering at least once a week	1 = respondent attends religious gatherings at least once a week, 0 otherwise
Attend gathering more than once a month	1 = respondent attends religious gatherings more than once a month, 0 otherwise
Always consult scholar	1 = respondent always consults a Sheikh (religious scholar) for verifying the validity of financial products, 0 otherwise

Notes: The data are from the authors' own survey.

Table 2
Summary statistics and balance tests.

Variables	Overall sample		Control group			Treatment group			p-values
	N	Mean	N	Mean	SD	N	Mean	SD	
Demographics									
Female	2,220	0.87	1,108	0.87	0.33	1,112	0.87	0.33	0.92
Socio-Economic Factors									
Educ. abv. Median	1,972	0.46	980	0.45	0.50	992	0.46	0.50	0.94
Formal educ.	1,972	0.57	980	0.57	0.50	992	0.57	0.49	0.96
Read Urdu	2,147	0.53	1,073	0.53	0.50	1,074	0.54	0.50	0.68
Richest hh	2,106	0.33	1,051	0.34	0.48	1,055	0.32	0.47	0.19
Maj. inc. hh	2,106	0.81	1,051	0.81	0.39	1,055	0.80	0.40	0.86
HH own land	2,146	0.58	1,073	0.58	0.49	1,073	0.59	0.49	0.69
Religiosity									
Quran daily	2,144	0.22	1,072	0.21	0.41	1,072	0.23	0.42	0.37
Quran daily/weekly	2,144	0.47	1,072	0.46	0.50	1,072	0.48	0.50	0.39
Quran daily/weekly/monthly	2,144	0.66	1,072	0.67	0.47	1,072	0.66	0.47	0.86
Attend gathering at least once a week	2,143	0.48	1,071	0.48	0.50	1,072	0.48	0.50	0.98
Attend gathering more than once a month	2,143	0.76	1,071	0.77	0.42	1,072	0.75	0.43	0.25
Always consult scholar	1,753	0.12	875	0.11	0.32	878	0.13	0.34	0.36

Notes: The table reports the number of observations (N, where the unit of observation is an individual) and means for the overall sample and the number of observations (N, where the unit of observation is an individual), means, and standard deviations (SD) for the treatment and control groups. For each variable, the right-most column displays the p-value from a t-test on the equality of the means for the treatment and control groups. All variables are dummy variables. **Female:** 1 = female, 0 = male. **Educ. abv. Median:** 1 = respondent falls above the median (grade 3) in terms of the highest grade of schooling completed, 0 otherwise. **Formal educ.:** 1 = respondent's education is larger than 0 in terms of the highest grade of schooling completed, 0 otherwise. **Read Urdu:** 1 = respondent is able to read Urdu, 0 otherwise. **Richest hh:** 1 = respondent's overall monthly household income lies above 30,000 Pakistani Rupee (PKR), 0 otherwise. **Maj. inc. hh:** 1 = respondent's overall monthly household income is at least 20,000 PKR, 0 otherwise. **HH own land:** 1 = respondent or the respondent's household owns land, 0 otherwise. **Quran daily:** 1 = respondent is reading the Quran daily, 0 otherwise. **Quran daily/weekly:** 1 = respondent is reading the Quran daily or weekly, 0 otherwise. **Quran daily/weekly/monthly:** 1 = respondent is reading the Quran daily, weekly or on a monthly basis, 0 otherwise. **Attend gathering at least once a week:** 1 = respondent attends religious gatherings at least once a week, 0 otherwise. **Attend gathering more than once a month:** 1 = respondent attends religious gatherings more than once a month, 0 otherwise. **Always consult scholar:** 1 = respondent always consults a Sheikh (religious scholar) for verifying the validity of financial products, 0 otherwise. The data are from the authors' own survey.

Third, we consider the possibility that respondents might lie about the sensitive statement or otherwise manipulate their answers, conditional on having realized the mechanism behind the list experiment technique. The likelihood of such behavior increases when respondents expect their privacy to be violated, which is the case when affirmative responses to the sensitive statement are easy to identify, most prominently via the so-called ceiling and floor effects. Ceiling and floor effects refer, respectively, to respondents answering all control statements positively or negatively. To minimize the likelihood of such lies/answer manipulations, we chose negatively correlated pairs of control statements – the pairs being formed by the first two and the latter two statements – such that it was possible to answer both, one, or neither affirmatively. This, for all practical purposes, eliminated the possibility of ceiling and floor effects: Only 0.47% of the control group of 1,073 respondents agreed with zero control statements, and only 0.19% agreed with all control statements.

4. Empirical analysis

4.1. Usage of non-Islamic finance

We first estimate the proportion of individuals using non-Islamic finance according to the direct survey question and the list experiment. To do so, we use the Seemingly Unrelated Regressions (SUR) technique (see Wooldridge (2010), pp. 185–191), in order to account for the fact that the same respondents were asked the direct question and the list question (which might have led to correlation between the responses). Briefly, SUR estimates the regression models for different dependent variables (in our case, the response to the direct question and the count of the list response) simultaneously while allowing them to be correlated through a (possibly) correlated error term. The usefulness of SUR in our application stems from the fact that the covariates in each model can be different, yet statistical comparisons of coefficients can be made across the two models. The dependent variable in the first regression of the system is a dummy that takes the value 1 if the respondent reports that they are or may become a customer of non-Islamic financial institutions (0 otherwise; see Table 1), which is simply regressed on a constant (i.e., a vector of 1 s).

$$\text{Direct Question} = \alpha + \varepsilon_d$$

The intercept from this regression can be interpreted as the proportion of respondents using non-Islamic finance according to the direct survey question.

The dependent variable for the second regression in the SUR system is the count variable *list experiment* (see Table 1) which shows the number of affirmative responses to the itemized questions in the list experiment.

$$\text{List Experiment} = g + \gamma_1 \text{List Treatment} + \varepsilon_l$$

The proportion of subjects who, according to the list experiment, use non-Islamic financial products or services (whether formal or informal) can then be estimated as the coefficient on the dummy variable *list treatment* that takes the value 1 if the respondent is part of the treatment group that is presented with 5 statements (i.e., the sensitive statement and the 4 statements received by the control group). The constant of the above regression can be interpreted as the mean of the dependent variable for the control group that is presented with 4 statements. Note that the SUR technique allows ε_d and ε_l to be correlated.

The results of estimating the two-equation SUR system are reported in Table 3. The number reported in the first (“Direct survey question”) panel of Table 3, which is the constant α , shows that 19.6% use non-Islamic finance according to the direct survey

question. The number reported in the second (“List experiment”) panel of Table 3 is the coefficient γ_1 , which suggests that 37.1% of subjects – according to the list experiment – use non-Islamic financial products or services (whether formal or informal).⁹ Corresponding SUR standard errors and p-values from a two-sided z-test are reported below all the estimates. The panel labeled “Test on equality of SUR coefficients” reports a χ^2 statistic and p-value from a test on the equality of the proportions estimated from the direct question and the list experiment using SUR. The χ^2 statistic and p-value from the test on the equality of these two proportions (which takes into account that the subjects are the same for the direct question and for the list experiment), suggest that the proportions are significantly different at the 1%-level. Thus, the SUR estimates suggest that a sizably larger fraction of the sample, 37.1%, uses non-Islamic financial products according to the list experiment compared to the estimate from the closely related direct question from our survey, 19.6%, which points to potential underreporting of the use of non-Islamic finance due to social desirability bias. In terms of economic magnitude, almost twice as many individuals answer affirmatively in the case of the list experiment. It is worth reminding the reader that this difference in responses to the two questions establishes a lower bound for underreporting since (due to their respective wording; see the “Methodology” sub-section) responses to the direct question might include actual and potential usage, while responses to the sensitive question only include actual usage.

The evidence presented in Table 3 shows that there is substantial underreporting in the overall sample when using the direct question. Notably, we now present evidence that the direct question is also not able to correctly depict the pattern of heterogeneity across sub-samples of the data. To do so, we once again utilize a system of two Seemingly Unrelated Regressions:

$$\text{Direct Question} = \alpha + \beta_1 \text{Explanatory variable} + \varepsilon_d$$

$$\text{List Experiment} = g + \gamma_1 \text{List Treatment} + \gamma_2 \text{Explanatory variable} + \gamma_3 \text{List Treatment} * \text{Explanatory variable} + \varepsilon_l$$

In the above regressions, the *Explanatory variable* is a dummy indicating a particular sub-sample of the data (e.g., the sub-sample of individuals who read the Quran daily). As noted earlier, SUR allows ε_d and ε_l to be correlated. For this analysis, we are mainly interested in the coefficients β_1 and γ_3 , which have similar interpretations. β_1 is the difference in the proportion using non-Islamic finance according to the direct survey question, between the sub-sample indicated by the dummy variable being 1 compared to the sub-sample indicated by the dummy variable being 0 (e.g., a comparison of the proportion in the sub-sample of individuals who read the Quran daily versus those who do not). Similarly, γ_3 is the difference in the proportion using non-Islamic finance according to the list experiment, between the sub-sample indicated by the dummy variable being 1 compared to the sub-sample indicated by the dummy variable being 0 (e.g., a comparison of the list experiment estimated proportion in the sub-sample of individuals who read the Quran daily versus those who do not).

The results of the SUR estimation are reported in Table 4. The numbers reported in the row labeled “Direct survey question”, which are the coefficients β_1 shown for three sub-samples of the data that are defined by the three dummy variables described in the table notes. The numbers reported in the row labeled “List experiment interaction” of Table 4 are the coefficients γ_3 , shown for the aforementioned three sub-samples of the data. The row

⁹ The constant of this regression, g , which can be interpreted as the mean of the dependent variable for the control group that is presented with 4 statements, is reported below the proportion from the list experiment and is found to have a value of 1.978.

Table 3

Proportion engaging in sensitive behavior.

Direct survey question		
- Proportion of subjects that are or may become customers of non-Islamic financial institutions		0.1995
	SE	(0.009)
	P> z	0.000
	N	2,145
List experiment (N = 2,145)		
- Proportion from list experiment (list treatment)		0.371
	SE	(0.020)
	P> z	0.000
- Mean of control group		1.978
	SE	(0.014)
	P> z	0.000
Test on equality of SUR coefficients		
- χ^2 test on the equality of SUR-estimated proportions from direct survey question and list experiment	χ^2 statistic	61.26
	Prob > χ^2	0.000

Notes: The table reports the comparison between the results from the direct survey question and the list experiment for the full sample using SUR. The proportion reported for the “Direct survey question” panel is the mean, estimated as the constant from a SUR regression, of a dummy that takes the value 1 if the respondent reports that they are or may become a customer of non-Islamic financial institutions (and 0 otherwise). The proportion reported in the “List experiment” panel can be interpreted as the proportion of subjects who, according to the list experiment, use non-Islamic financial products or services (whether formal or informal). It is estimated as the coefficient of the *list treatment* dummy variable that takes the value 1 if the respondent is from the treatment group and therefore presented with 5 statements (i.e., the sensitive statement in addition to the 4 non-sensitive statements received by the control group), from a SUR regression. The dependent variable in this SUR regression is the count variable *list experiment*, which shows the number of affirmative responses to the itemized questions in the list experiment. The constant of this regression, which can be interpreted as the mean of the dependent variable for the control group that is presented with 4 non-sensitive statements, is reported below the proportion from the list experiment. Corresponding SUR standard errors and p-values from a two-sided z-test are reported below both proportion estimates. The “Test on equality of SUR coefficients” panel reports a χ^2 statistic and p-value from a test on the equality of the proportions estimated from the direct question and the list experiment using SUR.

labeled “Test on equality of SUR coefficients” reports a χ^2 statistic and p-value from a test on the equality of β_1 and γ_3 for the corresponding column.

Consider the last two columns of Table 4. The estimates of β_1 and γ_3 in Column 2 show that the proportion of respondents reporting using non-Islamic financial products is significantly higher for the sub-sample of those who reside in households with higher income (monthly income of at least 20,000 PKR) versus those with lower income, only according to the list experiment data (13.9 percentage points higher, significant at 1%). In turn, the estimates of β_1 and γ_3 in Column 3 show that the proportion of respondents reporting using non-Islamic financial products is significantly lower for the sub-sample of those who are religiously inclined (i.e., respondents reading the Quran daily) versus those not religiously inclined, only according to the list experiment data (9.1 percentage points lower, significant at 10%). On the other hand, the estimates of β_1 and γ_3 in Column 1 show that the proportion of respondents reporting using non-Islamic financial products is significantly lower for the sub-sample of those who are religiously inclined by a different measure (i.e., respondents always consulting a Sheikh or religious scholar for verifying the validity of financial products) for both the list experiment data and the direct question. These differing patterns between the direct question and the list experiment for different sub-samples implies that direct survey questions cannot accurately reflect the patterns of demand for Islamic (micro-)finance, in addition to not being able to capture its magnitude (as discussed for Table 3).

Thus, if we want to conduct heterogeneity analysis for policy recommendations it appears preferable to use estimates from list experiments. With the preceding results from Tables 3 and 4 in mind, we proceed with heterogeneity analysis of the list experiment data. These are estimated using Nonlinear Least Squares and Maximum Likelihood (NLS and ML, respectively; see (Imai,

2011)), but for ease of interpretation we also present OLS estimates. As the patterns of heterogeneity for the OLS and NLS (and, to a lesser degree, ML) estimates are broadly similar, we interpret the OLS coefficients in the text.¹⁰

Before proceeding to the interpretation of the results, we briefly describe the NLS and ML estimators for the list experiment data. The NLS uses a two-step approach: (1) using data for the control group only, the number of affirmative responses is modeled as a function of the covariates; (2) next, the response to the sensitive item in the treatment group is modeled using the parameter estimates from step (1). The NLS estimator uses a logistic function to restrict predicted values within the unit interval. The ML estimates are based on the Imai (2011) Maximum Likelihood estimator, which estimates parameters for two separate binomial models (the number of affirmative responses to the control list, given the covariates, and the probability of a positive response to the sensitive item, given the number of affirmative responses to the control list and the covariates) using the expectation maximization (EM) algorithm, which treats the response to the sensitive item as partially missing data.

Motivated by the literature (Höglinger et al., 2016), we proceed by conducting a number of heterogeneity analyses with

¹⁰ Note that while Imai’s ML estimator is statistically more efficient than the NLS estimator, a drawback of the ML estimator as compared to the NLS estimator is that the ML estimator is extremely complex and difficult to optimize, occasionally leading to convergence problems and differences in outcomes depending on the particular set of start values used. All ML estimates that are presented in this paper are done with the new STATA package *kict* (see (Tsai, 2019)). The *kict* ML STATA package includes the possibility to report results with ML using a set of optimizations with different randomly-selected initial values (using the `protect()` command). If a large number (say, 50) is set, it is highly likely that a global maximum, rather than a local maximum, results.

Table 4

		Always consult scholar	Majority income household	Quran daily
- Direct survey question		-0.056	0.034	0.001
	SE	(0.026)	(0.022)	(0.021)
	P> z	0.030	0.124	0.961
- List experiment interaction		-0.118	0.139	-0.091
	SE	(0.064)	0.052	0.049
	P> z	0.067	0.007	0.061
- Test on equality of SUR coefficients	χ^2 statistic	0.81	3.51	3.02
	Prob > χ^2	0.368	0.061	0.082
Number of observations		1,753	2,105	2,144

Notes: The table reports the comparison between the results from the direct survey question and the list experiment for various sub-samples using SUR. The “Direct survey question” panel reports the coefficient from a SUR regression on a dummy variable that takes a value of 1 for the sub-sample indicated in the corresponding column, where the dependent variable is a dummy that takes the value 1 if the respondent reports that they are or may become a customer of non-Islamic financial institutions (0 otherwise). The “List experiment interaction” panel reports the coefficient from a SUR regression on the interaction of the *list treatment* dummy variable and a dummy variable that takes a value of 1 for the sub-sample indicated in the corresponding column. The dependent variable in this SUR regression is the count variable *list experiment*, which shows the number of affirmative responses to the itemized questions in the list experiment. Corresponding SUR standard errors and p-values from a two-sided z-test are reported below both proportion estimates. The “Test on equality of SUR coefficients” panel reports a χ^2 statistic and p-value from a test on the equality of the two SUR coefficients reported in the preceding panels of the same column. The sub-samples are defined based on the following three dummy variables. **Always consult scholar:** 1 = respondent always consults a Sheikh (religious scholar) for verifying the validity of financial products, 0 otherwise. **Majority income household:** 1 = respondent’s overall monthly household income is at least 20,000 PKR, 0 otherwise. **Quran daily:** 1 = respondent is reading the Quran daily, 0 otherwise.

respect to these estimates along demographic and other dimensions.¹¹ The purpose is to estimate the proportion of individuals using non-Islamic finance for various sub-groups. In each regression, we are interested in the coefficient on the interaction between an indicator variable for the treatment group and an indicator variable for the relevant dimension of heterogeneity (e.g., gender).

We begin the analysis by focusing on a key demographic variable, gender. As the interaction coefficients in Table 5 show, the propensity to adopt non-Islamic finance is not significantly different for females, compared to males. This is in line with the stylized fact reported in the sub-section on Pakistan, that men and women do not differ statistically along most religious dimensions.¹²

Interestingly, in line with the self-declared importance of religion in Pakistan being roughly the same in different age groups (as noted in the country background information), the proportion of respondents using non-Islamic finance is not significantly different for persons above median age (37 years) compared to those below (results not shown for brevity). Other demographic factors (results not shown for brevity), including household size measured by total number of household members and number of children in the household, do not significantly affect the number of affirmative responses.

Formal education in terms of highest grade of schooling completed does not matter at various cut-offs (Columns 1–4 of Table 6). However, it is important to note that about 48% of our

¹¹ For full transparency we also provide a multivariate analysis (OLS, NLS and ML for the list data as well as Logit for the direct question) in Appendix A. Note, however, that we could only include a subset of the variables used in the heterogeneity analyses presented in the main body of this paper, since several of these variables are subsets of one another by definition.

¹² While we cannot completely rule out that we may be underpowered for picking up a gender effect (see Blair and Imai (2012), for a discussion of large standard errors and associated power issues in list experiments compared to direct questioning), this is unlikely since in our large sample of 2,220 individuals about 13% (i.e., almost 290 individuals) are male. Thus, the difference between males and females, if any, is unlikely to be large. In summary, we have no reason to expect that our main results do not also hold for men.

Table 5

Heterogeneity in sensitive behavior: Gender.

	OLS (1)	NLS (2)
List treatment # Female	0.083 (0.067)	0.373 (0.318)

Notes: The table reports the result of heterogeneity analysis using OLS, Nonlinear Least Squares (Imai, 2011) regressions. For ease of exposition only the relevant interaction coefficients from these regressions are reported (see main text for details). The dependent variable in these regressions is **List experiment:** Number of positive responses from the list experiment. The other variables are defined as **List treatment:** 1 = respondent is in the treatment group and received the sensitive statement, 0 otherwise; and **Female:** 1 = female, 0 = male. Note that the Maximum Likelihood (Imai, 2011) estimator for this specification does not converge. Robust standard errors in parentheses. * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$.

respondents have no formal education, so the median of the school grade completed by the respondents is grade 3. Due to these characteristics, and given the local context, passive literacy measured by reading skills may be a more suitable proxy, in our sample, for being more educated. We find that those who can read Urdu are significantly more likely, by about 7 percentage points, to state that they adopt non-Islamic finance (Columns 5 and 6).

Next, we study the impact of several variables indicating economic status on the number of affirmative responses in the list experiment. Columns 1–4 of Table 7 show that for the top two categories of household income, the affinity towards non-Islamic finance is larger: for the highest income group (above 30,000 Pakistani Rupee or PKR, the local currency, per month) it is about 11 percentage points higher and for the top two groups combined (at or above 20,000 PKR per month) it is about 13 percentage points higher. Note that roughly 33% and 81%, respectively, of our sample lie in these income categories. This finding is in line with the stylized fact from the WVS, noted in our background

Table 6
Heterogeneity in sensitive behavior: Education.

	OLS (1)	NLS (2)	ML (3)	OLS (4)	NLS (5)	ML (6)	OLS (7)	NLS (8)	ML (9)
List Treatment # Educ. Abv. Median	0.001 (0.043)	0.185 (0.182)	1.261 (1.189)						
List Treatment # Formal Educ.				0.059 (0.043)	0.254 (0.183)	-			
List Treatment # Read Urdu							0.072 (0.041)*	0.309 (0.177)*	1.189 (1.185)

Notes: The table reports the result of heterogeneity analysis using OLS, Nonlinear Least Squares and Maximum Likelihood (Imai, 2011) regressions. For ease of exposition only the relevant interaction coefficients from these regressions are reported (see main text for details). The dependent variable in these regressions is **List experiment**: Number of positive responses from the list experiment. The other variables are defined as **List treatment**: 1 = respondent is in the treatment group and received the sensitive statement, 0 otherwise; **Educ. abv. Median**: 1 = respondent falls above the median (grade 3) in terms of the highest grade of schooling completed, 0 otherwise; **Formal educ.**: 1 = respondent's education is larger than 0 in terms of the highest grade of schooling completed, 0 otherwise; and **Read Urdu**: 1 = respondent is able to read Urdu, 0 otherwise. Robust standard errors in parentheses. * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$.

Table 7
Heterogeneity in sensitive behavior: Economic status.

	OLS (1)	NLS (2)	ML (3)	OLS (4)	NLS (5)	ML (6)	OLS (7)	NLS (8)	ML (9)
List Treatment # Richest HH	0.107 (0.044)**	0.451 (0.186)***	2.211 (1.203)*						
List Treatment # Maj. Inc. HH				0.133 (0.051)***	0.609 (0.253)***	-			
List Treatment # HH Own Land							0.067 (0.042)	0.290 (0.183)	-0.218 (1.034)

Notes: The table reports the result of heterogeneity analysis using OLS, Nonlinear Least Squares and Maximum Likelihood (Imai, 2011) regressions. For ease of exposition only the relevant interaction coefficients from these regressions are reported (see main text for details). The dependent variable in these regressions is **List experiment**: Number of positive responses from the list experiment. The other variables are defined as **List treatment**: 1 = respondent is in the treatment group and received the sensitive statement, 0 otherwise; **Richest hh**: 1 = respondent's overall monthly household income lies above 30,000 Pakistani Rupee (PKR), 0 otherwise; **Maj. inc. hh**: 1 = respondent's overall monthly household income is at least 20,000 PKR, 0 otherwise; and **HH own land**: 1 = respondent or the respondent's household owns land, 0 otherwise. Robust standard errors in parentheses. * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$.

information on Pakistan, that the importance of religion in Pakistan is higher in lower income groups, and suggests that the promotion of Islamic finance to individuals with lower incomes might be particularly effective. In contrast, a wealth indicator, such as owning land, does not seem to matter for the proportion that is willing to avail non-Islamic financial products (Columns 5 and 6).

Panel a of Table 8 shows that in terms of religiosity, people who read the Quran daily, at least weekly or at least monthly (compared to those who do not do so at all or do so only during Ramadan) are all less likely (about 11 percentage points on average across these three different groups) to adopt non-Islamic finance. These results are reassuring because they show that the list experiment is successful at capturing components of financial behavior that are correlated with religiosity. Note that the survey-based measures of religiosity that we use for the analysis in Table 8 are less likely to suffer from social desirability bias. Reciting the Quran or attending religious gatherings is not emphasized as mandatory in Islam, in contrast to prescribed observances such as praying five times a day or proscribed practices such as paying or charging/receiving interest. Likewise, seeking guidance from an Islamic scholar is left to personal choice or availability.

More policy relevant, however, are our findings reported in Panel b of Table 8. Columns 1–4 show that among those who

attend religious meetings at least once a week (compared to those who do not attend religious meetings each week) and among those who do so more than once a month (compared to those who do so at most once a month), the proportions of individuals adopting non-Islamic finance are about 11 percentage points and 8 percentage points lower, respectively. This suggests that religious gatherings might be ideal venues for promoting financial inclusion via directly offering Islamic (micro-)finance products or via raising financial literacy/awareness about such products. Another finding that suggests a cost-effective avenue for increasing financial inclusion is highlighted in Columns 5 and 6 of Table 8: among people who always consult religious scholars about financial arrangements, the proportion adopting non-Islamic finance is, as expected, lower (by about 13 percentage points). Thus, if such scholars are informed about formal Islamic financial products available to poorer segments of society on a regular basis, it is conceivable that they can play a vital role in increasing financial inclusion.

5. Conclusion and discussion

We conduct a list experiment as part of a larger survey to study the attitudes of the Muslim poor in Pakistan towards the usage of non-Islamic finance. In this concluding section, we briefly summarize our main results and discuss some (policy) implications of our findings.

Table 8
Heterogeneity in sensitive behavior – Religiosity.

Panel a									
	OLS (1)	NLS (2)	ML (3)	OLS (4)	NLS (5)	ML (6)	OLS (7)	NLS (8)	ML (9)
List Treatment # Quran Daily	-0.100 (0.050)**	-0.447 (0.233)*	1.027 (1.039)						
List Treatment # Quran Daily/Weekly				-0.117 (0.041)***	-0.505 (0.179)***	-0.194 (1.036)			
List Treatment # Quran Daily/Weekly/Monthly							-0.112 (0.044)**	-0.476 (0.184)***	-0.919 (1.047)
Panel b									
	OLS (1)	NLS (2)	ML (3)	OLS (4)	NLS (5)	ML (6)	OLS (7)	NLS (8)	ML (9)
List Treatment # Attend Gathering At Least Once a Week	-0.105 (0.041)**	-0.466 (0.181)***	-0.062 (1.035)						
List Treatment # Attend Gathering More Than Once a Month				-0.079 (0.046)*	-0.333 (0.190)*	-0.043 (1.198)			
List Treatment # Always Consult Scholar							-0.132 (0.065)**	-0.694 (0.396)**	-28.421 (0.538)***

Notes: The two panels of this table report the result of heterogeneity analysis using OLS, Nonlinear Least Squares and Maximum Likelihood (Imai, 2011) regressions. For ease of exposition only the relevant interaction coefficients from these regressions are reported (see main text for details). The dependent variable in these regressions is **List experiment**: Number of positive responses from the list experiment. The other variables are defined as **List treatment**: 1 = respondent is in the treatment group and received the sensitive statement, 0 otherwise; **Quran daily**: 1 = respondent is reading the Quran daily, 0 otherwise; **Quran daily/weekly**: 1 = respondent is reading the Quran daily or weekly, 0 otherwise; **Quran daily/weekly/monthly**: 1 = respondent is reading the Quran daily, weekly or on a monthly basis, 0 otherwise; **Attend gathering at least once a week**: 1 = respondent attends religious gatherings at least once a week, 0 otherwise; **Attend gathering more than once a month**: 1 = respondent attends religious gatherings more than once a month, 0 otherwise; and **Always consult scholar**: 1 = respondent always consults a Sheikh (religious scholar) for verifying the validity of financial products, 0 otherwise. Robust standard errors in parentheses. * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$.

First, we find that about 37% of respondents in our sample report using non-Islamic financial products and services via the list experiment. Thus, the first key takeaway is that a sizeable fraction of our sample of Islamic microfinance borrowers uses conventional finance as well. A policy recommendation emerging from this result is that, from an economic perspective, individuals will benefit if they can access both, Islamic and conventional finance. Furthermore, regulators should take into account the fact that a non-negligible proportion of the poor might be customers of Islamic as well as conventional financial organizations. Take the case of information sharing between financial institutions: In countries where credit bureaus do not collect information on microfinance customers, information sharing between Islamic and conventional microfinance providers should be facilitated due to the likely overlap in their customer base. Overall, our first result speaks to the larger literature in development economics which is concerned with the usage of microfinance, and its impact on financial development and growth (Cull and Morduch, 2017; Morduch, 1999; Kuran, 2018).

Second, compared to the substantial usage of non-Islamic finance uncovered by our list experiment, we find evidence of underreporting of usage of non-Islamic financial products and services in a direct survey measure. Due to the wording of our questions, this estimate of underreporting is expected to be a lower bound for the extent of underreporting. An immediate implication of this finding is that previous studies are likely to have underestimated the usage of non-Islamic finance, due to survey participants exhibiting social desirability bias.

Third, our results reveal ex-ante unpredictable heterogeneity in underreporting across sub-groups of the sample.

Last, and importantly, we find meaningful heterogeneity in the list-experiment-reported usage of non-Islamic financial products across several sub-groups defined by individual characteristics. In particular, we show that the proportion of respondents that use non-Islamic financial products: (i) does not depend on gender or age; (ii) increases in literacy; (iii) is higher for the highest household-income groups; (iv) is lower for those who read the Quran on a regular basis; and (v) is lower among those who frequently attend religious gatherings or consult religious scholars on a regular basis. These results have important implications for financial inclusion efforts aimed at underserved Muslim populations. For example, our finding (iii) should motivate the promotion of Islamic (micro-)finance particularly to people living at the bottom of the income distribution, while our finding (v) suggests that collaborations with religious leaders and organizations might be fruitful.

Our second and third result imply that survey questions, per se, are unsuitable not only for gauging the overall usage of and attitudes towards Shari'a-compliant financial products and services, but also to inform discussions about their cross-sectional patterns. Given recent policy debates and drives to increase financial inclusion, these results suggest that survey researchers in the field should exercise caution in relating questions used to elicit financial behavior to precise quantitative implications and clearly communicate the methodological shortcomings of surveys revealed by our analysis. This is important as costly financial

inclusion strategies are shaped by such data, and rely on them being of high quality to be successful and cost-effective.

We believe that the inclusion of list experiments or alternative, survey-based, techniques that are designed to overcome social desirability bias in global surveys on barriers to financial inclusion as standard practice can mitigate these issues and has considerable advantages over non-survey-based alternatives. As our study implemented the list experiment method, we wish to point out the advantages of this approach in particular over some commonly used non-survey-based alternatives – compiling data on actual usage of Shari'a-compliant and conventional financial products and the implementation of randomized controlled trials (RCTs): While actual usage data can offer important insights, the collection of such data from administrative sources is costly and difficult compared to the list experiment approach due to concerns emerging in the context of personal financial data (e.g., data security and informed consent). These difficulties might lead to smaller data sets as well as possible bias (e.g., caused by selection in who provides consent), which pose significant challenges in the use of administrative data, as they necessitate the highly problematic assumption of external validity.

Contrary to actual usage data, RCTs allow for the precise control of conditions, thus enabling the study of demand for Islamic finance and its properties in great detail. A good example is [El-Zoghbi et al. \(2016\)](#), which uses an RCT in Jordan to study the demand for Shari'a-compliant financial products, as well as its properties such as price elasticity, by marketing different loan types to several thousand individuals. However, external validity is a common concern in the case of RCTs, and the authors stress the need for conducting similar RCTs in other contexts for achieving external validity. Yet, replicating RCTs with different product offerings, different segments of the population and – as the authors suggest – across a multitude of countries, is a very (likely prohibitively) costly approach to studying the issue at hand.

While external validity is also an issue for any particular list experiment, solving the issue simply involves appending a set of questions to an otherwise standard survey (e.g., as part of regularly conducted global surveys).

Among the survey-based alternatives, self-administered questions have been shown to reduce social desirability bias in responses about religion ([Presser and Stinson, 1998](#)). However, they may not be universally suitable in the context of surveys regarding financial inclusion due to the low levels of literacy that characterize financially under-served populations, such as our study population in Pakistan.

An alternative approach that estimates the prevalence of affirmative responses to a sensitive statement within groups, similar to the list experiment, is the randomized response technique (RRT; ([Warner, 1965](#))), which preserves anonymity by introduction of random noise in the responses. While meta-studies ([Lensvelt-Mulders et al., 2005](#)) find that RRTs on average ameliorate social desirability bias, they are known to increase the variance of estimates (as does the list experiment, see below),¹³ invoke lower trust in participants due to their complexity, and are not easy to administer in the field ([Tourangeau, 2018](#)). In addition, their validity has been disputed ([Höglinger et al., 2016](#); [Wolter and Preisendörfer, 2013](#)). Other alternatives in this category (e.g., the “bogus pipeline” method, which uses a fake polygraph test) also suffer from similar drawbacks related to difficulty in invoking trust or prohibitive cost.

The list experiment technique seems like an informative, cost-effective, and internationally scalable alternative among existing

non-survey- and survey-based tools that can address social desirability bias in the context of barriers to financial inclusion for poor Muslim populations worldwide. Yet, list experiments are not exempt from criticism. Besides having been found to produce large standard errors ([Blair and Imai, 2012](#)),¹³ and there being a recent discussion about a possible false-positive bias ([Ahlquist, 2018](#); [Kuhn and Vivyan, 2022](#); [Wolter and Diekmann, 2021](#)), several meta-analyses ([Blair et al., 2020b](#); [Ehler et al., 2021](#); [Li and van den Noortgate, 2022](#)) point to a number of additional concerns, such as the success of a list experiment depending strongly on certain procedural aspects. We therefore acknowledge that while our paper displays one possible remedy – the list experiment technique – determining the most effective and efficient way is beyond the scope of our study, but certainly deserves further attention.

Thus, given: (i) the key role of financial inclusion in high-profile development agendas such as the SDGs; (ii) the importance of a focus on Muslim populations in reaching the goals; (iii) the importance of data on the attitudes towards the usage of (non-) Islamic finance in this regard; and (iv) the shortcomings of direct survey questions for collecting such data demonstrated in this paper, investigating various contenders to scalable solutions to the issue of social desirability bias as it pertains to the demand for non-Islamic (micro-)finance with the goal to implement such approaches in the context of large global surveys is of key importance.

CRediT authorship contribution statement

Syedah Ahmad: Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Project administration. **Robert Lensink:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Supervision. **Annika Mueller:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank participants of the North American ESA meeting, the American Association of Behavioral and Social Sciences Conference, and seminar participants at the Faculty of Economics and Business, University of Groningen, for helpful comments. We are also thankful to Roy Mersland for providing valuable suggestions.

Appendix A

See [Table A.1](#).

Appendix B. Supplementary material

Supplementary material associated with this article can be found in the online version <https://doi.org/10.34894/NXXGPK>.

¹³ Note in this context that the focus of our discussion is on very large, global surveys, where this is less of a concern. Of course, the issue may still be relevant for particular sub-group analyses.

Table A.1
Multivariate analysis.

	OLS	NLS	ML	Logit
Educ. abv. Median	-0.037 (0.072)	-0.120 (0.324)	0.994 (1.693)	-0.242 (0.224)
Read Urdu	0.122 (0.072)*	0.595 (0.334)*	1.040 (1.329)	0.773*** (0.237)
Richest hh	0.068 (0.051)	0.352 (0.230)	2.319 (1.271)*	-0.284* (0.156)
HH own land	0.055 (0.046)	0.288 (0.224)	0.576 (1.379)	0.403*** (0.151)
Quran daily/weekly/monthly	-0.128 (0.050)**	-0.587 (0.233)**	-2.030 (1.078)*	0.101 (0.163)
Always consult scholar	-0.136 (0.062)**	-0.648 (0.352)*	-30.425 (1.738)***	-0.582** (0.258)

Notes: The table reports the result of heterogeneity analysis using OLS, Nonlinear Least Squares and Maximum Likelihood (Imai, 2011) regressions for the list data as well as heterogeneity analysis using Logit for the direct question. All estimates are done with the STATA package kict (see Tsai (2019)). The dependent variable for the first three columns is **List experiment**: Number of positive responses from the list experiment. The dependent variable for the fourth column is **Client of non-Islamic financial institution**: 1 = respondent reports that they are or may become a customer of non-Islamic financial institutions, 0 otherwise. In case of the first three columns, the coefficients on the interactions of several variables used for heterogeneity analysis with **List treatment**: 1 = respondent is in the treatment group and received the sensitive statement, 0 otherwise, are reported. These variables are: **Educ. abv. Median**: 1 = respondent falls above the median (grade 3) in terms of the highest grade of schooling completed, 0 otherwise. **Read Urdu**: 1 = respondent is able to read Urdu, 0 otherwise. **Richest hh**: 1 = respondent's overall monthly household income lies above 30,000 Pakistani Rupee (PKR), 0 otherwise. **HH own land**: 1 = respondent or the respondent's household owns land, 0 otherwise. **Quran daily/weekly/monthly**: 1 = respondent is reading the Quran daily, weekly or on a monthly basis, 0 otherwise. **Always consult scholar**: 1 = respondent always consults a Sheikh (religious scholar) for verifying the validity of financial products, 0 otherwise. The fourth column presents the coefficients of these variables. Robust standard errors in parentheses. * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$.

References

- Ahart, A.M., Sackett, P.R., 2004. A new method of examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique. *Organ. Res. Methods* 7 (1), 101–114.
- Ahluquist, J.S., 2018. List experiment design, non-strategic respondent error, and item count technique estimators. *Political Anal.* 26 (1), 34–53.
- Ahmad, S., Lensink, R., Mueller, A., 2020. Uptake, use, and impact of Islamic savings: Evidence from a field experiment in Pakistan. In: SSRN Working Paper No. 3743428.
- Benson, B.V., Merolla, J.L., Geer, J.G., 2011. Two steps forward, one step back? Bias in the 2008 presidential election. *Elect. Stud.* 30 (4), 607–620.
- Biemer, P.P., Jordan, B.K., Hubbard, M., Wright, D., 2005. A test of the item count methodology for estimating cocaine use prevalence. In: Kennet, J., Gfroerer, J. (Eds.), *Evaluating and Improving Methods Used in the National Survey on Drug Use and Health, Vol. Methodology Series. Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Rockville, MD*, pp. 149–174.
- Blair, G., Coppock, A., Moor, M., 2020b. When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments. *Am. Political Sci. Rev.* 114 (4), 1297–1315.
- Blair, G., Imai, K., 2012. Statistical analysis of list experiments. *Political Anal.* 20, 47–77.
- Blattman, C., Jamison, J., Koroknay-Palicz, T., Rodrigues, K., Sheridan, M., 2016. Measuring the measurement error: A method to qualitatively validate survey data. *J. Dev. Econ.* 120, 99–112.
- Bullock, W., Imai, K., Shapiro, J.N., 2011. Statistical analysis of endorsement experiments: Measuring support for militant groups in Pakistan. *Political Anal.* 19 (4), 363–384.
- Cull, R., Morduch, J., 2017. Microfinance and economic development. In: Beck, T., Levine, R. (Eds.), *Handbook of Finance and Development*.
- Demirgüç-Kunt, A., Klapper, L., Randall, D., 2014. Islamic finance and financial inclusion: measuring use of and demand for formal financial services among Muslim adults. *Rev. Middle East Econ. Financ.* 17, 7–218.
- Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S., Hess, J., 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. World Bank Group.
- Ehler, I., Wolter, F., Junkermann, J., 2021. Sensitive questions in surveys: A comprehensive meta-analysis of experimental survey studies on the performance of the item count technique. *Public Opin. Q.* 85 (1), 6–27.
- El-Gamal, M., El-Komi, M., Karlan, D., Osman, A., 2014. Bank-insured RoSCA for microfinance: Experimental evidence in poor Egyptian villages. *J. Econ. Behav. Organ.* S56–S73.
- El-Zoghbi, M., 2013. *Islamic Microfinance and Clients: See What I Do, Not What I Say*. CGAP, Washington, D.C.
- El-Zoghbi, M., Karlan, D., Osman, A., Shammout, N., 2016. *Understanding Demand for Sharia-Compliant Loans: Results of a Randomized Experiment in Jordan*. World Bank, Washington, D.C.
- Eriksen, S., Lensink, R., 2015. Measuring the impact of an ongoing microcredit project: evidence from a study in Ghana. *J. Dev. Eff.* 7 (4), 519–529.
- Glynn, A., 2013. What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opin. Q.* 77, 159–172.
- Höglinger, M., Ben, J., Diekmann, A., 2016. Sensitive questions in online surveys: An experimental evaluation of different implementations of the randomized response technique and the crosswise model. *Surv. Res. Methods* 10 (3), 171–187.
- Imai, K., 2011. Multivariate regression analysis for the item count technique. *J. Amer. Statist. Assoc.* 106 (494), 407–416.
- Janus, A.L., 2010. The influence of social desirability pressures on expressed immigration attitudes. *Soc. Sci. Q.* 91 (4), 928–946.
- Kane, J.G., Craig, S.C., Wald, K.D., 2004. Religion and presidential politics in Florida: A list experiment. *Soc. Sci. Q.* 85, 281–293.
- Karim, N., Tarazi, M., Reille, X., 2008. *Islamic Microfinance: An Emerging Market Niche*. CGAP, Washington DC.
- Karlan, D., Zinman, J., 2012. List randomization for sensitive behavior: An application for measuring use of loan proceeds. *J. Dev. Econ.* 98 (1), 71–75.
- Kuhn, P., Vivyan, N., 2022. The misreporting trade-off between list experiments and direct questions in practice: Partition validation evidence from two countries. *Political Anal.* 30 (3), 381–402.
- Kuran, T., 2018. Islam and economic performance: Historical and contemporary links. *J. Econ. Lit.* 56 (4), 1292–1359.
- Kustin, B., 2015. *Islamic (Micro) Finance: Culture, Context, Promise, Challenges*. Bill & Melinda Gates Foundation.
- Lensvelt-Mulders, G., Hox, J., van der Heijden, P., 2005. Meta-analysis of randomized response research: Thirty-five years of validation. *Sociol. Methods Res.* 33 (3), 319–348.
- Li, J., van den Noortgate, W., 2022. A meta-analysis of the relative effectiveness of the item count technique compared to direct questioning. *Sociol. Methods Res.* 51 (2), 760–799.
- Makhlouf, M., 2017. *Islamic Banking Opportunities Across Small and Medium Enterprises in MENA*. World Bank.
- Morduch, J., 1999. The microfinance promise. *J. Econ. Lit.* XXXVII, 1569–1614.
- Pakistan Bureau of Statistics, 2020. Retrieved from <http://pbs.gov.pk/>.
- Pakistan Microfinance Network, 2020. *Pakistan microfinance review 2019*. Retrieved from Pakistan Microfinance Network: <https://pmn.org.pk/wp-content/uploads/2020/11/PAKISTAN-MICROFINANCE-review-2019-10-12-Updated-Text-Oct-29.pdf>.
- Paulhus, D.L., 1986. 'Self-deception and impression management in test responses. In: Angleimer, A., Wiggins, J.S. (Eds.), *Personality Assessment Via Questionnaire: Current Issues in Theory and Measurement*. Springer-Verlag, Berlin.
- Pew Research Center, 2016. *Religion and gender*. Pew research center. Retrieved from <https://www.pewresearch.org/wp-content/uploads/sites/7/2016/03/Religion-and-Gender-APPENDIX-B.pdf>.
- Preisendorfer, P., Wolter, F., 2014. Who is telling the truth? A validation study on determinants of response behavior in surveys. *Public Opin. Q.* 78 (1), 126–146.
- Presser, S., Stinson, L., 1998. Data collection mode and social desirability bias in self-reported religious attendance. *Am. Sociol. Rev.* 63, 137–145.

- Rasmussen, S., 2018. Pakistan Enigma: Why Is Financial Inclusion Happening So Slowly? CGAP Blog.
- Redlawsk, D.P., Tolbert, C.J., Franko, W., 2010. Voters, emotions, and race in 2008: Obama as the First Black President. *Political Res. Q.* 63, 875–889.
- Sudman, S., Bradburn, N.M., 1974. *Response Effects in Surveys: A Review and Synthesis*. Aldine Publishing Company, Chicago.
- Tourangeau, R., 2018. Maintaining respondent trust and protecting their data. In: Vannette, D., Krosnick, J. (Eds.), *The Palgrave Handbook of Survey Research*. Palgrave Macmillan, Cham, pp. 135–141.
- Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. *Psychol. Bull.* 133 (5), 859–883.
- Tsai, C.L., 2019. Statistical analysis of the item-count technique using Stata. *Stata J.* 19 (2), 390–434.
- van der Heijden, P.G., van Gils, G., Bouts, J., Hox, J., 2000. A comparison of randomized response, CASAQ, and direct questioning: eliciting sensitive information in the context of social security fraud. *Sociol. Methods Res.* 32 (4), 384–410.
- Warner, S.L., 1965. Randomised response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* 60 (309), 63–69.
- Wolter, F., Diekmann, A., 2021. False positives and the more-is-better assumption in sensitive question research : New evidence on the crosswise model and the item count technique. *Public Opin. Q.* 85 (3), 836–863.
- Wolter, F., Preisdörfer, P., 2013. Asking sensitive questions: An evaluation of the randomized response technique versus direct questioning using individual validation data. *Sociol. Methods Res.* 42 (3), 321–353.
- Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, p. 2010.
- World Bank, 2018. Financial inclusion. In: *Financial Inclusion Is a Key Enabler to Reducing Poverty and Boosting Prosperity*. Retrieved from The World Bank.
- World Bank, 2020. World development indicators. Retrieved from <https://databank.worldbank.org/reports.aspx?source=2&country=PAK>.
- World Bank, Islamic Development Bank Group, 2017. *Global Report on Islamic Finance : Islamic Finance - A Catalyst for Shared Prosperity?* World Bank, Washington, DC.
- World Values Survey, 2020. Retrieved from <http://www.worldvaluessurvey.org/WVSONline.jsp>.