Consumer Preferences, Bayesian Statistics, and Contract Design: Three Essays in Applied Microeconomics



Nikolai Schaffner
Lincoln College
University of Oxford

July 2025

Submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Economics
Estimated word count: 49,300

Contents

Acknowledgements **5**

Introductory Remarks **6**

I Consumer Preference Elicitation in the Field: A Bundle-Comparison Method **7**

1 Introduction **8**

2 Data **12**
2.1 Population and Context 12
2.2 Survey Design and Treatment Assignment 13

3 Empirical Results **20**
3.1 Distribution of price estimates 20
3.2 Behavioural effects 23
3.3 Regression analysis 25
3.3.1 Relative popularity of goods 25
3.3.2 Treatment impact 32
3.3.3 Interactions between goods 40
3.4 Robustness and rationality 44

4 Conclusion **48**

Appendices **53**

A Appendix – Auxiliary Results **53**

II Utilising Gaussian Processes for Consumer Preference Elicitation and Modelling **61**

1 Introduction **62**

2 Methodology **69**
2.1 Problem statement 69
2.1.1 Estimation target 69
2.1.2 Setting example parameters 70
2.2 Simulated Maximum Likelihood Estimation Approach 75
2.2.1 Correct specification 75
2.2.2 Incorrect specification 80
2.3 Gaussian Process Estimation Approach 84
2.3.1 Introduction to Gaussian Processes 84
2.3.2 Applying Gaussian Processes to Bundle Choice Data 87

3 Empirical Results **95**
3.1 Estimating policy bundles from empirical data 95
3.2 Estimation outcomes 97

4 Conclusion **106**

III Heterogeneous Agent Matching and Microfinance Contract Structure in Bangladesh **111**

1 Introduction **112**

2 Data **118**
2.1 Experimental Design: Principles 118
2.2 Study Sample 127

3 Results **132**
3.1 Ranking behaviour 132
3.2 Participant outcomes 140

4 Structural Estimation **148**
4.1 Modelling Framework 148
4.2 Empirical Estimation (Maximum Simulated Likelihood) 153
4.3 Estimation Results 155

5 Conclusion **160**

Appendices **165**

A Appendix – Auxiliary Results **165**

B Appendix – Question Pool **172**
B.1 Graphic-Design Questions 172
B.2 English Questions 172
B.3 Screenshots of Main Game Stages 173

C Appendix – Communication to Participants **174**
C.1 Recruitment message 174
C.2 Elicitation of consent 175

Concluding Remarks **176**

Acknowledgements

This thesis would not have been possible without the significant support of many people and institutions. First and foremost, I want to express my sincere gratitude to my advisor, Dr Simon Quinn, for his tireless and inspiring guidance. I thank the field teams of *MOMODa Foundation* in Dhaka and Gaibandha as well as *BRACU* in Dhaka – especially Dr Wahedur Rahman – for the excellent support with surveying and experiment execution. I owe gratitude to the *Centre for Studies of African Economies (CSAE)* as well as Lincoln College for funding my fieldwork through multiple grants. Many thanks also to Dr Sanjay Jain for continuous empowerment, to Dr Muhammad Meki for performing a highly constructive Confirmation of Status and to my two DPhil examiners for their rigorous review of my work. Last but not least, I want to thank my partner Rebecca for always being by my side.

Introductory Remarks

In this thesis, I combine applied microeconomics methodology with two novel datasets from Bangladesh which I collected with partners in the field. The first dataset (collected in 2022) is focused on the elicitation of consumer preferences, whereas the second dataset (collected in 2019) resulted from an incentivised lab-in-the-field game focused on microfinance contract structure. I compile the resulting analysis in three self-contained chapters which make up the remainder of this thesis. In the first chapter, I subject the preference elicitation dataset to reduced-form analysis to validate surveying based on binary bundle comparisons. I furthermore provide insights on the uncovered preference relations of survey participants. In the second chapter, I move on to analyse that same dataset using Gaussian Processes – a tool from Bayesian statistics that enables a highly flexible, non-parametric method to capture complex empirical patterns. I illustrate the improvement of using Gaussian Processes compared to conventional methods and include additional insights into the preferences of survey participants. Finally, in the third chapter, I analyse my second dataset, building on the emerging literature on microequity contracts. I show how such microequity contracts could enable partner-matching in the context of a developing country and thereby provide a path towards long-term growth.

Chapter I
Consumer Preference Elicitation in the Field: A Bundle-Comparison Method

Abstract

In this chapter, I execute and evaluate a novel elicitation technique for consumer preferences, using data from a survey I conducted in Bangladesh. Prompting binary comparisons of two bundles of various types of goods at a time, I illustrate that the preferences over 51 different goods expressed by a population of microfinance recipients can be approximated with the output of a low number of salient survey prompts. Processing the resulting rich dataset, I present easily applicable reduced-form analysis that verifies the empirical viability of bundle comparison surveys for preference elicitation and I provide actionable guidance on how to achieve optimal results through such surveys.

1 Introduction

In this chapter, I propose a hitherto underutilised technique for consumer preference elicitation over a large set of goods that is efficient both with respect to time and cost expended by surveyors and participants: preference elicitation through surveys consisting of binary bundle comparisons. That is, presenting participants with two bundles at a time, each consisting of various types of goods and recording participants' expressed preference between the two bundles. Understanding patterns of consumer preferences is a pre-condition to solving many of the most important ongoing points of interest in the wider applied microeconomics literature – entrepreneurial mavericks versus mass participation, cash versus capital goods, rational versus irrational. A significant barrier in doing so is a very practical one: it is challenging to rigorously elicit consumer preferences. Given the importance of preference data for empirical development economics, there is a narrow range of methods available to elicit such data. If market data is available, revealed preference methods are a well-documented pathway for estimation (Varian (2014)). However, in the context of many situations of interest to applied microeconomics, the availability of reliable market data is a rare occurrence. Thus, economists must frequently rely on the study of stated preference data. The principal methodological concept for doing so are choice experiments: participants are presented with a set of hypothetical alternatives, each with varying attributes. They are asked to choose between these alternatives, and the trade-offs between the attributes allow researchers to estimate how much consumers may be willing to pay for specific goods and their attributes. Choice experiments have enjoyed particular popularity in fields such as development economics, environmental economics, health economics, and public policy (Carlsson, 2010, Loomis (2011), Hensher and Johnson (2018)). The preference-elicitation surveying method based

on bundle comparisons, which I propose in this chapter, fills this methodological gap at the forefront not only of microfinance but also other applied microeconomics fields. I evaluate novel data from a survey of Bangladeshi microfinance recipients I designed and conducted with the support of BRAC – a large Bangladesh-based NGO with a self-stated aid budget of USD 131.8M in 2021, of which USD 86.7M were dedicated to microfinance programmes (BRAC (2021)). I show the suitability of bundle comparison surveys as a means of measuring complex consumer preferences in the field. Further, I examine the method's robustness to some commonly observed behavioural biases and I provide insightful patterns among consumption preferences of microfinance recipients easily accessible through common reduced-form analysis. The development economics literature, in particular, has raised a number of key areas of study that benefit from a robust understanding of consumer preferences. For example, Angelucci et al. (2018) use preference data in the context of conditional cash transfers to analyse how recipient preferences for specific elements of a policy (such as conditionality) affect household welfare, demonstrating the importance of having a robust idea of participants' preferences before designing interventions. Chetty et al. (2014) use preference analysis to examine how low-income individuals value educational opportunities, providing valuable insights for policy design that aims to enhance human capital development. Bordering on the institutions literature, Khan et al. (2022) further show the need to measure individual preferences to design mechanisms geared to public goods creation. The disaster relief literature constitutes another example of the methodological need for capturing consumer preferences in a setting with excess demand, but no price mechanism in position to convey this demand to potential suppliers (market or non-market) (Strömberg (2007)). While the general lack of supply of goods is an obvious cause of welfare loss in such a setting, this is then compounded by a potential welfare loss from supply-side and demand-

side agents not being able to coordinate. As a primary focal point of research during the last three decades, the microfinance literature has also uncovered a growing set of use cases for better data on individual preferences. The main channel through which microfinance is posited to create impact is in making capital markets accessible to economic agents previously out of reach from interacting with them. Thus, microfinance interventions can enable consumption smoothing and additional investment in business activities, leading to welfare improvements (Armendariz and Labie (2011)). However, with the field maturing, evidence has emerged suggesting that the effects of microfinance interventions are less universally applicable than initially believed – with particularly the long-term effectiveness being called into question (Roodman (2012), Banerjee et al. (2015)). This has resulted in a shift of focus on heterogeneous treatment effects in microfinance – an extensive new literature has spawned numerous new insights illustrating a rich set of treatment effect distinctions for various sub-populations (Crépon et al. (2015), Meager (2019), Bari et al. (2024), Cueva et al. (2024), Meki and Quinn (2024b)). A strand that has gathered particular momentum (De Mel et al. (2008), Karlan and Zinman (2010), Quinn and Woodruff (2019)) shows that microfinance loans are particularly effective for sub-populations of borrowers with a certain set of expressed preferences or abilities which are closely linked to entrepreneurial activity. Karlan et al. (2014) demonstrate that the presence of microfinance recipients' entrepreneurial proclivities means that providing capital goods, such as agricultural inputs, led to significant positive long-term effects. Similarly, the study of trade-offs between cash transfers and in-kind support (as discussed by Fafchamps et al. (2014), Cunha et al. (2019), Jeong and Trako (2022)) highlights a profound need for understanding individual preferences and characteristics when designing microfinance interventions: cash transfers, which provide immediate liquidity, offer flexibility to allocate resources dynamically and with limited friction.

However, behavioural effects may also divert this flexibility, leading to non-rational financial decisions. Banerjee et al. (2015), for example, found that while cash transfers increased consumption among recipients, these did not typically translate into long-term improvements of financial conditions, suggesting that cash transfers may be susceptible to cognitive biases or external financial pressures (and certainly highlighting that the role of cash ought to be paid special attention to in microfinance intervention design). The behavioural literature has observed a range of bias patterns which have subsequently been linked to microfinance decision-making. For example, the concept of "mental accounting", according to which individuals are more likely to allocate resources to specific goods after having earmarked funds for that respective purpose is particularly influential (Thaler (1985)). Moreover, "endowment effects" also feature prominently in the microfinance literature – according to which individuals place a higher value on goods they already possess, making them less likely to sell or misuse capital goods once acquired (Kahneman et al. (1991)). Mullainathan and Shafir (2013) observe that high-stress situations induce participants to place increased focus on the short-term future – a phenomenon they term "tunnelling". Analysing the effects of participant attention – or lack thereof – on decision-making in a microfinance context, Karlan et al. (2016), show the importance of a pro-active consideration of attention effects when designing interventions. Not just in the context of these biases but many others as well: the role of researchers having a detailed understanding of individual preferences in order to counter-act behavioural biases becomes ever clearer.

This chapter is structured as follows: section 2 describes both the population from which the data in this paper was elicited as well as the design of this elicitation. Section 3 features an analysis of that data and section 4 summarises all insights gained towards the key research questions described above.

2 Data

2.1 Population and Context

A crucial requirement of the experiment evaluated in this chapter was the involvement of a relevant participant population. A proof-of-concept lacking this component would struggle to articulate its external validity. I was fortunate to be able to count on the support of researchers at BRACU[1] to integrate a module containing the bundle-comparison elicitation technique I designed within a large-scale research programme conducted by BRACU, which is otherwise not directly related to the experiment evaluated in this chapter. This research programme – termed *WEE-DiFine* – seeks to shed light on links between women's economic empowerment and digital financial services. Integral to many other components and initiatives is a longitudinal survey – the *frame survey* – (rolled out in numerous rounds between 2020 and 2024) of households participating in various other microfinance projects by BRAC BIGD (2020). The data underlying to this chapter was exclusively collected as part of one such round of that surveying panel – specifically a follow up visit in March 2022. All questions evaluated in this chapter were shown in a separate designated section of the survey and featured no interaction with other elements of the follow-up survey.[2] The respondent population of WEE-DiFine is a subset of individuals enrolled in one of BRAC's microfinance programmes across Bangladesh. The operative set up of BRAC's microfinance programme is built on more than 2,000 branch offices, which interact with Village Organisations (VO) in their respective catchment area (approx-

[1] A special thanks to Dr Wahedur Rahman (BRACU and Oxford Martin School) is due for his support in facilitating the execution of this study. BRACU is an institute embedded in BRAC with the aim to promote leading academic research concerned with microfinance.

[2] Please note that there is certain overlap between the section below and an entry to the AEA's RCT registry: https://www.socialscienceregistry.org/trials/9885

imately five to six kilometres). According to the basic microfinance model operated by BRAC, every VO consists of 15-25 women and it constitutes the counterparty for microfinance recipients' repayment deposits as well as further savings. The typically women-led VOs also function as a hub of information and host events designed to increase financial literacy. As of late 2022, BRAC reported a total of 283,955 VOs serving 5.4 million members across Bangladesh. As part of the WEE-DiFine frame survey, BRAC's microfinance programme representatives selected a random subset of 250 branches (from 30 districts). The microfinance clients of these branches subsequently form the population for the various WEE-DiFine interventions (not linked to this chapter in any further way). On average, VOs consist of 24 loan recipients each. From those 250 branches, BRAC randomly selected a subset of 50 branches for inclusion in the experiment underlying to this chapter. From each of those 50 branches, two VOs were randomly selected by BRAC to generate the data used in this chapter. The households making up the survey population were visited by enumerators in a randomised order. If no member of any given household could be located (or no member consented to being surveyed), then enumerators would move on to the next household associated with that particular VO.[3] This randomised rostering technique, therefore, yields a dataset consisting of survey responses sourced from exactly 500 participating households associated with 100 different VOs. Surveying took place between the 16th of March 2022 and the 2nd of April 2022.

2.2 Survey Design and Treatment Assignment

The purpose of the survey described in the section above is to provide a testing platform for preference elicitation through binary bundle preferences. Individual

[3]Recall that enumerator gender was randomly assigned to each participating household with each enumerator completing an average of 2.5 questionnaires per day, on average.

bundle comparisons, therefore, form the core building block of the deployed questionnaire. To provide additional insight on underlying mechanisms driving participants' responses, four types of slight variations were introduced when assembling questions into questionnaires – reflecting four different treatments that allow both within-individual and between-individual variation:

1. RED-DEC treatment: 15 questions of type (1 good, 1 good), then 15 questions of type (1 good, cash sum) – all drawn from 25-good universe.

2. RED-INC treatment: 15 questions of type (1 good, cash sum), then 15 questions of type (1 good, 1 good) – all drawn from 25-good universe.

3. FULL-DEC treatment: 15 questions of type (1 good, 1 good), then 15 questions of type (1 good, cash sum) – all drawn from 51-good universe.

4. FULL-INC treatment: 15 questions of type (1 good, cash sum), then 15 questions of type (1 good, 1 good) – all drawn from 51-good universe.

Questionnaires were prepared as pre-calculated and pre-assembled sets of bundle comparison questions paths – please see Figure 1 for a visual overview. In a first step, I generated a total of 1,440 fixed bundle comparisons (each consisting of two bundles each) by allocating the budget constraint of BDT 50,000 to integer amounts of randomly chosen types of goods according to the constraints imposed by the treatment design described above. In preparation of what would become the four treatment arms, bundle comparisons were generated in one of four marginally different manners. One half of the bundles were drawn from a universe of 51 types of goods[4] (the

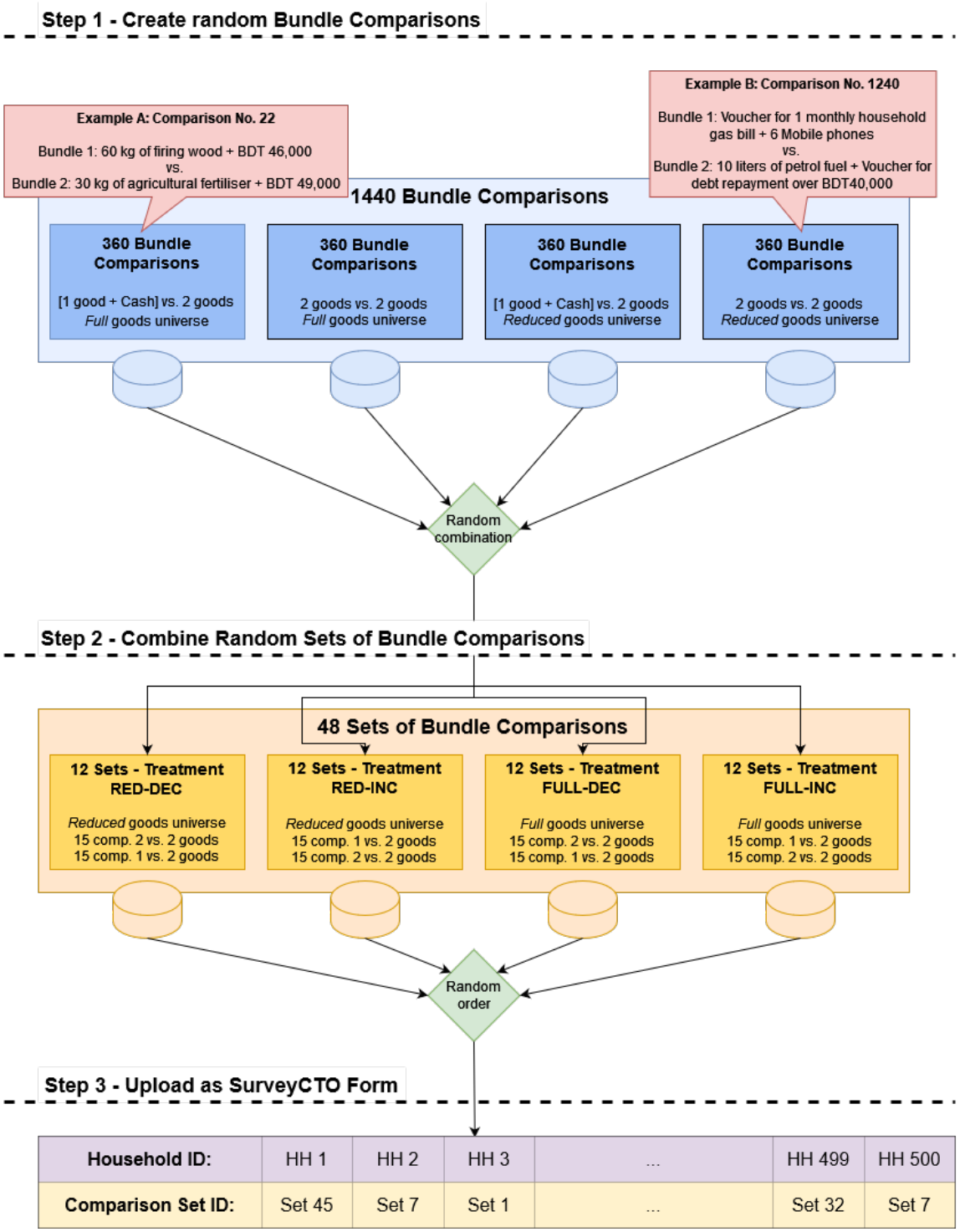
[4]The 51 types of goods – and the price per unit that was assumed for the generation of these bundles – were received by BRACU. They correspond to the 51 most commonly stated goods purchasing intentions for microfinance participants.

full universe), while the other half of the bundles was drawn from a reduced universe of a randomly selected, fixed subset of 25 types of goods. Intersected with the goods-universe distinction described above, a second split into randomly allocated halves was performed. One half of comparisons was constructed as the combination of two types of goods being featured in each presented bundle – that is, each bundle was of the shape *(1 good, 1 good)*. Meanwhile, the other half of comparisons was designed as the combination of one type of a good as well as a cash sum – that is, each bundle was of the shape *(1 good, cash sum)*. With this distinction, I identify effects stemming from a different composition of bundles – and particularly from a more prominent role of cash.[5]

[5]The following rules were applied in the generation of all bundles:

1. The market value of all bundles must be in a range of 90-100% of budget (which was BDT 50,000 – a typical amount of a microfinance loan provided by BRAC).
2. Within each comparison (consisting of two bundles), no bundle can be strictly dominated by the other bundle.
3. At most six units of each good may be part of each bundle.

Figure 1: Description of survey design and implementation in the field



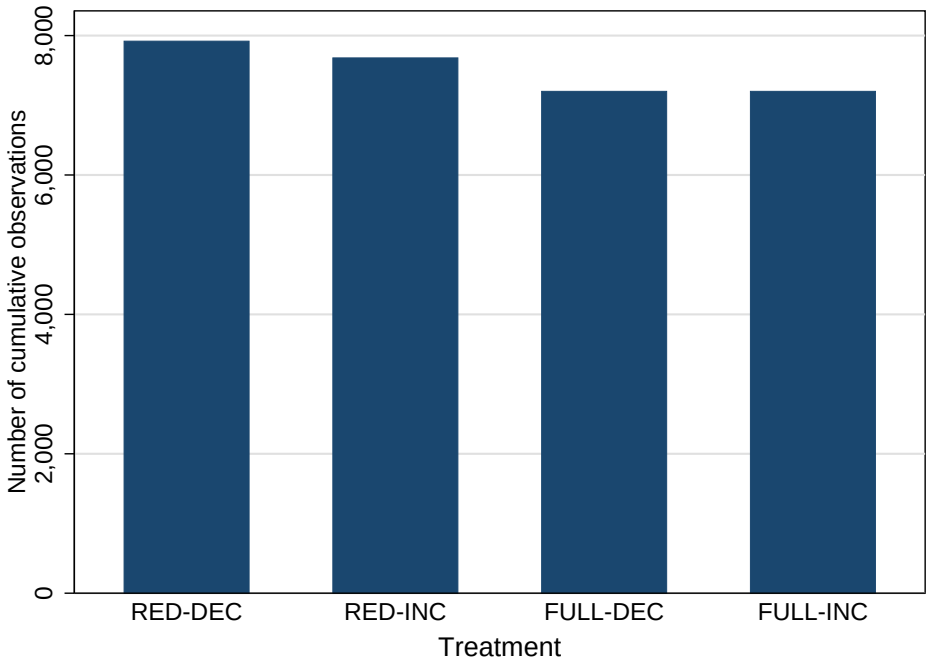
Household ID:	HH 1	HH 2	HH 3	...	HH 499	HH 500
Comparison Set ID:	Set 45	Set 7	Set 1	...	Set 32	Set 7

The last question of each 15-good block contains a choice that is almost identical to the first question featured in that block, but a) with an opposite order of bundles presented within the comparison; and b) one randomly selected type of good missing from the respective bundle. This allows an assessment of the degree to which decisions were made guided by rational thought. The resulting allocation of observations across these four treatments is illustrated in Figure 2. Note that the surveying method did not result in the same amount of bundle-level observations for each treatment: households were continuously surveyed (using a randomly allocated set of comparisons) until the surveying budget was reached, unless the household refused participation or was otherwise not available for surveying.[6] In total, 48 such sets of 30 comparisons each, which consist of two bundles each, were prepared. Survey participants were allocated randomly to exactly one such set. The content and order of questions was constant within each set, so was the order of bundles within comparisons as well as order of goods within bundle. The purpose of having pre-randomised question sets was to a) facilitate a replicable technical implementation in a setting without constant internet access; and b) to ensure a balanced menu of types of goods variance across treatments. After having asked 30 bundle comparison questions, participants were asked to provide their estimate on a random selection of four types of goods. This allowed an ex-post validation of the accuracy of price estimates used to construct bundles. The survey was implemented through *SurveyCTO* and all questions were translated into Bengali by local BRACU personnel before survey deployment. Covariate balance – shown in Table 1[7] – indicates that treatment randomisation was successful.

[6]Table 6 in Appendix A shows an overview of the 51 types of goods used for bundle construction together with their respective spread throughout bundle-level observations in all four treatments.Please note that two goods – *Downpayment for a CNG vehicle* and *Dowry voucher* – were not included in any bundles despite being included in the original universe of goods. The reason for this is that the typical price assumed for these goods was higher than the allocated budget for each bundle.

[7]As well as Table 20 in Appendix A

Figure 2: Allocation of observations



Note: Bars show the allocation of bundle-level observations across treatments.

Table 1: Balance of covariates (Adjusted)

Group	All	(1) RED-DEC	(2) RED-INC	(3) FULL-DEC	(4) FULL-INC
No. of adults in HH	2.94	2.81	3.05	2.86	3.04
	[3]	[3]	[3]	[3]	[3]
	(1.16)	(1.17)	(1.22)	(1.06)	(1.16)
No. of children in HH	1.62	1.68	1.55	1.5	1.76
	[2]	[2]	[1]	[2]	[2]
	(1.12)	(1.14)	(1.07)	(1.08)	(1.17)
No. of females in HH	2.21	2.18	2.26	2.08	2.34
	[2]	[2]	[2]	[2]	[2]
	(1.07)	(1.05)	(1.07)	(1.01)	(1.15)
No. of pensioners in HH	.23	.2	.23	.25	.24
	[0]	[0]	[0]	[0]	[0]
	(.49)	(.44)	(.49)	(.54)	(.48)
Avg. age in HH	29.11	28.01	29.5	30.77	28.26
	[28]	[27]	[29.37]	[28.27]	[27.75]
	(9.11)	(9)	(8.44)	(10.48)	(8.26)
Is HH head female?	.17	.19	.17	.17	.17
	[0]	[0]	[0]	[0]	[0]
	(.38)	(.39)	(.38)	(.37)	(.37)
Is HH head married?	.9	.92	.89	.91	.89
	[1]	[1]	[1]	[1]	[1]
	(.3)	(.28)	(.31)	(.29)	(.31)
HH head yrs. of edu.	6.38	6.08	7.03	6.15	6.29
	[5]	[5]	[7]	[5]	[5]
	(5.05)	(3.28)	(5.60)	(5.73)	(5.51)
Top edu. in HH (in yrs.)	10.75	11.38	11.83	9.51	10.17
	[9]	[9.5]	[10]	[9]	[9]
	(9.6)	(10.95)	(11.37)	(7.18)	(7.86)
Observations	500	132	128	120	120

P-value testing	(1)=(2)	(1)=(3)	(1)=(4)	(2)=(3)	(2)=(4)	(3)=(4)	(1)=(2)=(3)=(4)
No. of adults in HH	.123	.757	.106	.143	.933	.284	.221
No. of children in HH	.327	.121	.575	.716	.066	.088	.176
No. of females in HH	.506	.524	.314	.132	.502	.12	.318
No. of pensioners in HH	.596	.357	.509	.832	.925	.918	.765
Avg. age in HH	.211	.025	.795	.331	.28	.066	.13
Is HH head female?	.669	.635	.675	.898	.924	1	.96
Is HH head married?	.478	.749	.509	.597	.982	.63	.837
HH head yrs. of edu.	.124	.91	.73	.209	.394	.859	.365
Top edu. in HH (in yrs.)	.763	.267	.429	.054	.166	.476	.273

Note: Main figure shown for each variable is sample mean; sample median is in square brackets and sample standard error is in parentheses. P-values in the lower panel refer to the confidence in the null-hypothesis of equal distributions between sub-samples (standard errors are clustered at the level of Districts). For the joint tests in the lower panel, significance was calculated using a Bonferroni correction, accounting for the simultaneity of 9 or 27 (in the last column) hypotheses being tested, respectively. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

3 Empirical Results

In the following section, I present the analysis of the data collected in the survey described in the previous section. Before moving to regression-based evaluation of demand patterns and treatment effects, I examine the validity of key assumptions involved in the design of this experiment.

3.1 Distribution of price estimates

Recall that every participant was asked to respond to four randomly selected price-estimation questions (just after the bundle comparison questions). These questions each contained a random amount of a randomly selected good type from the respective goods universe and asked the participant to provide their estimate of how much that particular quantity of that particular good would cost them if they were to procure the specified unit of it. Participants were sourced from 26 districts across Bangladesh, covering both rural and urban populations as well as a high geographical diversity. Therefore, it is very likely that true market prices faced by participants featured significant variance throughout the sample of survey participants, making it probable to measure a certain price variation in participants' estimates. Figure 3 provides an overview of the distribution of participants' price estimates as well as the price that was used in the creation of the survey questions.[8] While there is a moderate spread of price estimates concerning most goods, there appears to be a well-defined consensus area for most goods (represented by the interquartile range in Figure 3).[9]

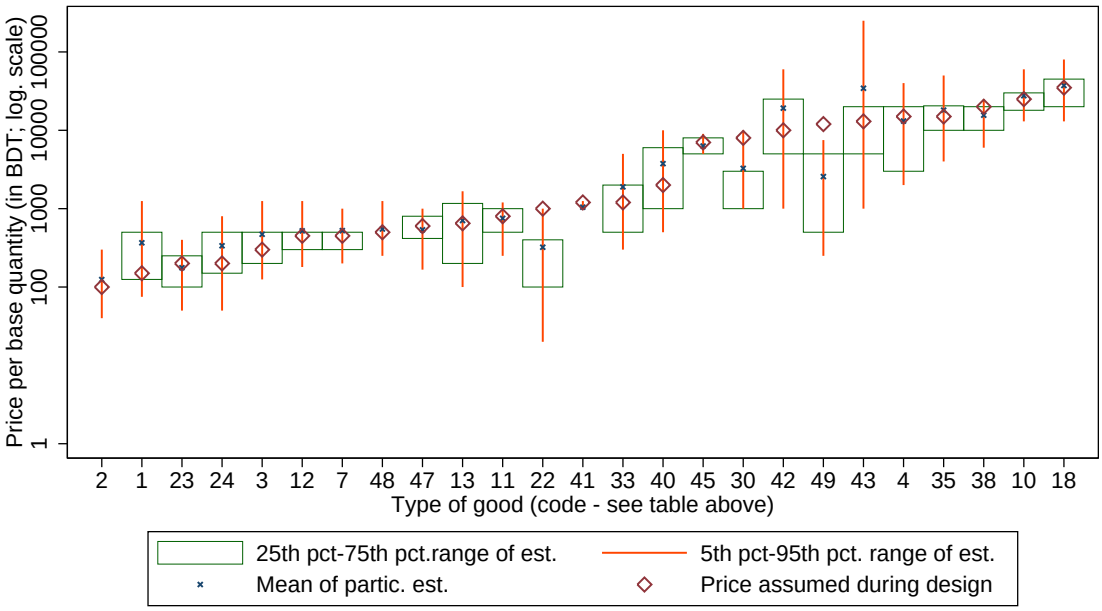
[8]Recall that bundles were constructed to be as close as possible to a constant monetary budget equivalent of BDT 50,000 to avoid arbitrage motives distorting data on preference expression.

[9]Figure 3 presents price data on a logarithmic vertical axis to enable an easily legible overview over unit prices ranging from BDT 1 to BDT 100,000.

While participants' price estimates for most goods are i) on average close to the price assumed when designing the survey and ii) within a compact range, there is a subset of goods for which there was a wider range of estimates or a mean that differs from the assumed value. For some goods, participants provided higher price estimates, on average, than was assumed. This presents a potential challenge for interpreting survey responses as indicative of consumer preferences, if goods associated with a higher-than-assumed price estimate are linked with an elevated average preferences. More rarely, participants also provided lower estimates. However, when they did so, the deviation was relatively large – for example for fishing rods (Good 22) or irrigation tubes (Good 49). Possible reasons for such deviations could include, for example, geographic variation in the underlying price structure or individual participants' low familiarity with certain types of goods. The presence of price estimate variation could compromise subsequent preference elicitation analysis by having introduced arbitrage effects. That is, if participants perceived that one of the two presented bundles carries a substantially higher monetary value than the other, they may have indicated preferring that bundle for that reason rather than by virtue of their consumer preferences. Bearing in mind the possible presence of such confounding effects, I conclude that hypothetical arbitrage motives (stemming from inaccurate or imprecise price estimates) were unlikely to be a dominant driver of participants' bundle preference decisions.[10]

[10]Possible additional analysis to further increase robustness of subsequent analysis to arbitrage effects could be to drop good types that exhibited the highest price estimate variation and to introduce fixed effects at the level of village organisations to capture the highest resolution of geographical demand variation.

Figure 3: Participants' price estimates of goods



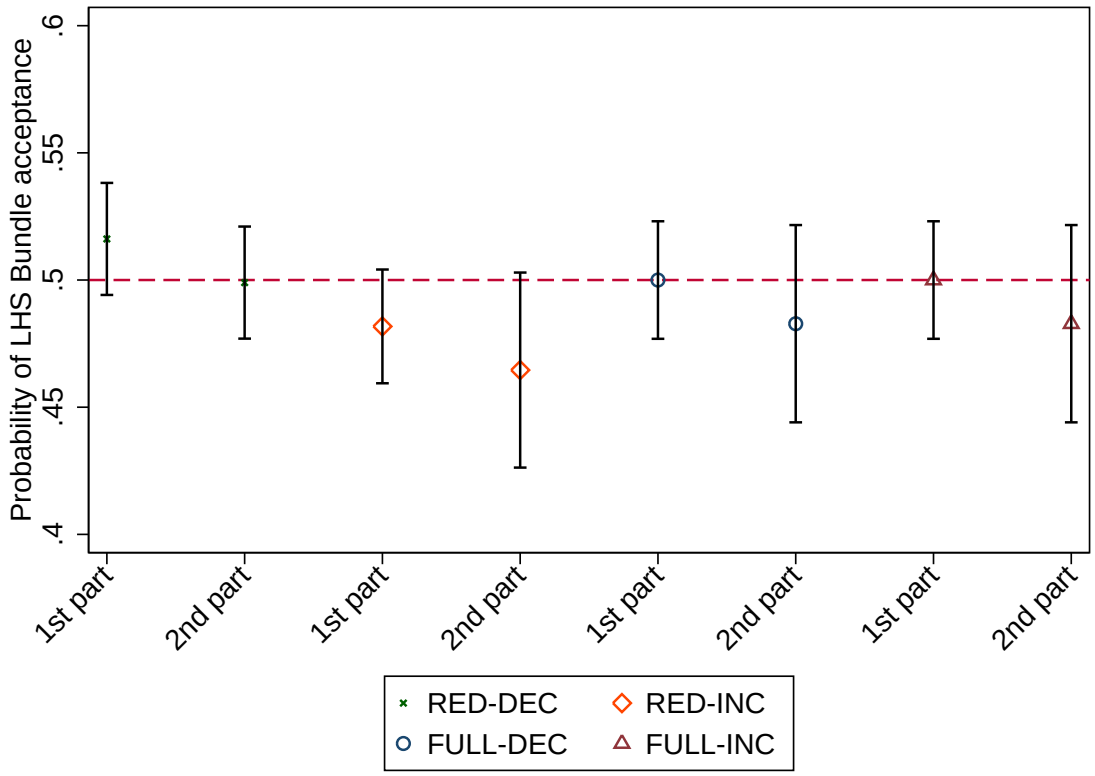
Note: Comparison of price estimate (per respective base quantity) provided by participants versus price assumed when survey was create. For a key to types of goods referred to by codes, please see Table 6.

3.2 Behavioural effects

Behavioural biases during surveying are a well-documented phenomenon. One significant category of such biases, particularly relevant in economic research, is mechanical biases. These biases arise due to the inherent structure and design of surveys, influencing respondents' choices in systematic ways that deviate from rational behaviour and prevent respondents from expressing their true underlying preferences. Common examples include: 1) primacy effect, where respondents tend to prefer the first response option presented in a survey (Kahneman and Tversky (1974), Van Erkel and Thijssen (2016)); 2) acquiescence bias – the tendency for respondents to agree with statements regardless of their actual beliefs (McClendon (1991)); 3) anchoring effects, where respondents are influenced by an initial reference point when making decisions (Ariely et al. (2003)). The bundles used for comparisons in this experiment were created randomly – explicitly including which of the two bundles was presented on which side (left-hand side or right-hand side) of the survey interface used to interact with the participants. A potential concern could be that participants systematically chose bundles on one particular side due to behavioural effects that are not directly related to their consumer preferences over the shown goods. If such patterns were impacting participants' responses, that would weaken the claim that data from a bundle-comparison survey can be used to infer consumer preferences. Figure 4 shows the share of bundle comparisons in which participants indicated that they prefer the left-hand side bundle.[11] Overall, no significant deviations from the (expected) 50% threshold are observed with no patterns across treatment. I therefore conclude that behavioural biases stemming from survey design and / or question setting were unlikely to be dominant influences on participants' responses.

[11] PRE indicates the first set of 15 bundle comparison questions, while POST indicates the second set of 15 bundle comparison questions.

Figure 4: Test for LHS bundle bias



Note: Mean probability of preferring LHS bundle by treatment and period (PRE refers to the 15 bundle prompts shown before the question structure switch, POST refers to the subsequent 15 prompts). Red dashed line marks 50% mark. Confidence intervals shown corresponding to 95% range around mean estimate. Standard errors clustered at the level of the interaction between household ID and bundle comparison ID.

3.3 Regression analysis

In the previous section, I established the validity of the most important assumptions underpinning the interpretation of survey data from this bundle comparison survey as an expression of consumer preferences. Building on this, I now move on to analyse consumer preferences using survey data.

3.3.1 Relative popularity of goods

Recall that the basic building block of this survey was a binary comparison between two bundles of goods. Participants were asked to indicated which of the two bundles they would prefer in a hypothetical scenario where they could have either of them. In the resulting dataset, every observation corresponds to a bundle (consisting of either one or two types of goods) with respect to which a participant recorded a choice. That is, the dataset analysed in this section features observations at the level of individual bundles shown to participants. The dependent variables is a binary indicator variable encoding the choice between the two presented bundles – every bundle-level observation was therefore either preferred (variable assumes value of one) or not preferred (variable assumes value of zero). Linking this bundle preference indicator variable to the goods contained in the respective bundle (goods were allocated through exogenous sources of variance), provides a way to analyse the relative popularity of types of goods. This link is the primary part of the subsequent analysis, through which I illustrate patterns of consumer preferences unveiled by this survey. Regarding sample size, 500 respondent IDs provided responses to 30 bundle-comparison questions each – with each of the two presented bundles creating exactly one resulting observation for the regression analysis. Thus, the datasets contains $500 \times 30 \times 2 = 30,000$ observations, with every respondent ID generating

60 bundle-level observation. Treatment was assigned at the level of respondent IDs and survey covariates covering various household characteristics are also available at the level of respondent IDs. I include household characteristics as regression covariates in order to reduce resulting standard errors – given successful treatment randomisation, no effect on coefficient estimates is to be expected. Unless stated otherwise, the regression analysis below is performed at the level of these 30,000 bundle-level observations with the outcome being the bundle preference indicator variable and the covariates (among others) being the good type container variables, indicating whether or not and – if yes – how much of a good was in the bundle shown to the respective participants. Standard errors are clustered at the level of household IDs. Fixed effects refer to indicators at the level of enumerator, surveying date, and district-level, respectively. Tables 2 and 3 show the impact of the presence of a type of good in a bundle on the probability of that particular bundle being preferred. Tables show regression coefficients of goods container variables obtained through the following estimation equation:

$$y_{i,j} = \beta_0 + \beta_1 \cdot G_j + \beta_2 \cdot D_i + \epsilon_{i,j} \tag{1}$$

where $y_{i,j}$ is the (binary) bundle preference indicator showing whether participant i expressed a preference for specific, randomly allocated bundle j; for every participant i, 60 bundles were shown as part of a randomly selected subset of 30 binary bundle comparison questions (drawn randomly from a pool of 1440 questions); G_j is a (1 x n) matrix where n is the number of types of goods in the survey universe, G_j denotes the quantities of each type of good contained in bundle j; D_i represents covariates at the level of the household of participant i who expressed the respective

preference[12]. For the Probit specification, the linear base model corresponds to the same estimating equation. To summarise the resulting coefficients, Figure 5 provides a list of goods impact estimates, sorted by ascending size of coefficient. I summarise four key findings revealed by these mean-effect estimations:

1. On average, most goods do not affect the probability of the bundle in which they are contained to be preferred. A small set of goods is universally popular - raising the average probability of all bundles they are contained in.

2. Some types of goods are unpopular to the degree that their presence even decreases the probability of a bundle being preferred. Together with the previous finding, this results in an S-shaped distribution of coefficient estimates across goods.

3. Compared to other goods, cash does not carry an outsize popularity.

Tables 2 and 3 show a wide range of regression specifications, illustrating the high robustness of the associated findings. Appendix A Tables 8 and 9 provide an additional robustness check – they show the estimation equivalent to Equation 1, but under a specification where all input variables denoting amounts of goods in bundles were multiplied by the price per unit of the respective good. The purpose of this price adjustment is to identify a potential bias of participants towards goods perceived as particularly cheap or expensive (i.e. not an expression of a consumer preference, but execution of an arbitrage opportunity). Comparing coefficients between the price-adjusted and the non-adjusted specifications suggests that such a bias is unlikely to be driving a significant share of expressed bundle preferences.

[12]Covariates include: number of adults and children in household, average age of household, education level of household head; Fixed effects include: enumerator, surveying date, and district-level.

The coefficients in Tables 2 and 3 capture the causal impact estimate of a given good's presence on the probability of a bundle containing this good being preferred. I propose to read these effects as measures of the respective good's average popularity among respondents. This interpretation is economically valuable as it can serve as a proxy for demand - that is, a utility-maximising allocation or policy bundle would most likely contain the most popular types of goods (subject to budget constraint).

A caveat associated with this economic interpretation of regression estimates is the entirely hypothetical context of surveying. This could, for example, lead to the potential manifestation of experimenter demand effects. That is, respondents may overstate their preference for goods associated with pro-social behaviour and in turn understate preferences for "vice goods". This may lead to a biased estimate of preferences for affected goods. Over the course of survey execution, there was no qualitative evidence of such effects recorded by surveyors familiar with the local cultural context. Nevertheless, in order to establish certainty quantitatively, I recommend implementing appropriate rigorous surveying modules dedicated to the detection of experimenter demand effects in potential follow-up studies.

Table 2: Avg. bundle preference probability effect of good presence (not price-adjusted) *(1/2)*

Good Type	**Part A** – Linear regression						**Part B** – Probit (marg. effects shown)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Cash	0.00	0.00	0.32**	0.42**	1.79	1.33	0.00	0.00	0.81**	1.05**	4.50	3.37
	(0.00)	(0.00)	(0.15)	(0.18)	(1.56)	(1.80)	(0.00)	(0.00)	(0.38)	(0.44)	(3.91)	(4.53)
Boat rental	0.28	0.09	1.96	1.64	3.05	1.07	0.69	0.23	4.91	4.13	7.68	2.71
	(0.55)	(0.65)	(1.55)	(1.81)	(2.57)	(2.99)	(1.38)	(1.63)	(3.89)	(4.54)	(6.45)	(7.50)
Bricks	0.06*	0.09**	1.10**	1.76***	2.78	3.98	0.16*	0.24**	2.76**	4.42***	6.99	10.00
	(0.04)	(0.05)	(0.51)	(0.59)	(2.13)	(2.48)	(0.10)	(0.11)	(1.28)	(1.49)	(5.35)	(6.22)
Fertilizer	0.00	0.02	0.70	1.01**	2.14	1.90	0.01	0.06	1.75	2.53**	5.37	4.79
	(0.03)	(0.04)	(0.44)	(0.51)	(1.87)	(2.17)	(0.08)	(0.10)	(1.11)	(1.29)	(4.69)	(5.43)
Irrigation pump	-0.95	-1.52	0.34	0.06	-1.20	-3.10	-2.40	-3.84	0.85	0.13	-3.02	-7.83
	(1.43)	(1.65)	(1.93)	(2.23)	(2.47)	(2.86)	(3.59)	(4.15)	(4.85)	(5.61)	(6.21)	(7.20)
Construct. labour	0.12	0.39	1.36	2.46*	0.85	1.19	0.30	0.98	3.42	6.18*	2.14	3.01
	(0.37)	(0.42)	(1.12)	(1.29)	(1.94)	(2.24)	(0.92)	(1.06)	(2.81)	(3.23)	(4.86)	(5.61)
Chicken feed	0.14**	0.17**	1.71**	2.34***	4.68*	5.70*	0.35**	0.42**	4.31**	5.90***	11.76*	14.37*
	(0.07)	(0.08)	(0.76)	(0.90)	(2.83)	(3.36)	(0.17)	(0.20)	(1.92)	(2.29)	(7.12)	(8.47)
Cattle feed	-0.08**	-0.10**	-0.44	-0.47	-2.28	-3.58	-0.20**	-0.25**	-1.11	-1.17	-5.70	-8.97
	(0.04)	(0.05)	(0.50)	(0.58)	(2.00)	(2.33)	(0.10)	(0.12)	(1.25)	(1.45)	(5.01)	(5.84)
Chicken	-0.00	0.03	0.52	0.80	1.15	0.53	-0.01	0.07	1.31	2.01	2.91	1.34
	(0.05)	(0.05)	(0.56)	(0.65)	(2.14)	(2.48)	(0.12)	(0.14)	(1.41)	(1.62)	(5.36)	(6.20)
Goat	-1.42**	-1.35*	-3.13*	-2.70	-6.80**	-7.96**	-3.58**	-3.41*	-7.89*	-6.81	-17.16**	-20.08**
	(0.69)	(0.79)	(1.75)	(2.01)	(2.89)	(3.34)	(1.74)	(2.00)	(4.43)	(5.09)	(7.34)	(8.48)
Cow	2.03	0.99	7.50**	7.00*	5.34*	2.72	5.09	2.49	18.83**	17.60*	13.42*	6.84
	(2.52)	(2.90)	(3.28)	(3.75)	(2.80)	(3.23)	(6.33)	(7.26)	(8.22)	(9.39)	(7.04)	(8.10)
School uniform	0.44	0.70	2.21	3.17	3.15	3.40	1.11	1.75	5.56	7.97	7.93	8.55
	(0.61)	(0.71)	(1.72)	(2.00)	(2.85)	(3.29)	(1.52)	(1.78)	(4.31)	(5.02)	(7.14)	(8.25)
Tuition fee	1.13*	1.99***	4.17**	6.90***	4.93*	7.27**	2.84*	5.04***	10.48**	17.43***	12.39*	18.38**
	(0.63)	(0.72)	(1.73)	(1.98)	(2.59)	(2.98)	(1.60)	(1.86)	(4.37)	(5.06)	(6.52)	(7.56)
CNG fuel	-0.13	-0.12	-0.82	-0.67	-3.16	-4.04	-0.33	-0.31	-2.05	-1.68	-7.92	-10.13
	(0.08)	(0.10)	(0.87)	(1.04)	(3.08)	(3.63)	(0.21)	(0.25)	(2.20)	(2.61)	(7.74)	(9.13)
E-Bike	0.00	0.00	0.78	1.55**	4.60	8.92	0.00	0.00	1.95	3.91**	11.56	22.51
	(0.00)	(0.00)	(0.64)	(0.73)	(6.09)	(6.93)	(0.00)	(0.00)	(1.60)	(1.84)	(15.28)	(17.58)
Vehicle repair	-0.00	0.00	0.34	0.60**	2.35	3.69	-0.00	0.00	0.85	1.51**	5.92	9.28
	(0.00)	(0.00)	(0.21)	(0.24)	(2.18)	(2.53)	(0.00)	(0.00)	(0.52)	(0.60)	(5.47)	(6.35)
Rickshaw	0.00	0.00	0.27	0.43**	1.71	2.07	0.00	0.00	0.68	1.08**	4.30	5.21
	(0.00)	(0.00)	(0.17)	(0.20)	(1.99)	(2.29)	(0.00)	(0.00)	(0.44)	(0.50)	(4.98)	(5.74)
Laptop	-5.32	-7.84*	-4.30	-6.40	-4.52	-7.43**	-13.38	-19.79*	-10.81	-16.18	-11.36	-18.73**
	(3.71)	(4.29)	(4.00)	(4.63)	(2.82)	(3.27)	(9.33)	(10.82)	(10.07)	(11.74)	(7.10)	(8.28)
Debt payment	0.00	-0.00	0.33	0.36	1.41	0.09	0.00	-0.00	0.83	0.90	3.55	0.24
	(0.00)	(0.00)	(0.24)	(0.28)	(2.46)	(2.85)	(0.00)	(0.00)	(0.61)	(0.71)	(6.16)	(7.14)
Diesel fuel	-0.08	-0.06	-0.71	-0.32	-3.68	-3.53	-0.20	-0.16	-1.78	-0.79	-9.24	-8.87
	(0.06)	(0.07)	(0.74)	(0.87)	(2.83)	(3.31)	(0.15)	(0.17)	(1.86)	(2.18)	(7.12)	(8.30)
Fish feed	0.01	0.04	0.45	1.05	0.50	1.47	0.02	0.10	1.12	2.64	1.27	3.71
	(0.07)	(0.08)	(0.85)	(0.98)	(3.09)	(3.58)	(0.19)	(0.21)	(2.13)	(2.46)	(7.73)	(8.95)
Fishing rod	-1.02**	-0.91*	-1.47	-0.73	-2.02	-1.96	-2.57**	-2.28*	-3.69	-1.82	-5.06	-4.91
	(0.44)	(0.51)	(1.28)	(1.48)	(2.20)	(2.55)	(1.11)	(1.28)	(3.22)	(3.72)	(5.52)	(6.40)
Fruit tree	0.14	0.30	0.92	1.54	-0.51	-1.10	0.36	0.75	2.32	3.87	-1.27	-2.74
	(0.38)	(0.43)	(1.13)	(1.30)	(2.00)	(2.30)	(0.94)	(1.08)	(2.83)	(3.25)	(5.00)	(5.77)
Soil	-0.09**	-0.11***	-0.42	-0.60	-1.95	-3.82*	-0.22**	-0.28***	-1.06	-1.50	-4.89	-9.59*
	(0.03)	(0.04)	(0.47)	(0.54)	(1.95)	(2.26)	(0.09)	(0.10)	(1.17)	(1.35)	(4.88)	(5.66)
N	30000	22320	30000	22320	30000	22320	30000	22320	30000	22320	30000	22320

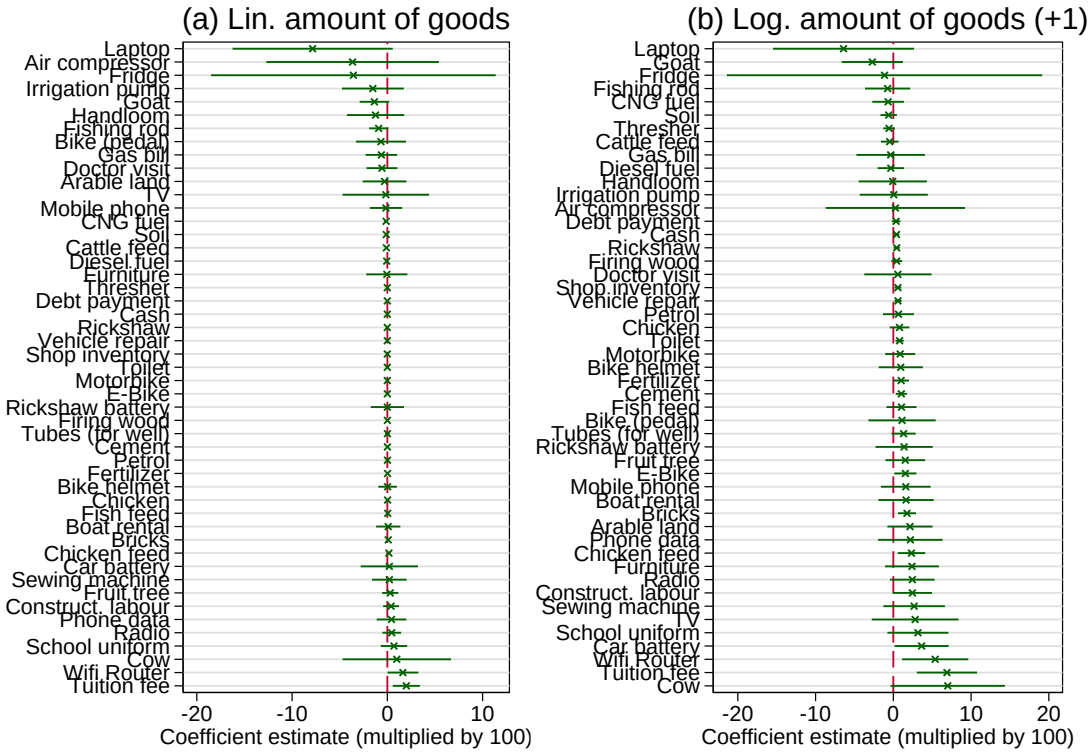
Note: Table shown above represents part 1 out of 2 – please see next page for notes.

Table 3: Avg. bundle preference probability effect of good presence (not price-adjusted) *(2/2)*

Good Type	**Part A** – Linear regression						**Part B** – Probit (marg. effects shown)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fridge	-2.28	-3.56	-0.06	-1.13	-1.22	-3.38	-5.71	-8.96	-0.15	-2.89	-3.07	-8.54
	(6.47)	(7.62)	(8.78)	(10.34)	(6.16)	(7.25)	(16.24)	(19.16)	(22.09)	(26.08)	(15.48)	(18.29)
Gas bill	-0.28	-0.62	0.30	-0.34	-0.11	-2.72	-0.68	-1.55	0.80	-0.79	-0.25	-6.77
	(0.74)	(0.84)	(1.95)	(2.25)	(2.91)	(3.37)	(1.85)	(2.12)	(4.90)	(5.64)	(7.29)	(8.47)
Arable land	-0.19	-0.29	1.60	2.14	0.89	0.17	-0.48	-0.73	4.02	5.40	2.24	0.44
	(1.01)	(1.17)	(1.29)	(1.48)	(1.80)	(2.07)	(2.54)	(2.93)	(3.23)	(3.72)	(4.50)	(5.20)
Toilet	0.00	0.00	0.55**	0.81***	3.98*	5.20**	0.00	0.00	1.37**	2.05***	10.02*	13.11**
	(0.00)	(0.00)	(0.23)	(0.27)	(2.25)	(2.63)	(0.00)	(0.00)	(0.58)	(0.68)	(5.65)	(6.62)
Mobile phone	0.09	-0.14	1.77	1.60	1.10	-0.89	0.22	-0.36	4.44	4.03	2.77	-2.23
	(0.74)	(0.86)	(1.39)	(1.63)	(2.37)	(2.78)	(1.86)	(2.17)	(3.50)	(4.09)	(5.95)	(6.97)
Phone data	-0.73	0.44	-0.90	2.20	-1.92	0.57	-1.83	1.10	-2.24	5.53	-4.80	1.46
	(0.68)	(0.79)	(1.83)	(2.12)	(2.82)	(3.24)	(1.70)	(1.98)	(4.58)	(5.30)	(7.08)	(8.11)
Motorbike	0.00	0.00	0.95	0.88	6.15	2.77	0.00	0.00	2.38	2.23	15.46	6.99
	(0.00)	(0.00)	(0.81)	(0.98)	(7.50)	(9.12)	(0.00)	(0.00)	(2.03)	(2.46)	(18.85)	(22.82)
Bike helmet	-0.08	0.03	0.35	0.96	-1.07	-1.81	-0.21	0.07	0.90	2.42	-2.68	-4.52
	(0.42)	(0.50)	(1.24)	(1.44)	(2.11)	(2.45)	(1.06)	(1.24)	(3.09)	(3.62)	(5.29)	(6.13)
Air compressor	-3.80	-3.65	-1.12	0.27	-2.10	-2.65	-9.56	-9.19	-2.82	0.67	-5.26	-6.66
	(4.00)	(4.62)	(3.95)	(4.57)	(2.76)	(3.20)	(10.05)	(11.60)	(9.91)	(11.48)	(6.92)	(8.02)
Car battery	0.46	0.21	3.23**	3.64**	2.47	1.52	1.16	0.52	8.13**	9.16**	6.22	3.84
	(1.33)	(1.53)	(1.53)	(1.77)	(1.92)	(2.22)	(3.33)	(3.85)	(3.84)	(4.45)	(4.80)	(5.57)
Rickshaw battery	0.32	0.00	1.64	1.39	0.12	-1.54	0.81	0.01	4.11	3.49	0.31	-3.85
	(0.75)	(0.89)	(1.55)	(1.88)	(2.63)	(3.13)	(1.88)	(2.22)	(3.88)	(4.70)	(6.59)	(7.85)
Bike (pedal)	-0.36	-0.67	1.12	1.12	-0.53	-2.08	-0.91	-1.69	2.81	2.81	-1.33	-5.26
	(1.16)	(1.34)	(1.89)	(2.20)	(2.67)	(3.10)	(2.90)	(3.37)	(4.75)	(5.55)	(6.69)	(7.80)
TV	0.05	-0.16	2.27	2.81	0.13	-1.08	0.12	-0.39	5.72	7.08	0.34	-2.69
	(2.01)	(2.32)	(2.48)	(2.85)	(2.37)	(2.73)	(5.05)	(5.81)	(6.22)	(7.14)	(5.93)	(6.85)
Radio	0.31	0.46	1.81	2.45*	2.23	1.63	0.77	1.16	4.54	6.15*	5.61	4.10
	(0.43)	(0.50)	(1.27)	(1.46)	(2.16)	(2.50)	(1.08)	(1.25)	(3.18)	(3.67)	(5.42)	(6.26)
Wifi Router	1.09	1.63**	3.74**	5.38**	4.45	5.11	2.75	4.12**	9.41**	13.57**	11.19	12.87
	(0.72)	(0.83)	(1.89)	(2.17)	(2.97)	(3.42)	(1.81)	(2.10)	(4.75)	(5.49)	(7.45)	(8.62)
Cement	0.01***	0.01**	1.01***	1.03***	4.57**	3.45	0.03***	0.03**	2.53***	2.59***	11.49**	8.67
	(0.00)	(0.00)	(0.31)	(0.36)	(2.03)	(2.37)	(0.01)	(0.01)	(0.78)	(0.91)	(5.10)	(5.94)
Furniture	-0.21	-0.05	1.33	2.40	0.01	-0.32	-0.53	-0.13	3.36	6.05	0.04	-0.80
	(0.95)	(1.10)	(1.53)	(1.77)	(2.17)	(2.52)	(2.39)	(2.76)	(3.84)	(4.44)	(5.44)	(6.31)
Handloom	-0.29	-1.24	1.22	-0.07	-0.23	-3.10	-0.72	-3.10	3.07	-0.16	-0.56	-7.76
	(1.32)	(1.53)	(1.92)	(2.24)	(2.26)	(2.63)	(3.31)	(3.84)	(4.82)	(5.61)	(5.67)	(6.61)
Thresher	-0.00	-0.00*	-0.50	-0.55	-6.38*	-8.29**	-0.00	-0.00*	-1.26	-1.41	-16.10*	-20.93**
	(0.00)	(0.00)	(0.33)	(0.38)	(3.35)	(3.91)	(0.00)	(0.00)	(0.83)	(0.97)	(8.51)	(9.97)
Sewing machine	-0.37	0.21	0.33	2.68	-0.61	2.04	-0.92	0.53	0.84	6.73	-1.50	5.13
	(0.78)	(0.93)	(1.68)	(2.01)	(2.84)	(3.39)	(1.97)	(2.32)	(4.21)	(5.05)	(7.12)	(8.49)
Petrol	-0.05	0.01	-0.21	0.66	-1.60	-0.06	-0.13	0.03	-0.53	1.67	-4.00	-0.14
	(0.07)	(0.09)	(0.85)	(1.01)	(3.02)	(3.61)	(0.18)	(0.22)	(2.13)	(2.54)	(7.58)	(9.04)
Firing wood	0.00	0.01	0.32	0.43	0.59	-0.14	0.01	0.01	0.80	1.09	1.49	-0.33
	(0.00)	(0.00)	(0.31)	(0.35)	(2.07)	(2.38)	(0.01)	(0.01)	(0.77)	(0.88)	(5.17)	(5.95)
Doctor visit	-0.38	-0.56	0.72	0.59	1.37	-0.18	-0.96	-1.42	1.81	1.48	3.46	-0.45
	(0.72)	(0.83)	(1.91)	(2.21)	(2.92)	(3.39)	(1.80)	(2.07)	(4.78)	(5.53)	(7.32)	(8.50)
Tubes (for well)	0.01	0.01	0.99	1.32*	1.72	1.41	0.03	0.02	2.50	3.33*	4.33	3.55
	(0.11)	(0.13)	(0.68)	(0.80)	(2.29)	(2.68)	(0.27)	(0.32)	(1.70)	(1.99)	(5.75)	(6.72)
Shop inventory	0.00	0.00	0.44**	0.60***	2.82	2.86	0.00	0.00	1.11**	1.50***	7.08	7.20
	(0.00)	(0.00)	(0.19)	(0.22)	(1.78)	(2.07)	(0.00)	(0.00)	(0.47)	(0.55)	(4.48)	(5.18)
N	30000	22320	30000	22320	30000	22320	30000	22320	30000	22320	30000	22320
FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
GoodsVar	Lin.	Lin.	Log.(x+1)	Log.(x+1)	I(x>0)	I(x>0)	Lin.	Lin.	Log.(x+1)	Log.(x+1)	I(x>0)	I(x>0)

Note: Average impact of presence of good – **divided by 100** for more accessible display of results – (or log. transformation thereof) in bundle on bundle preference probability. Standard errors clustered at the level of household ID. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

Figure 5: Avg. bundle preference probability effect of good presence



Note: Regression coefficients and associated 95% confidence intervals. Sub-figure (a) corresponds to regression specification show in Column (2) in Tables 2 and 3. Sub-figure (b) corresponds to regression specification show in Column (4) in Tables 2 and 3.

3.3.2 Treatment impact

As outlined above, behavioural effects confounding preferences measurement constitutes a central practical concern with preference elicitation. Bundle comparison surveying reduces complexity of eliciting preferences, but behavioural biases could persist. To enable a rigorous analysis of exactly this impact, I constructed the bundle comparison survey elements in one of four marginally different ways – which was encoded by the four treatments in this experiment. Tables 4 and 5 evaluate what effect the different survey design choices reflected in the four treatments have had on the expressed preferences over each type of good. This represents the fully interacted specification shown below:

$$
\begin{aligned}
&y_{i,j} = \beta_0 + \beta_1 \cdot G_j + \beta_2 \cdot P_{i,j} + \\
&\sum_{k \in \{1,4\}} \Big[\beta_{3,k} \cdot I\{treatment = k\}_{i,j} + \beta_{4,k} \cdot I\{treatment = k\}_{i,j} \cdot G_j + \\
&\beta_{5,k} \cdot I\{treatment = k\}_{i,j} \cdot P_{i,j} + \beta_{6,k} \cdot I\{treatment = k\}_{i,j} \cdot G_j \cdot P_{i,j} \Big] + \\
&\beta_7 \cdot D_i + \epsilon_{i,j}
\end{aligned}
\tag{2}
$$

where $y_{i,j}$ is the (binary) bundle preference indicator showing whether participant i expressed a preference for specific, randomly allocated bundle j; for every participant i, 60 bundles were shown as part of a randomly selected subset of 30 binary bundle comparison questions (drawn from a pool of 1440 questions); G_j is a (1 x n) matrix where n is the number of types of goods in the survey universe, G_j denotes the quantities of each type of good contained in bundle j; D_i represents covariates at

the level of the household of participant i who expressed the respective preference[13]; $P_{i,j}$ is a binary indicator assuming a value of one if participant i answered bundle comparison question j during the second of the two periods (each consisting of 15 bundle comparison questions) making up each surveying instance, and zero otherwise; $I\{treatment = k\}_{i,j}$ is a binary indicator assuming a value of one if bundle comparison question j was answered by participant i under treatment k, and zero otherwise. For the Probit specification, the linear base model corresponds to the same estimating equation.

In order to highlight the interaction terms between goods containers and treatments, I ran a slightly modified set of regressions that does not feature a full interaction between treatment, period flag, and goods container. Instead, this second set of regressions only contained an interaction between treatment and goods container – but period fixed effects on top of that. That is, a regression that corresponds to the following estimating equation:

$$\begin{aligned} y_{i,j} &= \beta_0 + \beta_1 \cdot G_j + \beta_2 \cdot P_{i,j} + \\ &\sum_{k \in \{1,4\}} \Big[\beta_{3,k} \cdot I\{treatment = k\}_{i,j} + \beta_{4,k} \cdot I\{treatment = k\}_{i,j} \cdot G_j \Big] + \\ &\beta_7 \cdot D_i + \epsilon_{i,j} \end{aligned} \tag{3}$$

where $y_{i,j}$ is the (binary) bundle preference indicator showing whether participant i expressed a preference for specific, randomly allocated bundle j; for every participant i, 60 bundles were shown as part of a randomly selected subset of 30 binary

[13]Note that covariates are not interacted with the treatment due to treatment randomisation being successfully exogenous (see analysis of balance above). Covariates include: number of adults and children in household, average age of household, education level of household head; further, also enumerator, surveying date, and district-level fixed effects.

bundle comparison questions (drawn from a pool of 1440 questions); G_j is a (1 x n) matrix where n is the number of types of goods in the survey universe, G_j denotes the quantities of each type of good contained in bundle j; D_i represents covariates at the level of the household of participant i who expressed the respective preference[14]; $P_{i,j}$ is a binary indicator assuming a value of one if participant i answered bundle comparison question j during the second of the two periods (each consisting of 15 bundle comparison questions) making up each surveying instance, and zero otherwise; $I\{treatment = k\}_{i,j}$ is a binary indicator assuming a value of one if bundle comparison question j was answered by participant i under treatment k, and zero otherwise. For the Probit specification, the linear base model corresponds to the same estimating equation.

I display the crucial coefficients for the interaction between treatments and goods containers – $\beta_{4,k}$ in Equation 3 above – in Figures 6 and 7. Please note that the FULL-INC treatment (full goods universe; 1 vs. 1 questions before 2 vs. 2 questions) was chosen as a baseline – and displayed effects therefore correspond to the impact of a good *differential* to the impact of the same good under the FULL-INC treatment. Two findings stand out: firstly, from the perspective of a survey conductor, reducing the goods universe over which preferences are queried does not represent an obviously favourable trade-off. (That is, at least when comparing a scenario with 51 goods vs. 25 goods.) While both parts of Sub-Figure 6 (b) only review a reduced amount of goods compared to Figure 6 (while keeping other elements of survey structure unchanged), there are very few differences in effect estimates. The few significant changes only further accentuate results already available under the

[14]Covariates include: number of adults and children in household, average age of household, education level of household head; further, also enumerator, surveying date, and district-level fixed effects.

FULL-INC treatment.[15] Second, comparing FULL-DEC to FULL-INC in Figure 7 – that is, visualising the causal effect of switching the comparison set-up from bundles of the type (1 good + cash) to bundles of the type (1 good + 1 good) – most goods' popularity has decreased. Under the assumption that treatments did not have any causal impact on consumer preferences themselves, participants' ability to express such preferences was indeed affected by the treatments. One way to interpret the broad decrease of goods' popularity could then be that participants were able to focus their choices more accurately under the FULL-DEC treatment. Recall the S-shaped form of mean goods popularity estimates in Figure 5: a further decrease of relative popularity of already unpopular goods exacerbates the unequal focus on very popular types of goods. Therefore, the presence of cash did have an impact, even though cash per se – as its own type of good, presented in bundles together with other types of goods – was neither particularly popular or unpopular (see Figure 5).

[15]Notably, coefficient estimates are all drawn from the same joint estimation – that is, estimates pertaining to goods from the reduced goods universe (RED-DEC and RED-INC treatments) are based on considerably more non-zero observations than those associated with goods that only feature in the full universe. Consequently, one might expect estimates for reduced-universe goods to be more precise than those only contained in the full universe. Comparing confidence intervals between Figures 6 and 7, this does not appear to be the case.

Table 4: Treatment effect on bundle preference probability
(not price-adjusted, linear specification) *(1/2)*

Type of Good	Code	(1) Base effect for Type of Good	(2) **RED-DEC** X Goods Var.	(3) **RED-INC** X Goods Var.	(4) **FULL-DEC** X Goods Var.
Cash	0	0*	0**	0*	0*
		(0)	(0)	(0)	(0)
Boat rental voucher	1	-1.04	0	0	2.68*
		(.87)	(0)	(0)	(1.37)
Building brick	2	.11	-.12	.14	-.1
		(.14)	(.16)	(.16)	(.19)
Agricultural fertilizer	3	.04	-.01	.03	-.1
		(.11)	(.12)	(.13)	(.15)
Irrigation motor (diesel-fuelled water pump)	4	5.73*	0	0	-17.13**
		(3.06)	(0)	(0)	(7.89)
Construction (unskilled) labour voucher	5	1.84	-1.99	-1.2	-1.64
		(1.25)	(1.42)	(1.48)	(1.67)
Chicken feed	6	.33***	0	0	-.39**
		(.11)	(0)	(0)	(.16)
Goat / Cow feed	7	-.16	.05	.21	-.06
		(.12)	(.15)	(.15)	(.16)
Young chicken / duck (female)	8	.03	.04	.07	-.3
		(.14)	(.17)	(.17)	(.2)
Young goat (female)	9	.86	0	0	-5.41**
		(1.23)	(0)	(0)	(2.67)
Young cow (female)	10	10*	-13.75*	-9.08	-18.33
		(5.9)	(7.51)	(7.74)	(14.67)
School uniform	11	.2	0	0	.63
		(1.04)	(0)	(0)	(1.5)
School tuition fee voucher	12	3.08***	0	0	-2.81*
		(.92)	(0)	(0)	(1.57)
CNG fuel	13	-.29**	0	0	.26
		(.14)	(0)	(0)	(.2)
Downpayment voucher for CNG vehicle	14	0	0	0	0
		(0)	(0)	(0)	(0)
Downpayment voucher for Easybike	15	0	0	0	0
		(0)	(0)	(0)	(0)
Vehicle annual repair voucher	16	0	0	0	0*
		(0)	(0)	(0)	(0)
Downpayment voucher for Rickshaw	17	0	0	0	0
		(0)	(0)	(0)	(0)
Laptop computer	18	-4.31	-1.71	-4.16	-36.87*
		(10.77)	(12.67)	(13.06)	(20.71)
Debt payment voucher	19	0**	0	0	0**
		(0)	(0)	(0)	(0)
Diesel fuel	20	-.12	0	0	.04
		(.1)	(0)	(0)	(.15)
Fish feed	21	.08	0	0	-.12
		(.12)	(0)	(0)	(.18)
Fishing rod	22	-1.18	1.85	.27	-1.73
		(1.17)	(1.52)	(1.42)	(1.92)
Fruit tree seedling	23	-2.07*	1.85	4.08***	3.14**
		(1.12)	(1.33)	(1.42)	(1.52)
Fertile soil	24	.13	-.31***	-.27**	-.25*
		(.1)	(.12)	(.12)	(.14)
Dowry / wedding cost payment voucher	25	0	0	0	0
		(0)	(0)	(0)	(0)
		Model output continued on next page			

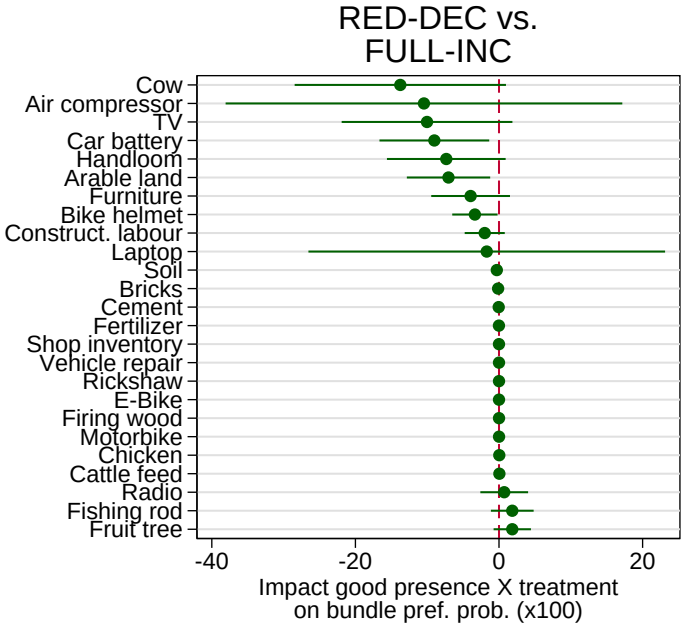
Note: Table shown above represents part 1 out of 2 – please see next page for notes.

Table 5: Treatment effect on bundle preference probability
(not price-adjusted, linear specification) *(2/2)*

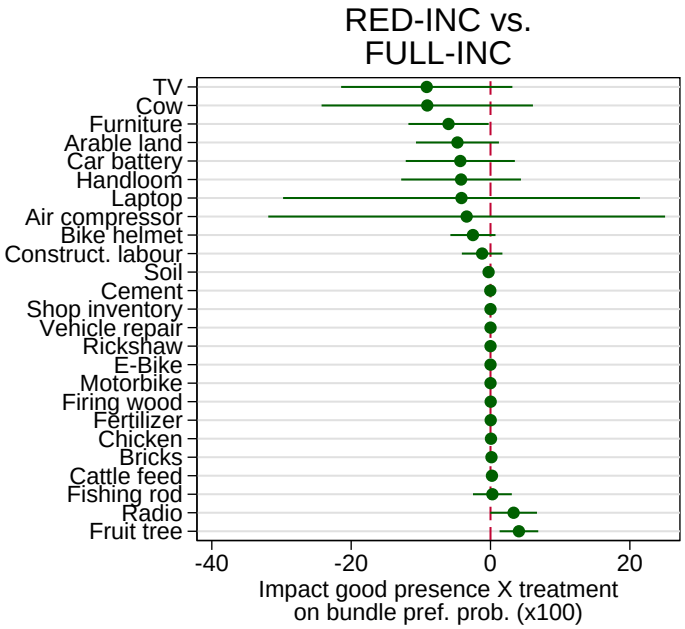
Type of Good	Code	(1) Base effect for Type of Good	(2) **RED-DEC** X Goods Var.	(3) **RED-INC** X Goods Var.	(4) **FULL-DEC** X Goods Var.
New household fridge	26	13.59	0	0	-40.17*
		(11.97)	(0)	(0)	(20.58)
Voucher for household gas bill	27	-.5	0	0	-1.58
		(1.22)	(0)	(0)	(1.93)
Downpayment for purchase / lease of arable land	28	4.82**	-7.03**	-4.74	-13.49**
		(2.28)	(2.96)	(3.04)	(6.01)
Voucher for construction of toilet	29	0***	0	0	0**
		(0)	(0)	(0)	(0)
Mobile phone	30	3.03*	0	0	-7.95**
		(1.57)	(0)	(0)	(4.04)
Mobile phone data plan	31	2.33**	0	0	-4.64***
		(1.06)	(0)	(0)	(1.65)
Motorbike downpayment	32	0***	.01***	0	0
		(0)	(0)	(0)	(0)
Motorbike helmet	33	2.59*	-3.36**	-2.52	-4.49**
		(1.39)	(1.61)	(1.65)	(1.94)
Electric air compressor	34	.41	-10.45	-3.42	-22.39
		(12.07)	(14.09)	(14.53)	(23.34)
Car battery	35	5.88*	-9.01**	-4.33	-16.21**
		(3.01)	(3.9)	(3.99)	(7.93)
Rickshaw battery	36	3*	0	0	-7.4**
		(1.61)	(0)	(0)	(3.74)
Bike	37	2.91	0	0	-10.06
		(2.41)	(0)	(0)	(6.14)
TV	38	6.67	-10.02*	-9.15	-12.1
		(4.79)	(6.07)	(6.27)	(11.31)
Radio	39	-1.4	.73	3.32*	3.51*
		(1.47)	(1.7)	(1.71)	(1.89)
Wifi router	40	2.07*	0	0	-.54
		(1.17)	(0)	(0)	(1.99)
Cement	41	.04***	-.04***	-.03**	-.04*
		(.01)	(.01)	(.01)	(.02)
Bed / Wardrobe / Table	42	4.45**	-3.95	-6.02**	-12.77**
		(2.2)	(2.8)	(2.94)	(5.4)
Handloom machine	43	3.66	-7.34*	-4.22	-11.43
		(3.44)	(4.22)	(4.38)	(7.29)
Paddy threshing machine downpayment	44	0	0	0	0
		(0)	(0)	(0)	(0)
Sewing machine	45	2.09	0	0	-5.62
		(1.54)	(0)	(0)	(3.86)
Petrol fuel	46	.14	0	0	-.52***
		(.11)	(0)	(0)	(.2)
Firing wood	47	0	0	.03*	-.01
		(.01)	(.01)	(.02)	(.02)
Doctor visit voucher	48	.27	0	0	-2.92
		(1.05)	(0)	(0)	(1.8)
Tubes for well construction	49	.51**	0	0	-1.31**
		(.23)	(0)	(0)	(.61)
Shop inventory refill voucher	50	0***	0***	0**	0**
		(0)	(0)	(0)	(0)
Fixed Effects		Yes	Yes	Yes	Yes
Covariates		Yes	Yes	Yes	Yes
Goods variable		Lin.	Lin.	Lin.	Lin.

Note: Interaction effect of treatment with presence of good – **divided by 100** for more accessible display of results – in bundle on bundle preference probability. Treatment effects expressed relative to FULL-INC treatment. Covariates include among standard selection used in previous regression analysis, fixed effects for 1st or 2nd part of questionnaire. Standard errors clustered at the level of household ID. N=22,320. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

Figure 6: Treatment effect interacted with good presence



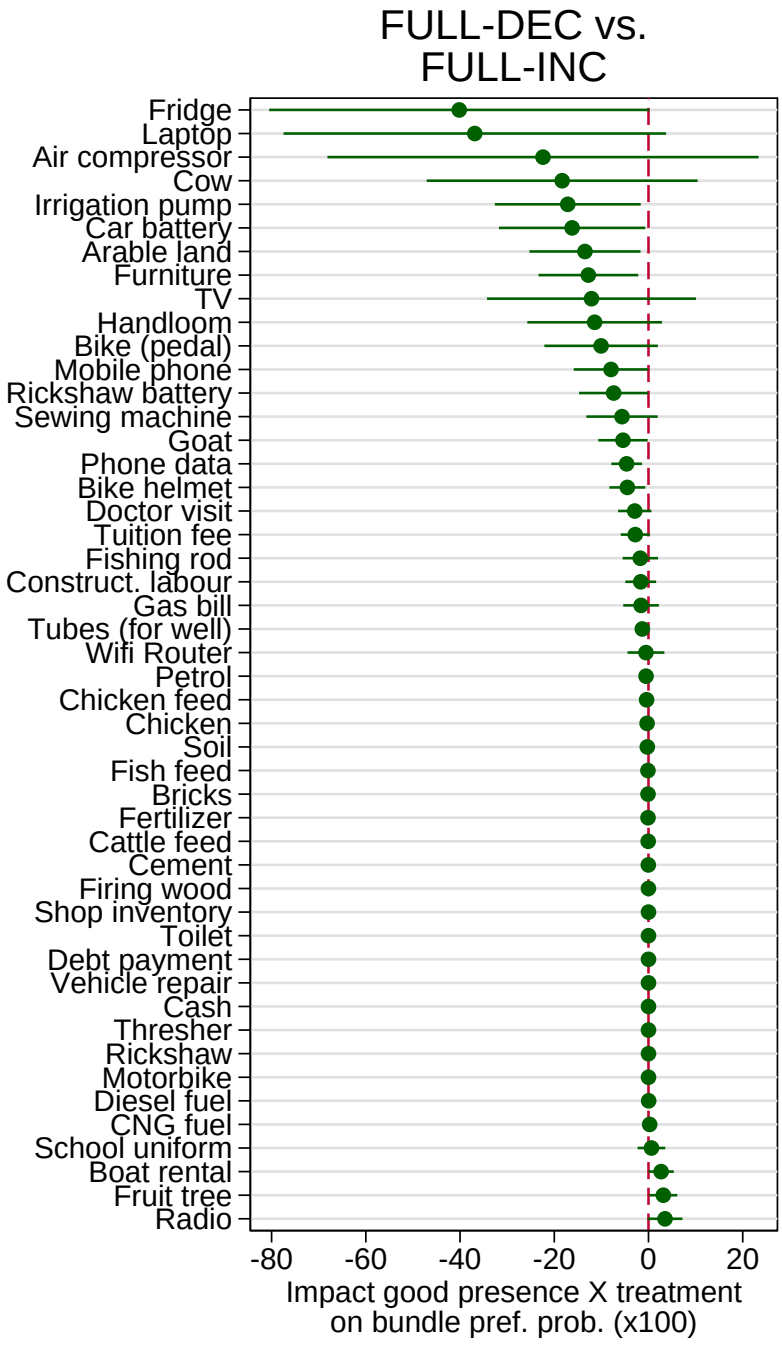
(a) Rel. effect of RED-DEC treatment



(b) Rel. effect of RED-INC treatment

Note: Indicated variables correspond to 95%-confidence intervals surrounding coefficient estimates for the interaction between the variable denoting the goods present in a bundle – **divided by 100** for more accessible display of results – and the assigned treatment. Treatment effects expressed relative to FULL-INC treatment. Covariates include among standard selection used in previous regression analysis, fixed effects for first or second part of questionnaire. Standard errors clustered at the level of household ID. N=22,320.

Figure 7: Treatment effect interacted with good presence



Note: Indicated variables correspond to 95%-confidence intervals surrounding coefficient estimates for the interaction between the variable denoting the goods present in a bundle – **divided by 100** for more accessible display of results – and the assigned treatment. Treatment effects expressed relative to FULL-INC treatment. Covariates include among standard selection used in previous regression analysis, fixed effects for first or second part of questionnaire. Standard errors clustered at the level of household ID. N=22,320.

3.3.3 Interactions between goods

A core property of preference relations is that they permit component goods to cross-influence each other. For example, if a school uniform functions as a strong complement to cash, then the presence of cash could influence the relative popularity of a bundle without this effect necessarily being attributed to cash as a type of good. To assess the magnitude of such goods interactions, the appropriate regression estimation would correspond to an expansion of Equation 1 along the following lines:

$$\begin{aligned} y_{i,j} = \beta_0 + \beta_1 \cdot G_j + \beta_1 \cdot \sum_{\forall\ k,l\ s.t.\ k \neq l} g_{j,k} \cdot g_{j,l} + \\ \beta_3 \cdot C_j + \beta_4 \cdot \sum_{\forall\ k,l\ s.t.\ k \neq l} c_{j,k} \cdot c_{j,l} + \\ \beta_5 \cdot D_i + \epsilon_{i,j} \end{aligned} \tag{4}$$

where $y_{i,j}$ is the (binary) bundle preference indicator showing whether participant i expressed a preference for specific, randomly allocated bundle j; for every participant i, 60 bundles were shown as part of a randomly selected subset of 30 binary bundle comparison questions (drawn from a pool of 1440 questions); G_j is a (1 x n) matrix where n is the number of types of goods in the survey universe, G_j denotes the quantities of each type of good contained in bundle j; $g_{j,k}$ is a scalar denoting the quantity of the k^{th} type of good within the universe contained in G_j; C_j is a (1 x n) matrix that denotes the quantities of each type of good contained in the bundle *opposing* bundle j in the binary choice; $c_{j,k}$ is a scalar denoting the quantity of the k^{th} type of good within the universe contained in C_j; D_i represents covariates at the level of the household of participant i who expressed the respective preference[16].

[16]Covariates include: number of adults and children in household, average age of household, education level of household head; further, also enumerator, surveying date, and district-level fixed

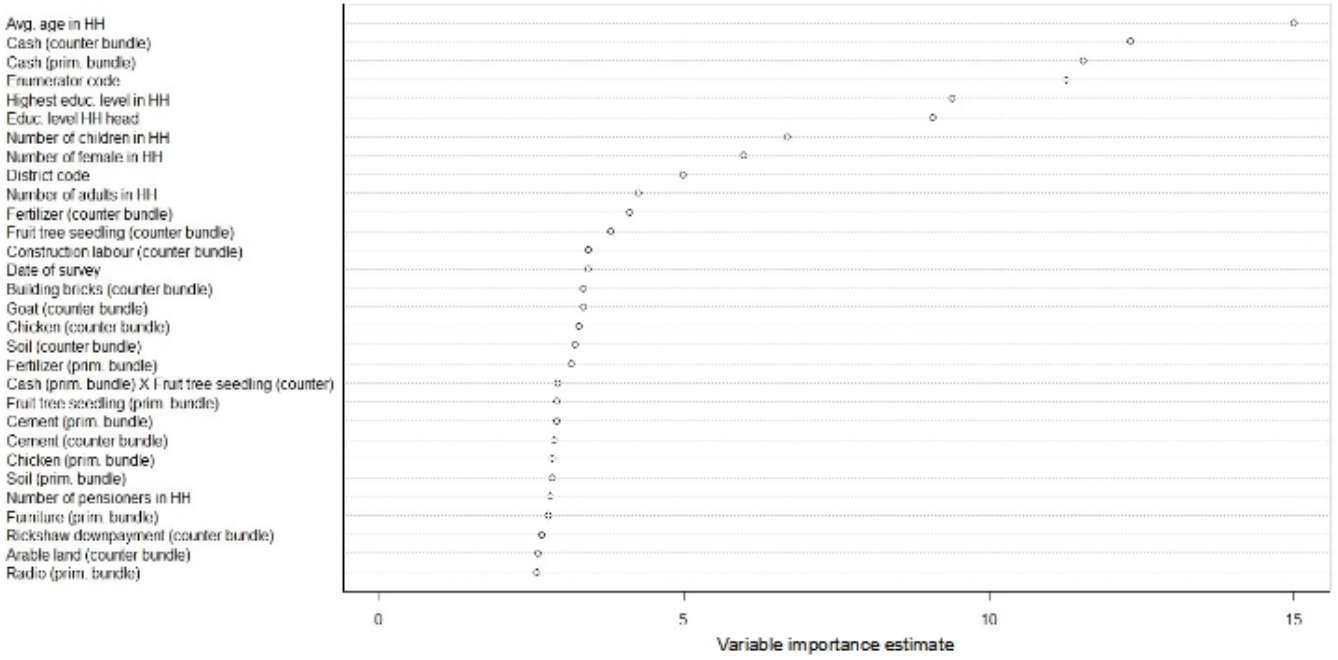
For the case of 51 types of goods, this results in more than 2,500 dependent variables. Coupled with the fact that the vast majority of these interacted variables exhibit primarily values of zero, this presents a sparseness problem for regression analysis, despite a rich dataset of 30,000 bundle observations. To overcome this technical challenge, I estimate a Random Forest algorithm equivalent to Equation 4. This machine-learning technique, popularised by Breiman (2001), estimates a large number of randomly generated decision trees and then aggregates predictions across an entire such "forest" of decision trees. This technique is particularly efficient in aggregating many dependent variables – and moreover, produces an estimates of each involved variables' importance.[17] That is, a measure of that variable's content in information that can be used to predict the outcome. Random Forest being based on a decision-tree approach – it involves creating a high number of randomly generated thresholds in dependent variables and then sequentially bifurcating the flow of predicting the outcome based on whether an observation does or does not reach the respective threshold. Since every decision tree consists of several such threshold bifurcations stacked on each other, this estimation captures (in a non-continuous way) the same variance in the outcome variable that would otherwise have been captured by a model of fully interacted covariates. Figure 8 shows the importance measures corresponding to a Random Forest estimation of Equation 4 for the 30 highest-impact dependent variables. The top 10 variables – which feature a steep decline in importance between them – consist of predominantly household-level covariates. However, cash (both in the primary and in the counter bundle) forms a unique exception to this pattern. This means that the presence of cash

effects.

[17] In the implementation of this algorithm in the *R* environment – which I used in this instance – importance corresponds to the mean loss of outcome accuracy between decision trees that include a given dependent variable and those which do not contain it.

was – on average – a significant driver of bundle preference expression. Additionally, interaction terms between types of goods are not, generally, an outcome driver that is high in importance. Only the interaction between cash (in the primary bundle) and fruit seedlings in the counter bundle is listed among the top 30 dependent variables.

Figure 8: Random Forest estimation



Note: Top 30 covariates by importance as estimated in Random Forest model of bundle preference probability. Horizontal axis displays variable importance as measured by the loss of mean prediction probability associated with respective variable.

3.4 Robustness and rationality

To corroborate the analysis presented previously, I quantify the degree to which respondents have engaged with the survey on the basis of rational decision-making. In microeconomic theory, rationality carries a very specific definition.[18] A common way to test empirical choice data for rationality is to check whether it violates the Weak Axiom of Revealed Preferences (WARP). A Walrasian demand vector $x(p, w)$ satisfies WARP if the following property holds for any (p, w) and (p', w'):

$$p \cdot x(p', w') \leq w \;\; and \;\; x(p', w') \neq x(p, w) \implies p' \cdot x(p, w) > w \tag{5}$$

where (p, w) and (p', w') are the price vectors and budget inducing the respective Walrasian demand $x(p, w)$ and $x(p', w')$.

A simplified and accentuated application of a WARP test was integrated in the design of this survey: at the end of each subset of 15 bundle-comparison questions, the final question was analogous to the first question shown in the respective subset – with two key changes. Firstly, the position of the two bundles was switched (that is, left-hand bundle became right-hand bundle); secondly, a randomly selected type of good (featured with a non-zero quantity in the original good) was dropped from the bundle. As a consequence, one of the two bundles presented in the first question always strictly dominated the opposing bundle in the last question (while the other bundle remained identical and switched sides). From the perspective of a rational, utility maximising agent, that has a very clear implication on bundle choice (through WARP): if the agent prefers the non-altered bundle the first time it was shown, then the agent must also prefer the non-altered bundle when it is shown compared to a bundle that is strictly inferior to the previously opposing bundle. This provides a

[18] That is, a preference relation is rational exactly if it is both transitive and complete.

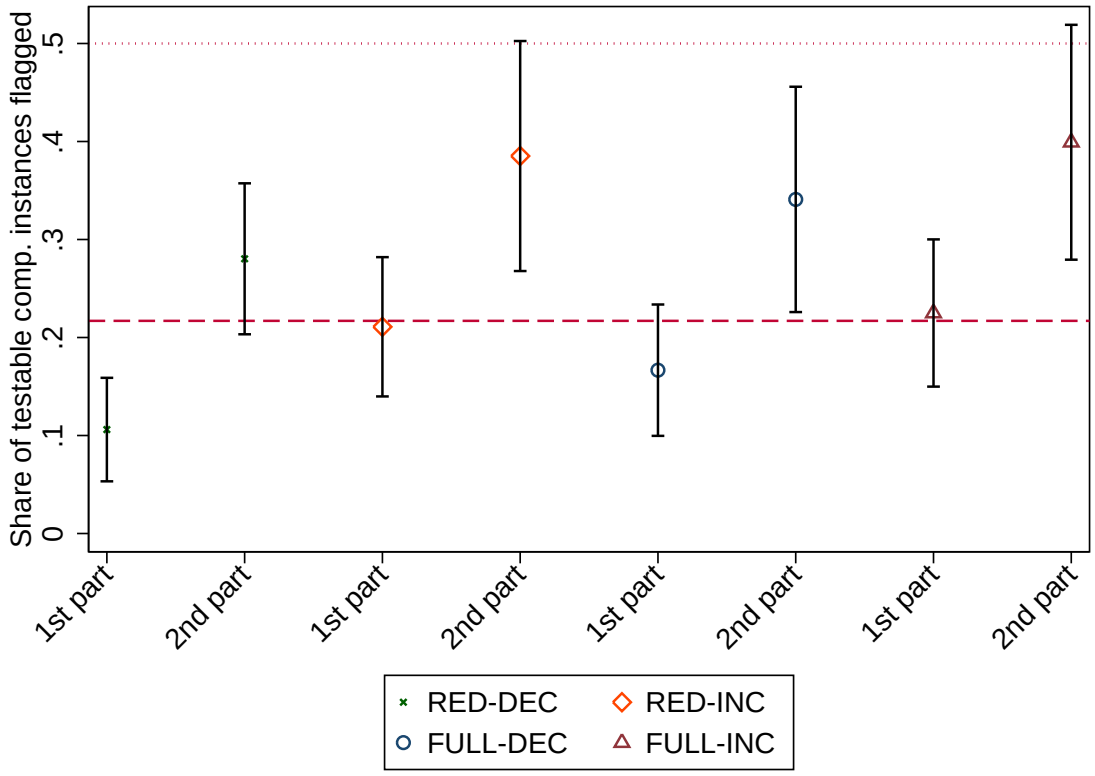
testable criterion for the validity of WARP.

Figure 11 illustrates the share of WARP-testable comparison instances where participants did choose a bundle that was inconsistent with WARP. Most strikingly, the share of WARP-inconsistent choices increases across all treatments by a very similar amount between the first 15 survey questions and the second 15 survey questions, on average. This forms a strong indication, that a decrease in participant attention drove an increase in the probability in overtly non-rational choices. Further, Figure 11 also hints at WARP-inconsistent choices having been made less frequently under treatments that started with bundle comparisons of the type (1 good + 1 good) and then moved to comparisons of the type (1 good + cash), compared to treatments that did this the other way around. That is particularly visible for the RED-DEC treatment. Overall, 21.7% of testable bundle choices were responded to in a way that is inconsistent with WARP – indicating that rational thought was very likely an overall significant driver for participants' expressed bundle choices. Notable caveats include: firstly, the degree of irrational choices seems to have increased over time; and secondly, treatments that placed an emphasis on cash in bundles seem to have exhibited a higher degree of irrational choices, on average. WARP is a prominent concept within microeconomics and many influential studies have studied its applicability in various empirical settings. While most studies identify a strong predictive power of WARP, it also appears that economic agents frequently violate WARP – especially over an extended time horizon.[19] The respondents in this sample therefore

[19]Echenique et al. (2011) find that almost all households involved in a long-term panel violate WARP at least partially. Blundell et al. (2003) also show that household panels can feature frequent departures from WARP over a long period of observation. In an experimental context, Andreoni and Miller (2002) find that up to 78% of participants violate WARP over the course of a budgeting exercise. Similarly Sippel (1997) finds commonplace violation of WARP in another experimental setting. Hoderlein and Stoye (2014) highlight especially frequent WARP violation in the case of three or more goods involved.

exhibit a relatively high degree of rationality comparable to that of other populations of economic agents. Nevertheless, I present the primary elements of subsequent analysis both including and excluding the sub-population of WARP non-adherent respondents as a robustness check. Appendix A Tables 10 and 11 show that while these effect gain in significance upon dropping agents that have been found to not adhere to WARP, the patterns of effects remains robust.

Figure 9: Test for instances of WARP violation



Note: Share of those WARP-testable observations for which participants chose bundles in violation of WARP. Red dashed line marks total sample average at the 21.7% mark; red dotted line indicates 50% mark. Confidence intervals shown corresponding to 95% range around mean estimate. Standard errors clustered at the level of the interaction between household ID and bundle comparison ID.

4 Conclusion

In this chapter, I demonstrate how bundle-comparison surveying can serve as an efficient method for consumer preference elicitation in the field. I show evidence from a pilot of this method that I designed and rolled out to the field in Bangladesh with the help of BRACU. The principal results are, firstly, a conceptual validation of bundle-comparison surveying – there is a robust set of data points implying that I was able to measure participants' consumer preferences over a broad universe of goods at a low cost. I analysed the resulting data by estimating the impact of the presence of a type of good in a bundle on the probability that this bundle is chosen by a participant when prompted with it using conventional reduced-form methods. These results showed that there is an identifiable subset of goods which attract a significant degree of popularity among respondents – important and directly actionable insight for policymakers. Building on the literature on microfinance heterogeneity, I initially provided a methodological contribution which can be used by applied microeconomists to study empirical preference relations.

I also added to the still developing literature on the role of cash versus in-kind for microfinance. Recent contributions, such as McIntosh and Zeitlin (2024), show that cash transfers are able to create outsize positive effects on consumption and investment among populations typically targeted by microfinance interventions – largely driven by behavioural mechanisms. The evidence I showed paints a nuanced picture of cash, composed of three levels: i) cash as a good: there was no average popularity effect dissimilar to that of other types of goods (Table 2); ii) cash as a focus generator: survey design variations with cash inclusion helped participants manifest clearer views on their preferences for non-cash goods (Table 4); iii) cash as a measuring stick: Random Forest analysis suggests that there are very strong high-dimensional goods-cash interaction effects that are obscured at the level of analysing average popularity effects (Figure 8). I interpret this as evidence that participants use the presence of cash as a powerful value-framing device – with the mere presence of cash in bundles

helping respondents set clearer priorities. Moreover, I contributed to the behavioural literature on surveying by quantifying the impact of certain biases and offering the following recommendations for the design of bundle-comparison surveys in the field based on these estimations. Firstly, attention management is critical as there appears to be a correlation between survey progress and WARP violation. Secondly, there seems to be a rather flat trade off between number of types of goods within the queried universe and the clarity of responses – encouraging researchers to open survey scope. I conclude by highlighting the significant potential for further study shown by the high-dimensional datasets involved in bundle comparison. Techniques best suited to extracting additional insight would need to be particularly comfortable with processing sparse data – as is the case with more recent machine-learning methods, for example.

References

Andreoni, J., & Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Angelucci, M., De Giorgi, G., & Rasul, I. (2018). Consumption and investment in resource pooling family networks. *The Economic Journal*, *128*(615), 2613–2651.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*(1), 73–106. https://doi.org/10.1162/00335530360535153

Armendariz, B., & Labie, M. (2011). *The handbook of microfinance*. World scientific.

Banerjee, A., Karlan, D., & Zinman, J. (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, *7*(1), 1–21. https://doi.org/10.1257/app.20140287

Bari, F., Malik, K., Meki, M., Quinn, S., Morduch, J., Ogden, T., & Said, F. (2024). Asset-based microfinance for microenterprises: Evidence from pakistan. *American Economic Review*, *114*(2), 534–574.

Blundell, R. W., Browning, M., & Crawford, I. A. (2003). Nonparametric engel curves and revealed preference. *Econometrica*, *71*(1), 205–240.

BRAC. (2021). *Annual report 2021 - brac* [Accessed on May. 17, 2023]. https://www.brac.net/downloads/ar2021/BRAC.pdf

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Crépon, B., Devoto, F., Duflo, E., & Pariente, W. (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in morocco. *American Economic Journal: Applied Economics*, *7*(1), 123–150.

Cueva, R. A., Osman, A., & Speer, J. D. (2024). Microfinance's transformational potential: Looking beyond average treatment effects. *Oxford Review of Economic Policy*, *40*(1), 71–81.

Cunha, J. M., De Giorgi, G., & Jayachandran, S. (2019). The price effects of cash versus in-kind transfers. *The Review of Economic Studies*, *86*(1), 240–281.

De Mel, S., McKenzie, D., & Woodruff, C. (2008). Returns to capital in microenterprises: Evidence from a field experiment. *The Quarterly Journal of Economics*, *123*(4), 1329–1372.

Echenique, F., Lee, S., & Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, *119*(6), 1201–1223.

Fafchamps, M., McKenzie, D., Quinn, S., & Woodruff, C. (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in ghana. *Journal of development Economics*, *106*, 211–226.

Hensher, D. A., & Johnson, L. W. (2018). *Applied discrete-choice modelling*. Routledge.

Hoderlein, S., & Stoye, J. (2014). Revealed preferences in a heterogeneous population. *Review of Economics and Statistics*, *96*(2), 197–213.

Jeong, D., & Trako, I. (2022). *Cash and in-kind transfers in humanitarian settings* (tech. rep.). World Bank, Washington, DC.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, *5*(1), 193–206.

Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Karlan, D., & Zinman, J. (2010). Expanding credit access: Using randomized supply decisions to estimate the impacts. *Review of Financial Studies*, *23*(1), 433–464.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management science*, *62*(12), 3393–3411.

Karlan, D., Osei, R., Osei-Akoto, I., & Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, *129*(2), 597–652.

Khan, A. Q., Khwaja, A. I., Olken, B. A., & Shaukat, M. (2022). *Rebuilding the social compact: Urban service delivery and property taxes in pakistan* (tech. rep.).

McClendon, M. J. (1991). Acquiescence and recency response-order effects in interview surveys. *Sociological Methods & Research*, *20*(1), 60–103.

McIntosh, C., & Zeitlin, A. (2024). Cash versus kind: Benchmarking a child nutrition program against unconditional cash transfers in rwanda. *The Economic Journal*, *134*(664), 3360–3389.

Meager, R. (2019). Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, *11*(1), 57–91.

Meki, M., & Quinn, S. (2024b). Microfinance: An overview. *Oxford Review of Economic Policy*, *40*(1), 1–7.

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.

Quinn, S., & Woodruff, C. (2019). Experiments and entrepreneurship in developing countries. *Annual Review of Economics*, *11*(1), 225–248.

Roodman, D. (2012). *Due diligence: An impertinent inquiry into microfinance*. Center for Global Development.

Sippel, R. (1997). An experiment on the pure theory of consumer's behaviour. *The Economic Journal*, *107*(444), 1431–1444.

Strömberg, D. (2007). Natural disasters, economic development, and humanitarian aid. *Journal of Economic perspectives*, *21*(3), 199–222.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, *4*(3), 199–214.

Van Erkel, P. F., & Thijssen, P. (2016). The first one wins: Distilling the primacy effect. *Electoral Studies*, *44*, 245–254.

Varian, H. R. (2014). *Intermediate microeconomics with calculus: A modern approach*. W.W. Norton & Company.

Appendices

A Appendix – Auxiliary Results

Table 6: Allocation of observations across types of goods

Good type	(1) Description	(2) Base quantity	(3) Est. price per base quantity (in BDT)	(4) No. of treatments	(5) No. of bundle sets	(6) No. of household IDs	(7) No. of bundle obs.
0	Cash	1 BDT	1	4	48	500	14546
1	Boat rental voucher	1 week	150	2	23	230	580
2	Building brick	10 pieces	100	4	43	450	1158
3	Agricultural fertilizer	10 kgs	300	4	47	490	1743
4	Irrigation motor (diesel-fuelled water pump)	1 pieces	15000	2	22	220	660
5	Construction (unskilled) labour voucher	1 person for 1 day	700	4	42	440	1482
6	Chicken feed	10 kgs	500	2	22	220	410
7	Goat / Cow feed	10 kgs	450	4	45	470	1433
8	Young chicken / duck (female)	10 pieces	1000	4	43	450	1064
9	Young goat (female)	1 pieces	4500	2	18	180	430
10	Young cow (female)	1 pieces	25000	4	33	349	574
11	School uniform	1 set for 1 child	800	2	20	200	450
12	School tuition fee voucher	1 for 1 child for 1 month	450	2	22	220	550
13	CNG fuel	10 litres	650	2	19	190	350
14	Downpayment voucher for CNG vehicle	10000 BDT	100000	0	0	0	0
15	Downpayment voucher for Easybike	10000 BDT	50000	3	7	76	86
16	Vehicle annual repair voucher	10000 BDT	10000	4	40	418	966
17	Downpayment voucher for Rickshaw	10000 BDT	5000	4	40	420	1433
18	Laptop computer	1 pieces	35000	4	29	307	434
19	Debt payment voucher	10000 BDT	12000	2	24	240	770
20	Diesel fuel	10 litres	800	2	22	220	440
21	Fish feed	10 kgs	550	2	19	190	360
22	Fishing rod	1 pieces	1000	4	40	420	978
23	Fruit tree seedling	1 pieces	200	4	45	470	1506
24	Fertile soil	10 kgs	200	4	45	470	1612
25	Dowry / wedding cost payment voucher	10000 BDT	100000	0	0	0	0
26	New household fridge	1 pieces	30000	2	6	60	70
27	Voucher for household gas bill	1 months	1350	2	20	200	380
28	Downpayment for purchase / lease of arable land	1 decimals	12000	4	45	470	3210
29	Voucher for construction of toilet	10000 BDT	15000	2	24	240	1010
30	Mobile phone	1 pieces	8000	2	24	240	750
31	Mobile phone data plan	1 months	450	2	19	190	430
32	Motorbike downpayment	10000 BDT	50000	4	5	52	52
33	Motorbike helmet	1 pieces	1200	4	44	460	1076
34	Electric air compressor	1 pieces	40000	4	34	359	477
35	Car battery	1 pieces	15000	4	47	490	2600
36	Rickshaw battery	1 pieces	7000	2	22	220	540
37	Bike	1 pieces	12000	2	20	200	540
38	TV	1 pieces	20000	4	37	390	743
39	Radio	1 pieces	600	4	45	470	1055
40	Wifi router	1 pieces	2000	2	21	210	380
41	Cement	100 kgs	1200	4	45	470	1323
42	Bed / Wardrobe / Table	1 pieces	10000	4	43	450	1252
43	Handloom machine	1 pieces	13000	4	39	409	834
44	Paddy threshing machine downpayment	10000 BDT	10000	2	15	150	280
45	Sewing machine	1 pieces	7000	2	20	200	450
46	Petrol fuel	10 litres	900	2	17	170	370
47	Firing wood	100 kgs	600	4	45	470	1322
48	Doctor visit voucher	1 pieces	500	2	19	190	430
49	Tubes for well construction	10 meters	12000	2	24	240	920
50	Shop inventory refill voucher	10000 BDT	15000	4	48	500	4037

Note: Overview of types of goods used to construct bundles which were presented to participants. Column (4) and (5) indicate the number of treatments and sets of bundle comparisons, respectively, in which each type of good has featured. Column (6) and (7) show the number of household-level and bundle-level observations, respectively, in which each type of good has featured.

Table 7: Balance of covariates

Group			All	(1) RED-DEC	(2) RED-INC	(3) FULL-DEC	(4) FULL-INC
No. of adults in HH			2.94	2.81	3.05	2.86	3.04
			[3]	[3]	[3]	[3]	[3]
			(1.16)	(1.17)	(1.22)	(1.06)	(1.16)
No. of children in HH			1.62	1.68	1.55	1.5	1.76
			[2]	[2]	[1]	[2]	[2]
			(1.12)	(1.14)	(1.07)	(1.08)	(1.17)
No. of females in HH			2.21	2.18	2.26	2.08	2.34
			[2]	[2]	[2]	[2]	[2]
			(1.07)	(1.05)	(1.07)	(1.01)	(1.15)
No. of pensioners in HH			.23	.2	.23	.25	.24
			[0]	[0]	[0]	[0]	[0]
			(.49)	(.44)	(.49)	(.54)	(.48)
Avg. age in HH			29.11	28.01	29.5	30.77	28.26
			[28]	[27]	[29.37]	[28.27]	[27.75]
			(9.11)	(9)	(8.44)	(10.48)	(8.26)
Is HH head female?			.17	.19	.17	.17	.17
			[0]	[0]	[0]	[0]	[0]
			(.38)	(.39)	(.38)	(.37)	(.37)
Is HH head married?			.9	.92	.89	.91	.89
			[1]	[1]	[1]	[1]	[1]
			(.3)	(.28)	(.31)	(.29)	(.31)
HH head yrs. of edu.			6.38	6.08	7.03	6.15	6.29
			[5]	[5]	[7]	[5]	[5]
			(5.05)	(3.28)	(5.60)	(5.73)	(5.51)
Top edu. in HH (in yrs.)			10.75	11.38	11.83	9.51	10.17
			[9]	[9.5]	[10]	[9]	[9]
			(9.6)	(10.95)	(11.37)	(7.18)	(7.86)
Observations			500	132	128	120	120
P-value testing	**(1)=(2)**	**(1)=(3)**	**(1)=(4)**	**(2)=(3)**	**(2)=(4)**	**(3)=(4)**	**(1)=(2)=(3)=(4)**
No. of adults in HH	.123	.757	.106	.143	.933	.284	.221
No. of children in HH	.327	.121	.575	.716	.066*	.088*	.176
No. of females in HH	.506	.524	.314	.132	.502	.12	.318
No. of pensioners in HH	.596	.357	.509	.832	.925	.918	.765
Avg. age in HH	.211	.025**	.795	.331	.28	.066*	.13
Is HH head female?	.669	.635	.675	.898	.924	1	.96
Is HH head married?	.478	.749	.509	.597	.982	.63	.837
HH head yrs. of edu.	.124	.91	.73	.209	.394	.859	.365
Top edu. in HH (in yrs.)	.763	.267	.429	.054*	.166	.476	.273

Note: Main figure shown for each variable is sample mean; sample median is in square brackets and sample standard error is in parentheses. P-values in the lower panel refer to the confidence in the null-hypothesis of equal distributions between sub-samples (standard errors are clustered at the level of Districts).

Table 8: Avg. bundle preference probability effect of good presence (price-adjusted) *(1/2)*

Good Type	**Part A** – Linear regression						**Part B** – Probit (marg. effects shown)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
main												
Cash	0.33***	0.33***	1250.36***	1202.03***	13056.20***	11937.38***	0.83***	0.85***	3173.66***	3050.07***	33222.41***	30328.80***
	(0.10)	(0.12)	(203.99)	(236.63)	(1956.73)	(2273.17)	(0.26)	(0.30)	(525.31)	(607.31)	(5076.90)	(5857.37)
Boat rental	0.87	-1.64	2012.28***	1563.28**	14754.31***	11075.66***	2.18	-4.18	5102.84***	3970.23**	37501.56***	28161.82***
	(4.80)	(5.69)	(527.13)	(616.37)	(3295.51)	(3854.49)	(12.03)	(14.34)	(1333.60)	(1560.40)	(8376.86)	(9780.79)
Bricks	0.98**	1.16**	1656.40***	1676.82***	14962.83***	14271.43***	2.45**	2.92**	4197.42***	4246.77***	38017.94***	36195.84***
	(0.49)	(0.58)	(321.70)	(373.91)	(2690.25)	(3124.94)	(1.23)	(1.45)	(818.36)	(949.50)	(6878.17)	(7956.41)
Fertilizer	0.11	0.20	1309.58***	1311.35***	13885.96***	13078.74***	0.28	0.50	3322.39***	3324.80***	35318.43***	33204.94***
	(0.14)	(0.16)	(249.62)	(288.94)	(2352.93)	(2725.02)	(0.35)	(0.41)	(636.37)	(735.04)	(6041.32)	(6962.46)
Irrigation pump	0.27**	0.27*	907.37***	857.83**	9336.39***	8216.56**	0.68**	0.69*	2308.44***	2181.11**	23838.71***	20932.94**
	(0.12)	(0.14)	(314.15)	(367.18)	(3193.60)	(3735.31)	(0.32)	(0.37)	(798.29)	(931.35)	(8130.63)	(9480.21)
Construct. labour	0.91	1.39*	1638.49***	1679.37***	14231.98***	13750.97***	2.29	3.51*	4153.24***	4254.65***	36173.06***	34886.71***
	(0.68)	(0.78)	(312.70)	(359.32)	(2449.51)	(2825.38)	(1.70)	(1.95)	(796.10)	(913.79)	(6275.83)	(7210.99)
Chicken feed	0.47***	0.47**	1862.42***	1913.55***	19514.89***	19384.74***	1.19***	1.18**	4719.84***	4850.69***	49526.61***	49153.21***
	(0.17)	(0.20)	(374.64)	(447.80)	(3639.84)	(4356.92)	(0.45)	(0.52)	(957.35)	(1144.33)	(9305.73)	(11127.63)
Cattle feed	-0.07	-0.12	908.06***	738.22**	10294.88***	7948.88***	-0.16	-0.29	2313.57***	1884.48**	26305.32***	20322.12***
	(0.12)	(0.13)	(262.52)	(305.76)	(2539.32)	(2963.78)	(0.29)	(0.34)	(667.77)	(776.82)	(6499.95)	(7557.43)
Chicken	0.05	0.10	1122.62***	1150.86***	13145.29***	12554.41***	0.12	0.26	2849.57***	2918.66***	33453.60***	31882.22***
	(0.06)	(0.07)	(263.81)	(305.99)	(2734.50)	(3177.42)	(0.15)	(0.18)	(669.16)	(775.42)	(6973.58)	(8075.80)
Goat	-0.10	-0.06	169.61	63.20	2581.07	702.19	-0.27	-0.16	437.30	159.69	6706.07	1806.31
	(0.20)	(0.23)	(391.65)	(452.28)	(3785.80)	(4368.69)	(0.51)	(0.60)	(1009.82)	(1168.23)	(9788.69)	(11311.86)
Cow	0.48***	0.48***	2134.29***	1922.21***	22928.24***	19772.39***	1.23***	1.22***	5417.33***	4872.63***	58329.82***	50179.13***
	(0.13)	(0.15)	(344.49)	(398.38)	(3506.46)	(4066.46)	(0.33)	(0.38)	(888.93)	(1018.98)	(9098.97)	(10427.72)
School uniform	1.08	1.73	1628.53***	1750.47***	14426.78***	14571.56***	2.72	4.37	4128.11***	4433.66***	36663.19***	36947.65***
	(1.01)	(1.18)	(462.60)	(541.07)	(3682.84)	(4300.33)	(2.55)	(2.99)	(1168.57)	(1367.27)	(9327.78)	(10875.40)
Tuition fee	2.98	4.92**	2171.38***	2517.94***	17202.68***	18788.98***	7.52	12.48**	5499.77***	6385.24***	43667.95***	47684.57***
	(1.86)	(2.16)	(472.91)	(552.32)	(3378.93)	(3956.33)	(4.71)	(5.58)	(1202.15)	(1415.04)	(8615.70)	(10140.50)
CNG fuel	-0.29*	-0.23	562.67	497.96	7562.94*	6037.95	-0.74*	-0.59	1443.20	1279.51	19426.20*	15520.81
	(0.16)	(0.19)	(403.99)	(477.68)	(3964.63)	(4693.41)	(0.41)	(0.49)	(1024.38)	(1210.69)	(10066.94)	(11901.12)
E-Bike	2.16**	2.94***	3138.70***	3793.40***	30080.78***	34729.39***	5.47**	7.47***	7956.74***	9635.78***	76432.52***	88299.92***
	(0.86)	(0.98)	(859.63)	(970.95)	(7846.17)	(8861.90)	(2.19)	(2.55)	(2178.68)	(2524.64)	(19918.91)	(23071.05)
Vehicle repair	0.38***	0.42***	1521.18***	1566.00***	16229.70***	15999.91***	0.95***	1.07***	3855.56***	3967.28***	41221.88***	40570.01***
	(0.13)	(0.15)	(279.93)	(320.71)	(2771.88)	(3178.12)	(0.32)	(0.37)	(712.87)	(815.91)	(7086.40)	(8102.11)
Rickshaw	0.19***	0.24***	1169.60***	1250.37***	13141.20***	13224.45***	0.48***	0.60***	2968.72***	3170.02***	33441.92***	33571.66***
	(0.07)	(0.08)	(244.02)	(279.66)	(2519.08)	(2888.56)	(0.18)	(0.21)	(621.82)	(711.61)	(6451.51)	(7370.39)
Laptop	0.19	0.18	722.71**	579.34	7775.39**	5583.34	0.49	0.46	1841.15**	1475.97	19882.18**	14261.13
	(0.13)	(0.16)	(344.80)	(398.28)	(3552.24)	(4105.04)	(0.35)	(0.40)	(875.53)	(1011.24)	(9040.19)	(10434.94)
Debt payment	0.32**	0.33**	936.54***	816.96**	9006.94***	7114.90*	0.82**	0.83**	2384.53***	2079.64**	23019.13***	18157.76*
	(0.14)	(0.16)	(316.81)	(369.68)	(3145.76)	(3675.27)	(0.36)	(0.42)	(805.82)	(939.32)	(8015.46)	(9345.85)
Diesel fuel	0.01	0.08	965.83***	1119.63***	11448.98***	12181.88***	0.02	0.19	2456.47***	2838.49***	29201.28***	30935.22***
	(0.10)	(0.12)	(364.38)	(425.64)	(3736.33)	(4345.89)	(0.26)	(0.30)	(918.13)	(1071.78)	(9445.32)	(10967.65)
Fish feed	0.19	0.41**	1285.41***	1765.71***	13923.11***	17837.65***	0.47	1.04**	3259.76***	4470.14***	35404.77***	45196.44***
	(0.17)	(0.20)	(402.46)	(468.65)	(3925.80)	(4575.13)	(0.43)	(0.50)	(1014.42)	(1191.88)	(9915.82)	(11636.19)
Fishing rod	0.12	0.28	1369.77***	1419.59***	12925.72***	12641.09***	0.31	0.72	3478.28***	3602.05***	32903.40***	32112.40***
	(0.57)	(0.66)	(342.16)	(396.80)	(2805.04)	(3253.71)	(1.42)	(1.65)	(867.78)	(1004.71)	(7152.67)	(8269.45)
Fruit tree	3.29	4.73*	1670.34***	1714.10***	12025.14***	11537.59***	8.29	11.90*	4243.20***	4350.61***	30649.14***	29339.11***
	(2.40)	(2.73)	(381.43)	(438.54)	(2524.60)	(2909.50)	(6.02)	(6.85)	(970.91)	(1114.46)	(6464.49)	(7418.33)
Soil	-0.50**	-0.67***	828.79***	575.80*	9237.15***	6244.98**	-1.27**	-1.68***	2117.15***	1477.02*	23646.47***	16024.94**
	(0.22)	(0.25)	(273.60)	(316.33)	(2464.25)	(2853.33)	(0.56)	(0.65)	(697.31)	(806.47)	(6320.36)	(7298.48)
N	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500

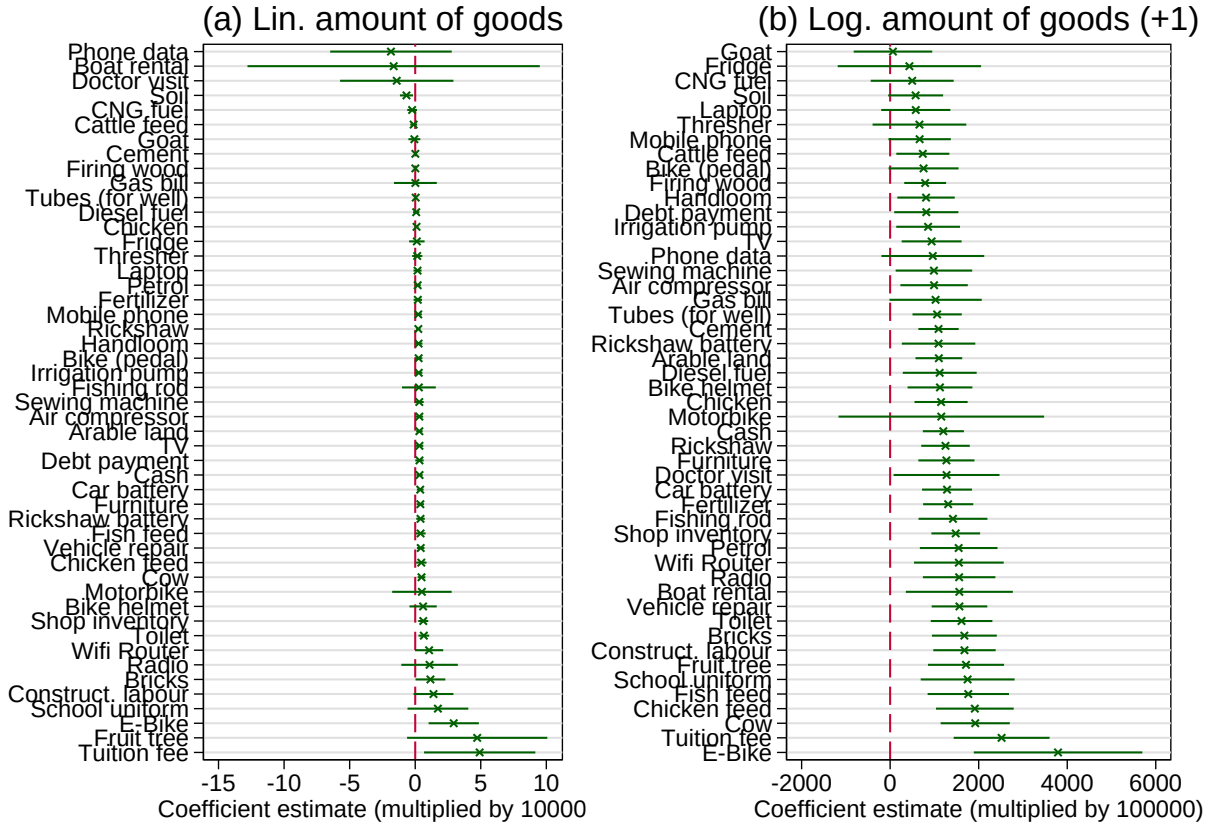
Note: Table shown above represents part 1 out of 2 – please see next page for notes.

Table 9: Avg. bundle preference probability effect of good presence (price-adjusted) *(2/2)*

Good Type	**Part A** – Linear regression						**Part B** – Probit (marg. effects shown)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
main												
Fridge	0.12	0.13	522.62	435.30	5874.29	4273.04	0.30	0.33	1343.38	1122.00	15173.70	11065.87
	(0.25)	(0.30)	(690.12)	(825.94)	(7103.89)	(8504.56)	(0.64)	(0.76)	(1750.63)	(2092.22)	(18029.88)	(21549.25)
Gas bill	0.36	0.01	1390.51***	1026.78*	13153.56***	9356.70**	0.91	0.06	3531.16***	2615.19*	33483.96***	23860.07**
	(0.74)	(0.83)	(463.29)	(531.85)	(3822.50)	(4400.06)	(1.84)	(2.09)	(1166.15)	(1339.54)	(9652.69)	(11105.23)
Arable land	0.31***	0.32**	1153.27***	1100.39***	11787.12***	10609.30***	0.78***	0.81**	2930.70***	2795.91***	30034.18***	26998.42***
	(0.11)	(0.12)	(233.40)	(269.46)	(2266.94)	(2619.82)	(0.28)	(0.32)	(597.23)	(687.91)	(5823.94)	(6699.61)
Toilet	0.58***	0.66***	1506.95***	1612.86***	15004.18***	15466.05***	1.48***	1.67***	3823.84***	4090.77***	38154.58***	39266.83***
	(0.18)	(0.21)	(300.18)	(355.03)	(2902.41)	(3439.55)	(0.46)	(0.54)	(764.54)	(903.65)	(7406.71)	(8761.24)
Mobile phone	0.31***	0.23*	1084.19***	666.21*	10975.28***	5988.78*	0.79**	0.60*	2755.14***	1697.91*	27976.48***	15317.66*
	(0.12)	(0.14)	(306.32)	(359.34)	(3061.86)	(3591.48)	(0.31)	(0.36)	(777.28)	(913.84)	(7786.61)	(9144.75)
Phone data	-4.15**	-1.85	664.87	964.26	6900.74*	8023.50*	-10.53**	-4.65	1708.73	2462.95*	17753.25*	20515.81*
	(1.99)	(2.36)	(511.16)	(592.22)	(3708.88)	(4288.19)	(5.13)	(5.97)	(1297.68)	(1493.45)	(9431.94)	(10832.57)
Motorbike	0.94	0.51	1818.88*	1155.93	17930.60**	10449.17	2.40	1.28	4634.31*	2922.78	45845.93**	26499.43
	(0.98)	(1.16)	(1001.65)	(1184.44)	(9145.20)	(10818.31)	(2.51)	(3.03)	(2551.20)	(3098.10)	(23323.16)	(28325.33)
Bike helmet	0.44	0.60	1188.73***	1125.78***	11113.27***	9722.01***	1.13	1.52	3023.56***	2863.85***	28356.91***	24782.80***
	(0.45)	(0.53)	(320.77)	(373.29)	(2679.48)	(3117.48)	(1.13)	(1.33)	(814.27)	(945.47)	(6837.48)	(7921.01)
Air compressor	0.26**	0.32**	933.18***	992.70**	9805.36***	9713.21**	0.67**	0.80**	2376.25***	2523.69**	25055.55***	24740.86**
	(0.13)	(0.15)	(337.31)	(388.45)	(3475.15)	(3999.65)	(0.33)	(0.38)	(853.88)	(981.27)	(8809.03)	(10108.45)
Car battery	0.38***	0.38***	1342.00***	1284.95***	13749.86***	12531.26***	0.95***	0.97***	3405.56***	3260.19***	34975.84***	31834.37***
	(0.11)	(0.13)	(248.11)	(288.14)	(2419.64)	(2811.37)	(0.29)	(0.34)	(634.54)	(735.00)	(6210.76)	(7182.31)
Rickshaw battery	0.42***	0.41**	1239.79***	1093.71***	12838.39***	10639.52**	1.07***	1.03**	3148.55***	2778.56***	32704.61***	27084.24**
	(0.14)	(0.17)	(347.37)	(423.21)	(3479.64)	(4239.40)	(0.36)	(0.43)	(878.79)	(1067.08)	(8823.74)	(10700.71)
Bike (pedal)	0.31**	0.26*	1035.24***	755.37*	10254.02***	6749.96*	0.80**	0.67*	2638.13***	1926.64*	26235.16***	17276.38*
	(0.12)	(0.14)	(342.43)	(403.03)	(3461.76)	(4072.82)	(0.32)	(0.37)	(867.54)	(1023.64)	(8784.27)	(10352.61)
TV	0.33**	0.32**	1046.76***	937.19***	11070.82***	9241.62***	0.83**	0.82**	2661.15***	2383.72***	28236.34***	23554.01***
	(0.13)	(0.15)	(299.64)	(345.18)	(3011.88)	(3474.76)	(0.33)	(0.38)	(760.21)	(874.89)	(7664.95)	(8821.96)
Radio	0.90	1.10	1633.49***	1558.76***	14098.25***	12676.03***	2.26	2.75	4141.74***	3951.90***	35841.61***	32186.36***
	(0.95)	(1.10)	(361.35)	(417.57)	(2764.60)	(3194.10)	(2.38)	(2.75)	(917.23)	(1057.59)	(7056.55)	(8117.04)
Wifi Router	0.55	1.06*	1318.02***	1552.02***	12908.47***	13913.82***	1.38	2.68*	3344.94***	3933.03***	32857.40***	35316.11***
	(0.46)	(0.55)	(438.75)	(516.73)	(3861.53)	(4553.12)	(1.17)	(1.39)	(1105.95)	(1305.00)	(9760.84)	(11507.63)
Cement	0.01***	0.01**	1175.43***	1093.09***	16477.45***	14749.12***	0.03***	0.03**	2977.69***	2769.12***	41837.19***	37406.91***
	(0.00)	(0.01)	(198.73)	(231.66)	(2589.85)	(3027.19)	(0.01)	(0.01)	(506.87)	(589.05)	(6639.79)	(7719.78)
Furniture	0.32***	0.40***	1210.24***	1270.71***	12835.76***	12543.22***	0.81**	1.01***	3075.35***	3226.00***	32709.45***	31893.61***
	(0.12)	(0.14)	(279.03)	(323.02)	(2752.40)	(3187.70)	(0.31)	(0.36)	(710.75)	(821.39)	(7040.05)	(8121.51)
Handloom	0.32**	0.26*	1104.19***	810.83**	11724.85***	8056.84**	0.81**	0.66*	2807.66***	2070.06**	29900.00***	20614.69**
	(0.13)	(0.15)	(287.20)	(330.95)	(2867.35)	(3308.47)	(0.33)	(0.38)	(728.34)	(839.05)	(7294.94)	(8401.24)
Thresher	0.18	0.15	780.27*	661.80	8520.45*	6698.39	0.47	0.39	1981.60*	1676.23	21720.23*	17016.14
	(0.17)	(0.20)	(453.14)	(540.17)	(4526.35)	(5391.43)	(0.43)	(0.51)	(1151.23)	(1371.41)	(11509.39)	(13681.72)
Sewing machine	0.22	0.31*	755.31**	989.51**	8200.24**	9927.80**	0.56	0.79*	1930.89**	2516.12**	21040.95**	25283.98**
	(0.14)	(0.17)	(365.19)	(440.90)	(3667.66)	(4432.79)	(0.37)	(0.44)	(924.79)	(1112.81)	(9306.99)	(11193.67)
Petrol	0.07	0.19	1147.78***	1548.29***	13246.13***	16540.50***	0.17	0.47	2912.95***	3918.43***	33708.93***	41910.09***
	(0.10)	(0.12)	(373.41)	(447.25)	(3822.69)	(4588.41)	(0.25)	(0.30)	(939.94)	(1131.44)	(9650.42)	(11618.54)
Firing wood	0.01	0.01	824.09***	790.88***	11528.37***	10272.93***	0.02	0.03	2094.79***	2009.78***	29393.02***	26154.87***
	(0.01)	(0.01)	(209.91)	(240.90)	(2621.33)	(3012.71)	(0.02)	(0.02)	(533.53)	(611.06)	(6701.94)	(7669.97)
Doctor visit	-1.63	-1.41	1304.10**	1274.41**	11606.29***	10594.35**	-4.11	-3.55	3317.84**	3239.80**	29590.86***	26957.79**
	(1.89)	(2.20)	(520.80)	(609.89)	(3813.94)	(4466.32)	(4.77)	(5.56)	(1312.33)	(1534.91)	(9644.72)	(11266.42)
Tubes (for well)	0.03***	0.04***	1067.75***	1061.31***	13325.17***	12650.22***	0.09***	0.09***	2709.34***	2691.81***	33904.23***	32128.04***
	(0.01)	(0.01)	(241.51)	(284.16)	(2946.40)	(3473.16)	(0.03)	(0.04)	(614.00)	(721.02)	(7509.56)	(8820.58)
Shop inventory	0.57***	0.61***	1465.84***	1481.71***	14580.49***	14118.32***	1.46***	1.56***	3717.54***	3756.03***	37056.73***	35820.40***
	(0.17)	(0.19)	(241.92)	(280.49)	(2256.33)	(2617.63)	(0.43)	(0.50)	(619.80)	(716.97)	(5805.61)	(6702.89)
N	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500
FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
GoodsVar	Lin.	Lin.	Log.(x+1)	Log.(x+1)	I(x>0)	I(x>0)	Lin.	Lin.	Log.(x+1)	Log.(x+1)	I(x>0)	I(x>0)

Note: Average impact of presence of good – **divided by 100,000** for more accessible display of results – (or log. transformation thereof) in bundle on bundle preference probability. SEs clustered at the level of household ID. All good quantities multiplied by respective price per unit to adjust for differences in affordability between goods. Stars for confidence: 0.1 *, 0.05 **, 0.01 ***

Figure 10: Avg. bundle preference probability effect of good presence



Note: Regression coefficients and associated 95% confidence intervals. Sub-figure (a) corresponds to regression specification show in Column (2) in Tables 8 and 9. Sub-figure (b) corresponds to regression specification show in Column (4) in Tables 8 and 9.

Table 10: Avg. bundle preference probability effect of good presence (dropping WARP non-compliant agents, not price-adjusted) *(1/2)*

Good Type	**Part A** – Linear regression						**Part B** – Probit (marg. effects shown)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
main												
Cash	0.00***	0.00***	0.87***	0.97***	13.06***	11.94***	0.00***	0.00***	2.19***	2.45***	33.22***	30.33***
	(0.00)	(0.00)	(0.19)	(0.22)	(1.96)	(2.27)	(0.00)	(0.00)	(0.49)	(0.57)	(5.08)	(5.86)
Boat rental	0.13	-0.25	4.00**	3.15	14.75***	11.08***	0.33	-0.63	10.07**	7.95	37.50***	28.16***
	(0.72)	(0.85)	(2.02)	(2.37)	(3.30)	(3.85)	(1.81)	(2.15)	(5.07)	(5.97)	(8.38)	(9.78)
Bricks	0.10**	0.12**	2.41***	2.82***	14.96***	14.27***	0.25**	0.29**	6.07***	7.10***	38.02***	36.20***
	(0.05)	(0.06)	(0.65)	(0.76)	(2.69)	(3.12)	(0.12)	(0.15)	(1.64)	(1.91)	(6.88)	(7.96)
Fertilizer	0.03	0.06	1.87***	2.23***	13.89***	13.08***	0.08	0.15	4.71***	5.63***	35.32***	33.20***
	(0.04)	(0.05)	(0.56)	(0.65)	(2.35)	(2.73)	(0.11)	(0.12)	(1.42)	(1.63)	(6.04)	(6.96)
Irrigation pump	4.01**	4.07*	4.43*	5.11*	9.34***	8.22**	10.20**	10.34*	11.15*	12.90*	23.84***	20.93**
	(1.84)	(2.14)	(2.51)	(2.92)	(3.19)	(3.74)	(4.75)	(5.52)	(6.31)	(7.36)	(8.13)	(9.48)
Construct. labour	0.64	0.98*	4.81***	5.90***	14.23***	13.75***	1.60	2.45*	12.10***	14.87***	36.17***	34.89***
	(0.48)	(0.54)	(1.45)	(1.65)	(2.45)	(2.83)	(1.19)	(1.36)	(3.63)	(4.15)	(6.28)	(7.21)
Chicken feed	0.23***	0.23**	3.74***	4.15***	19.51***	19.38***	0.59***	0.59**	9.45***	10.50***	49.53***	49.15***
	(0.09)	(0.10)	(0.99)	(1.18)	(3.64)	(4.36)	(0.22)	(0.26)	(2.52)	(3.01)	(9.31)	(11.13)
Cattle feed	-0.03	-0.05	1.03	0.90	10.29***	7.95***	-0.07	-0.13	2.60	2.28	26.31***	20.32***
	(0.05)	(0.06)	(0.64)	(0.75)	(2.54)	(2.96)	(0.13)	(0.15)	(1.61)	(1.87)	(6.50)	(7.56)
Chicken	0.05	0.10	1.83**	2.34***	13.15***	12.55***	0.12	0.26	4.62**	5.91***	33.45***	31.88***
	(0.06)	(0.07)	(0.73)	(0.84)	(2.73)	(3.18)	(0.15)	(0.18)	(1.82)	(2.11)	(6.97)	(8.08)
Goat	-0.47	-0.29	-1.63	-1.14	2.58	0.70	-1.19	-0.73	-4.19	-2.97	6.71	1.81
	(0.89)	(1.04)	(2.28)	(2.64)	(3.79)	(4.37)	(2.32)	(2.68)	(5.85)	(6.77)	(9.79)	(11.31)
Cow	12.11***	12.03***	19.01***	19.31***	22.93***	19.77***	30.65***	30.47***	47.88***	48.69***	58.33***	50.18***
	(3.23)	(3.70)	(4.18)	(4.77)	(3.51)	(4.07)	(8.30)	(9.52)	(10.57)	(12.02)	(9.10)	(10.43)
School uniform	0.86	1.38	4.98**	6.47**	14.43***	14.57***	2.18	3.50	12.53**	16.32**	36.66***	36.95***
	(0.81)	(0.95)	(2.26)	(2.65)	(3.68)	(4.30)	(2.04)	(2.39)	(5.69)	(6.69)	(9.33)	(10.88)
Tuition fee	1.34	2.21**	6.76***	9.48***	17.20***	18.79***	3.38	5.61**	17.03***	24.00***	43.67***	47.68***
	(0.84)	(0.97)	(2.28)	(2.65)	(3.38)	(3.96)	(2.12)	(2.51)	(5.76)	(6.81)	(8.62)	(10.14)
CNG fuel	-0.19*	-0.15	-0.15	0.10	7.56*	6.04	-0.48*	-0.38	-0.38	0.28	19.43*	15.52
	(0.10)	(0.12)	(1.13)	(1.34)	(3.96)	(4.69)	(0.27)	(0.32)	(2.86)	(3.38)	(10.07)	(11.90)
E-Bike	0.00**	0.00***	2.17***	3.13***	30.08***	34.73***	0.01**	0.01***	5.46***	7.93***	76.43***	88.30***
	(0.00)	(0.00)	(0.83)	(0.93)	(7.85)	(8.86)	(0.00)	(0.00)	(2.09)	(2.42)	(19.92)	(23.07)
Vehicle repair	0.00***	0.00***	1.07***	1.25***	16.23***	16.00***	0.00***	0.00***	2.69***	3.15***	41.22***	40.57***
	(0.00)	(0.00)	(0.27)	(0.30)	(2.77)	(3.18)	(0.00)	(0.00)	(0.67)	(0.77)	(7.09)	(8.10)
Rickshaw	0.00***	0.00***	0.74***	0.95***	13.14***	13.22***	0.00***	0.00***	1.87***	2.38***	33.44***	33.57***
	(0.00)	(0.00)	(0.22)	(0.26)	(2.52)	(2.89)	(0.00)	(0.00)	(0.56)	(0.64)	(6.45)	(7.37)
Laptop	6.77	6.28	4.67	4.78	7.78**	5.58	17.31	16.06	11.76	12.06	19.88**	14.26
	(4.72)	(5.46)	(5.06)	(5.84)	(3.55)	(4.11)	(12.11)	(14.01)	(12.76)	(14.76)	(9.04)	(10.43)
Debt payment	0.00**	0.00**	0.58*	0.63*	9.01***	7.11*	0.00**	0.00**	1.46*	1.58*	23.02***	18.16*
	(0.00)	(0.00)	(0.31)	(0.37)	(3.15)	(3.68)	(0.00)	(0.00)	(0.79)	(0.92)	(8.02)	(9.35)
Diesel fuel	0.01	0.06	1.32	2.12*	11.45***	12.18***	0.01	0.15	3.33	5.34*	29.20***	30.94***
	(0.08)	(0.10)	(1.00)	(1.18)	(3.74)	(4.35)	(0.20)	(0.24)	(2.50)	(2.95)	(9.45)	(10.97)
Fish feed	0.10	0.23**	2.21**	3.88***	13.92***	17.84***	0.26	0.57**	5.56**	9.79***	35.40***	45.20***
	(0.09)	(0.11)	(1.08)	(1.26)	(3.93)	(4.58)	(0.24)	(0.28)	(2.72)	(3.20)	(9.92)	(11.64)
Fishing rod	0.12	0.28	3.22*	4.14**	12.93***	12.64***	0.31	0.72	8.11**	10.45**	32.90***	32.11***
	(0.57)	(0.66)	(1.65)	(1.91)	(2.81)	(3.25)	(1.42)	(1.65)	(4.13)	(4.78)	(7.15)	(8.27)
Fruit tree	0.66	0.95*	4.33***	5.31***	12.03***	11.54***	1.66	2.38*	10.91***	13.40***	30.65***	29.34***
	(0.48)	(0.55)	(1.44)	(1.64)	(2.52)	(2.91)	(1.20)	(1.37)	(3.62)	(4.13)	(6.46)	(7.42)
Soil	-0.10**	-0.13***	0.49	0.20	9.24***	6.24**	-0.25**	-0.34***	1.24	0.53	23.65***	16.02**
	(0.04)	(0.05)	(0.60)	(0.69)	(2.46)	(2.85)	(0.11)	(0.13)	(1.50)	(1.73)	(6.32)	(7.30)
N	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500

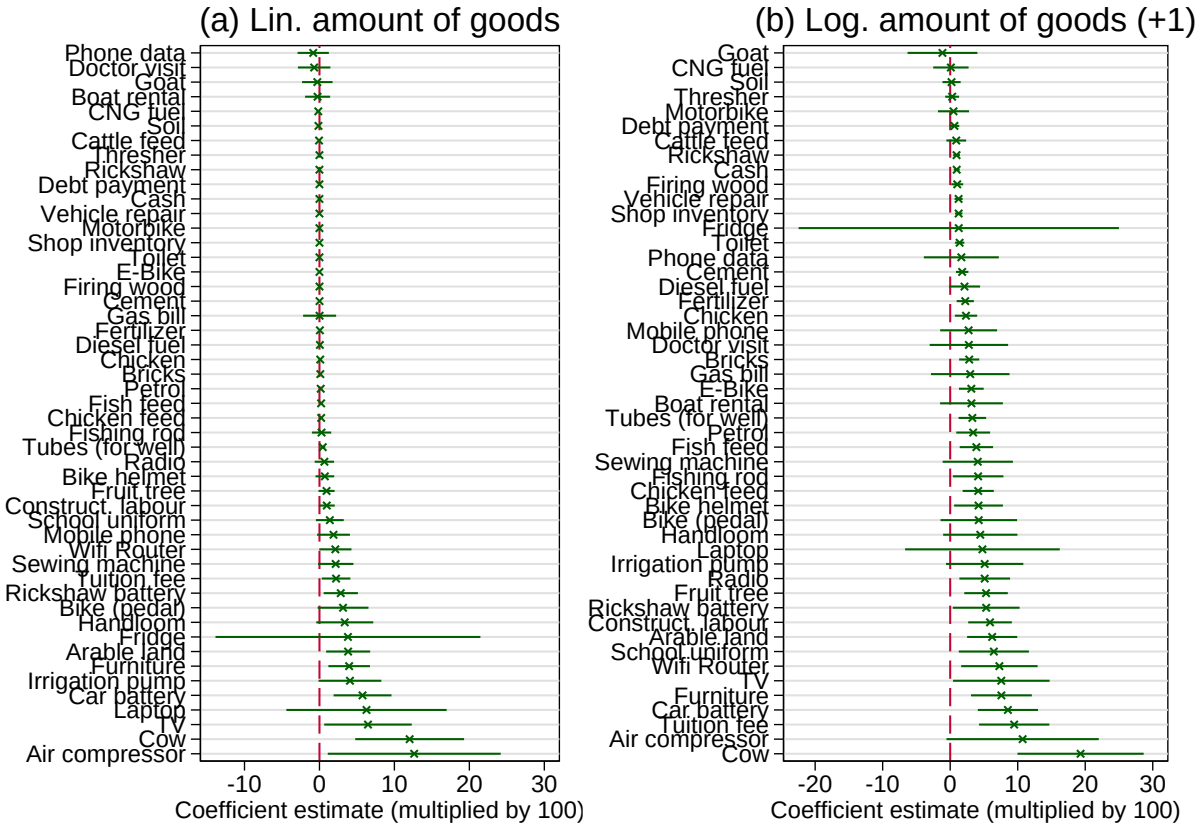
Note: Table shown above represents part 1 out of 2 – please see next page for notes.

Table 11: Avg. bundle preference probability effect of good presence (dropping WARP non-compliant agents, not price-adjusted) *(2/2)*

Good Type	**Part A** – Linear regression						**Part B** – Probit (marg. effects shown)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
main												
Fridge	3.52	3.80	0.65	1.27	5.87	4.27	9.10	9.82	1.68	3.29	15.17	11.07
	(7.54)	(9.01)	(10.12)	(12.10)	(7.10)	(8.50)	(19.18)	(22.88)	(25.64)	(30.63)	(18.03)	(21.55)
Gas bill	0.48	0.02	3.99	2.97	13.15***	9.36**	1.22	0.08	10.06	7.54	33.48***	23.86**
	(0.99)	(1.12)	(2.61)	(2.97)	(3.82)	(4.40)	(2.48)	(2.82)	(6.52)	(7.43)	(9.65)	(11.11)
Arable land	3.67***	3.82**	5.46***	6.22***	11.79***	10.61***	9.33***	9.72**	13.76***	15.70***	30.03***	27.00***
	(1.29)	(1.50)	(1.64)	(1.89)	(2.27)	(2.62)	(3.35)	(3.89)	(4.13)	(4.77)	(5.82)	(6.70)
Toilet	0.00***	0.00***	1.13***	1.40***	15.00***	15.47***	0.00***	0.00***	2.85***	3.53***	38.15***	39.27***
	(0.00)	(0.00)	(0.30)	(0.36)	(2.90)	(3.44)	(0.00)	(0.00)	(0.76)	(0.90)	(7.41)	(8.76)
Mobile phone	2.49***	1.87*	4.32**	2.73	10.98***	5.99*	6.33**	4.78*	10.88**	6.90	27.98***	15.32*
	(0.96)	(1.12)	(1.83)	(2.15)	(3.06)	(3.59)	(2.47)	(2.89)	(4.59)	(5.42)	(7.79)	(9.14)
Phone data	-1.87**	-0.83	-1.05	1.66	6.90*	8.02*	-4.74**	-2.09	-2.65	4.21	17.75*	20.52*
	(0.89)	(1.06)	(2.42)	(2.83)	(3.71)	(4.29)	(2.31)	(2.69)	(6.15)	(7.12)	(9.43)	(10.83)
Motorbike	0.00	0.00	0.85	0.49	17.93**	10.45	0.00	0.00	2.13	1.21	45.85**	26.50
	(0.00)	(0.00)	(0.99)	(1.17)	(9.15)	(10.82)	(0.00)	(0.00)	(2.52)	(3.06)	(23.32)	(28.33)
Bike helmet	0.53	0.72	3.51**	4.20**	11.11***	9.72***	1.35	1.83	8.85**	10.60**	28.36***	24.78***
	(0.54)	(0.64)	(1.58)	(1.84)	(2.68)	(3.12)	(1.36)	(1.59)	(3.97)	(4.62)	(6.84)	(7.92)
Air compressor	10.59**	12.64**	7.57	10.73*	9.81***	9.71**	26.97**	32.13**	19.09	27.09*	25.06***	24.74**
	(5.09)	(5.89)	(5.01)	(5.75)	(3.48)	(4.00)	(13.09)	(15.11)	(12.57)	(14.45)	(8.81)	(10.11)
Car battery	5.63***	5.75***	7.79***	8.55***	13.75***	12.53***	14.29***	14.59***	19.63***	21.58***	34.98***	31.83***
	(1.69)	(1.97)	(1.96)	(2.27)	(2.42)	(2.81)	(4.39)	(5.11)	(4.94)	(5.73)	(6.21)	(7.18)
Rickshaw battery	2.94***	2.84**	5.30***	5.33**	12.84***	10.64**	7.46***	7.21**	13.35***	13.45**	32.70***	27.08**
	(0.97)	(1.16)	(2.05)	(2.53)	(3.48)	(4.24)	(2.49)	(2.98)	(5.16)	(6.34)	(8.82)	(10.70)
Bike (pedal)	3.76**	3.15*	5.33**	4.24	10.25***	6.75*	9.57**	8.02*	13.45**	10.70	26.24***	17.28*
	(1.48)	(1.73)	(2.45)	(2.89)	(3.46)	(4.07)	(3.80)	(4.45)	(6.16)	(7.30)	(8.78)	(10.35)
TV	6.55**	6.48**	7.28**	7.57**	11.07***	9.24***	16.64**	16.45**	18.34**	19.11**	28.24***	23.55***
	(2.58)	(2.98)	(3.19)	(3.65)	(3.01)	(3.47)	(6.64)	(7.65)	(8.02)	(9.16)	(7.66)	(8.82)
Radio	0.54	0.66	4.57***	5.11***	14.10***	12.68***	1.36	1.65	11.50***	12.89***	35.84***	32.19***
	(0.57)	(0.66)	(1.66)	(1.92)	(2.76)	(3.19)	(1.43)	(1.65)	(4.17)	(4.81)	(7.06)	(8.12)
Wifi Router	1.09	2.12*	4.55*	7.29**	12.91***	13.91***	2.76	5.36*	11.46*	18.37**	32.86***	35.32***
	(0.93)	(1.10)	(2.45)	(2.89)	(3.86)	(4.55)	(2.33)	(2.77)	(6.14)	(7.28)	(9.76)	(11.51)
Cement	0.02***	0.01**	1.76***	1.78***	16.48***	14.75***	0.04***	0.04**	4.42***	4.48***	41.84***	37.41***
	(0.01)	(0.01)	(0.40)	(0.47)	(2.59)	(3.03)	(0.01)	(0.02)	(1.02)	(1.18)	(6.64)	(7.72)
Furniture	3.18***	3.96***	5.57***	7.59***	12.84***	12.54***	8.07**	10.05***	14.06***	19.16***	32.71***	31.89***
	(1.22)	(1.41)	(1.99)	(2.30)	(2.75)	(3.19)	(3.15)	(3.64)	(5.01)	(5.79)	(7.04)	(8.12)
Handloom	4.16**	3.37*	5.85**	4.46	11.72***	8.06**	10.56**	8.60*	14.76**	11.30	29.90***	20.61**
	(1.68)	(1.94)	(2.46)	(2.81)	(2.87)	(3.31)	(4.32)	(4.98)	(6.16)	(7.05)	(7.29)	(8.40)
Thresher	0.00	0.00	0.28	0.29	8.52*	6.70	0.00	0.00	0.69	0.73	21.72*	17.02
	(0.00)	(0.00)	(0.44)	(0.52)	(4.53)	(5.39)	(0.00)	(0.00)	(1.12)	(1.33)	(11.51)	(13.68)
Sewing machine	1.54	2.17*	1.89	4.09	8.20**	9.93**	3.95	5.52*	4.78	10.33	21.04**	25.28**
	(1.01)	(1.20)	(2.20)	(2.65)	(3.67)	(4.43)	(2.59)	(3.08)	(5.53)	(6.67)	(9.31)	(11.19)
Petrol	0.06	0.17	1.88*	3.41***	13.25***	16.54***	0.16	0.42	4.75*	8.58***	33.71***	41.91***
	(0.09)	(0.11)	(1.07)	(1.28)	(3.82)	(4.59)	(0.22)	(0.27)	(2.67)	(3.22)	(9.65)	(11.62)
Firing wood	0.00	0.01	0.88**	1.04**	11.53***	10.27***	0.01	0.02	2.23**	2.63**	29.39***	26.15***
	(0.00)	(0.01)	(0.39)	(0.45)	(2.62)	(3.01)	(0.01)	(0.01)	(0.99)	(1.13)	(6.70)	(7.67)
Doctor visit	-0.82	-0.70	2.08	2.77	11.61***	10.59**	-2.05	-1.77	5.27	7.01	29.59***	26.96**
	(0.94)	(1.10)	(2.53)	(2.96)	(3.81)	(4.47)	(2.39)	(2.78)	(6.34)	(7.42)	(9.64)	(11.27)
Tubes (for well)	0.42***	0.45***	2.85***	3.28***	13.33***	12.65***	1.06***	1.13***	7.17***	8.29***	33.90***	32.13***
	(0.14)	(0.16)	(0.88)	(1.04)	(2.95)	(3.47)	(0.36)	(0.42)	(2.21)	(2.62)	(7.51)	(8.82)
Shop inventory	0.00***	0.00***	1.08***	1.26***	14.58***	14.12***	0.00***	0.00***	2.73***	3.18***	37.06***	35.82***
	(0.00)	(0.00)	(0.24)	(0.28)	(2.26)	(2.62)	(0.00)	(0.00)	(0.61)	(0.71)	(5.81)	(6.70)
N	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500	18060	13500
FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
GoodsVar	Lin.	Lin.	Log.(x+1)	Log.(x+1)	I(x>0)	I(x>0)	Lin.	Lin.	Log.(x+1)	Log.(x+1)	I(x>0)	I(x>0)

Note: Average impact of presence of good – **divided by 100** for more accessible display of results – (or log. transformation thereof) in bundle on bundle preference probability. Standard errors clustered at the level of household ID. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

Figure 11: Avg. bundle preference probability effect of good presence



Note: Dropping WARP non-compliant agents, not price-adjusted. Regression coefficients and associated 95% confidence intervals. Sub-figure (a) corresponds to regression specification show in Column (2) in Tables 2 and 3. Sub-figure (b) corresponds to regression specification show in Column (4) in Tables 2 and 3.

Chapter II
Utilising Gaussian Processes for Consumer Preference Elicitation and Modelling

Abstract

Elicitation of consumer preferences is of significant importance in many applied microeconomics settings but it is challenging to execute empirically without relying on strong assumptions or incurring prohibitive surveying cost. In this chapter, I present an alternative method, driven by Gaussian Processes, that allows preferences elicitation across a high-dimensional universe of 51 goods. First, I show estimation approach on a illustrative simulation – I also showcase performance in comparison to conventional MLE approaches. Then, I present estimation outcomes based on empirical data collected through preference data of 500 Bangladeshi respondents. I conclude by highlighting familiar Cobb-Douglas preference patterns which this method was able to reconstruct from survey data alone.

1 Introduction

Consumer preference elicitation is a principal ingredient to the academic study of decision-making processes and market behaviour, offering valuable insights for both theoretical and applied microeconomics. Above and beyond its time-tested impact, the literature emphasising the significance of preference elicitation in guiding economic choices continues to grow (Clithero et al. (2023)). Specifically in the era of evidence-based policy-making, the ability to accurately assess consumer preferences is a key input for the design of markets (Arrow et al. (2010)) or ascertain the nature of biases departing from conventional rational-choice theory (Kahneman and Tversky (1974), Kahneman and Tversky (1979)). Especially in evidence-based development economics, consumer preference data often constitutes important precursor analysis for a subsequent assessment of the impact that a specific policy or mechanism may be generating. For example, Banerjee and Duflo (2020) argue that understanding consumer preferences is crucial for effective policy design, particularly in poverty alleviation. In their contribution, they use preference data from microfinance programs to demonstrate how tailoring financial services to local preferences can improve welfare outcomes. Similarly, Karlan et al. (2016) explore microfinance's impact by eliciting preferences regarding credit access, revealing substantial heterogeneity in treatment effects based on consumer preferences – highlighting the importance of having data on individual preferences when assessing policy or intervention impact. Similarly, Angelucci et al. (2018) use preference data in the context of conditional cash transfers to analyse how recipient preferences for specific elements of a policy (such as conditionality) affect household welfare, demonstrating the importance of having a robust idea of participants' preferences before designing interventions. Chetty et al. (2014) use preference analysis to examine how low-income individuals

value educational opportunities, providing valuable insights for policy design that aims to enhance human capital development. Bordering on the institutions literature, Khan et al. (2022) further show the need to measure individual preferences to design mechanisms geared to public goods creation.

Given the importance of preference data for empirical development economics, there is a surprisingly thin range of methods available to elicit such data. If market data is available, revealed preference methods are a well-documented pathway for estimation (Varian (2014)). However, in the context of many other situations of interest for applied microeconomics, the availability of reliable market data is a rare occurrence. Thus, economists must frequently rely on the study of stated preference data. The principle methodological concept for doing so are choice experiments: participants are presented with a set of hypothetical alternatives, each with varying attributes. They are asked to choose between these alternatives, and the trade-offs between the attributes allow researchers to estimate how much consumers are willing to pay for specific goods and their attributes. Such methods have enjoyed particular popularity in fields such as development economics, environmental economics, health economics, and public policy (Carlsson (2010), Loomis (2011), Hensher and Johnson (2018)). Going one step further, Cunningham et al. (2010) also propose a medical application of adaptive choice-based conjoint analysis that features choices designed dynamically based on a respondent's previous responses, leveraging strong assumptions on both the underlying preference structure as well as the statistical distribution of measurement error.[20] Despite these developments, the conventional analysis of stated-preference choice experiment continues to be a staple of applied

[20]As outlined by Hossain and List (2012), there is a particular endogeneity challenge related to hybrid models, which may be challenging many results generated by such hybrid methods.

microeconomics. Building on data generated by choice experiments, discrete choice modelling is commonly used to estimate market demand, consumer welfare, and predict the impact of policy interventions (McFadden and Train (2000), Train (2009)). Note that such estimation approaches are almost exclusively based on strong functional assumptions about preference structure and error distributions.

There are further methodological challenges associated with the empirical implementation of consumer preference elicitation methods in applied microeconomics. One of the most persistent issues is that hypothetical – non-incentivised – choices lend themselves to responses not fully in line with participants' economic preferences. This can lead to a biased estimation of consumers' willingness to pay (Carson and Groves (2007), Hensher (2010)). A series of articles has proposed survey design mechanisms created to mitigate such bias – for example the introduction of certainty scales – but these changes generally incur the cost of additional complexity (Ledyard et al. (1994), List (2006)). The additional complexity incurred by these advanced methods can, in turn, affect respondents' decision-making and lead them to ignore or misinterpret some attributes of presented choices – for example due to cognitive overload. The resulting bias (termed attribute non-attendance in the literature) can lead to distorted preference estimates (Hensher and Johnson (2018)). An emerging literature has begun developing advanced estimators to detect and correct for this phenomenon, emphasising the need for survey design with a careful focus on possible attention-driven effect (Alemu et al. (2013)). Beyond the concerns specific to the elicitation of stated preferences, the general concerns related to the external validity of survey-based data apply as well. While there are proposed methods for weight and sampling techniques to correct for imbalance resulting from non-responses, that points to the comparative advantage of lab or lab-in-the-field experiments for prefer-

ence elicitation – despite possible challenges related to result generalisation (Viceisza (2016)).

The increasing penetration of applied microeconomics by machine-learning techniques has inspired many methodological advances (Mullainathan and Spiess (2017), Athey et al. (2019), Baiardi and Naghi (2024)). I argue that particularly in the application of preference elicitation, additional improvement potential compared to conventional discrete choice modelling is possible – making available techniques less dependent on functional assumptions, upgrading the dimensionality of queried preference structures, and reducing the requirements for costly data generation in the field. To highlight one area of improvement: in many conventional preference elicitation methods, researchers are forced to assume preference homogeneity across participants due to degrees-of-freedom constraints. While analytically elegant, these assumptions are not only not realistic – they furthermore obscure a significant motivation for the study of preferences in the first place: uncovering differences between agents. Machine learning methods – particularly Bayesian latent class models and clustering algorithms – can solve this challenge by allowing individual-level study of preference data. These techniques both detect and adaptively model different segments of agents with similar preferences, even in large and complex datasets, thus providing much more nuanced insights into consumer behaviour (Bujosa et al. (2010)). T. Chen and Guestrin (2016) provide an overview of early advances made by introducing random forests, gradient boosting, and neural networks to the study of very large datasets with a high number of variables and complex interactions between them. Improvements in estimation capability have been significant: for instance, the random-forest algorithm *XGBoost* has been shown to significantly outperform traditional econometric models in terms of predictive accuracy in preference

elicitation tasks, particularly when the data include non-linearities and interactions among attributes that are difficult to specify in conventional models (Borrellas and Unceta (2021)). The ability to handle large, high-dimensional datasets with numerous attributes and choices alternatives has thusly opened up previously underutilised datasets to economic analysis.

Recent economic research has applied both the econometric capabilities of machine-learning methods, as well as, the probabilistic view of agent choice to consumer preference elicitation. The underlying "engine" of discrete-choice modelling still continues to be the focal point of these otherwise improved methods. In this chapter, I propose the use of Gaussian Processes as a way to take the concept of Bayesian statistics in preference measurement to its proximate conclusion. Gaussian Processes are a powerful machine-learning concept for non-parametric regression and classification tasks – bearing significant advantages in contexts where the underlying relationship between variables is unknown or complex as well as high dimensional. For the purpose of preference elicitation, Gaussian Processes are of particular value due to their ability to model non-linear preferences for goods without relying on functional form assumptions. While conventional discrete choice models frequently include explicit and hard-to-verify assumptions of i) agents' utility function, and ii) distribution of error terms, Gaussian Processes agnosticly view relationships between input variables – for example stated or expressed agent preferences – as the realisation of a latent probabilistic expression of a random function (Rasmussen and Williams (2006)). Since the incremental estimation step of a Gaussian Process – conditional updating of a Gaussian function – can be expressed analytically, Gaussian Process training and inference is a frugal employer of computational resources. Next to the effective processing of large and high-dimensional datasets, this also opens up the possibility

of advanced techniques that take into consideration data imperfections – for example sparse input data (Gal et al. (2014)).

Moreover, Gaussian Processes – yielding Gaussian distributions in every iteration rather than naked point estimates – carry a native measure of (un-)certainty. That is, as a Gaussian Process is trained on an increasing observation base, both the (multi-)dimensional posterior as well as a view of that posterior's certainty are simultaneously updated. This feature is particularly interesting for application where uncertainty is a well-described part of the economic framework of analysis – for example, decision-making under uncertainty or the study of biases in decision-making. A prediction made by a Gaussian Process, therefore, carries a dual interpretation. On one hand, posterior uncertainty can be understood in the econometric way – indicating confidence in the estimate, much like confidence intervals in standard estimation toolkits. On the other hand, the posterior can also be viewed as an exact estimate of an inherently probabilistic perspective on agency: the Gaussian Process produces a view on the structure of preference relations as well as the statistical rigidity of that structure. Combining i) a way to describe non-linear functions in a non-parametric way, ii) the ability to incorporate prior knowledge and sequentially uncovered observational data, and iii) a continuous view on estimate certainty even creates the possibility of a self-designing survey, in which each iteration of respondent feedback not only updates the researcher's estimation but also indicates the most high-yield next question to prompt.

In this chapter, I illustrate how Gaussian Processes can be used to generate actionable insight on consumer preference structure based on lab-in-the-field bundle comparison data I recorded with a population of Bangladeshi microfinance entrepreneurs.

In section 2, I demonstrate in a simplified theoretical example a step-by-step guide of how this data can be processed using Gaussian Processes. Furthermore, I explain where conventional discrete choice models would fail to process the same dataset. In section 3, I then provide a view on the bundles each of the 500 individuals would most likely prefer to all other feasible bundles, allocating a given budget to a universe of 51 different goods.

2 Methodology

In the following section, I state a basic microeconomic utility-maximisation problem with observations consisting of binary bundle comparisons. I first present a Maximum Likelihood Estimation (MLE) driven approach to estimating optimal consumption bundles, which rests on conventional methods and assumptions. Subsequently, I propose an alternative way, based on Gaussian Processes, to derive the analogous insights while relying on fewer (parametric) assumptions.

2.1 Problem statement

2.1.1 Estimation target

A cornerstone of microeconomic theory is the claim that constrained consumptions choices by rational agents, rooted in potentially complex preference relations, can be captured by demand functions for the goods making up the consumption choices. This permits the construction of an analytical framework that produces an answer to the question: Which affordable combination of goods is most preferred by a given economic agent?

This Walrasian demand, $\mathbf{x}$, is defined as the function yielding the demanded quantities of each indexed good x_i by an utility-maximising, rational agent subject to a budget constraint. The demand function calls for two exogenous inputs – a price vector, $\mathbf{p}$, and an available budget, w. It furthermore builds on a mathematically tractable expression of the agent's individual preferences, which is typically provided by the direct utility function, u. A standard approach to the utility maximising problem through Lagrangian will yield the following First Order Conditions

for optimality:

$$\begin{aligned} \mathbf{x}(\mathbf{p}, w) &= \underset{\mathbf{x}}{\text{argmax}}\ u(\mathbf{x})\ s.t.\ \mathbf{x}\cdot\mathbf{p} \leq w \\ \max \Lambda(\mathbf{x}, \mathbf{p}, w) &= u(\mathbf{x})\ + \lambda(w - \mathbf{x}\cdot\mathbf{p}) \\ \forall i, j:\ MRS_{i,j} &= \frac{\frac{\partial u}{\partial x_i}}{\frac{\partial u}{\partial x_j}} = \frac{p_i}{p_j} \end{aligned} \tag{6}$$

That is, demand is characterised by the bundle that equates the Marginal Rate of Substitution (MRS) between each pair of goods to the their respective relative prices (in the case of competitive market pricing). A bundle of goods represents a utility-maximising consumption choice for a given set of prices as well as a budget if it represents Walrasian demand for each of its constituent goods.

While most economists will be highly familiar with this basic insight, it is also the starting point for a number of practical challenges in an applied context. Direct utility is elusive and empirically intractable in the absence of explicit assumptions about its functional form. In the following, I will contrast two distinct ways to create an estimate of what the utility-maximising bundle of agents may be, based only on data in the form of binary bundle comparisons.

2.1.2 Setting example parameters

Recall that one of the principal motivations for this examination was to create a method of eliciting agents' preferences over a wide range of goods in a situation where market prices are not observable and data collection is subject to high cost. The surveying technique (both simulated and empirical) chosen for this is *binary*

bundle comparison. That is, and experimenter presents agents with two bundles at a time – each consisting of different types and amounts of goods. The agent then indicates which of the two bundles would be the preferred choice. I then created simulated sample data, on which estimations are performed, through the following steps:

Step 0 – Defining parameters Assume the existence of two perfectly rational agents – agent q and agent p – in a universe where there are exactly two goods from which both agents derive utility. Goods are only presented to the agents in the shape of two bundles at a time, each consisting of a given amount of each good. The rational agents always choose the bundle, which corresponds to the set of goods that yield the highest utility out of both presented bundles. Only 1) the presented bundles X_1 and X_2, and 2) the agents' choice of bundle, π, are observable to the experimenter. For the sake of this illustrative example, the following is assumed to constitute the agents' utility function:

$$
\begin{aligned}
&With\ X_1 = (x_1^A, x_1^B)\ and\ X_2 = (x_2^A, x_2^B) \\
&and\ u_l(X_i) = \begin{cases} (x_i^A)^{0.1} \cdot (x_i^B)^{0.9} & l = q \\ (x_i^A)^{0.7} \cdot (x_i^B)^{0.3} & l = p \end{cases} \\
&let\ \pi_l(X_1, X_2) = \begin{cases} 1 & u_l(X_1) \geq u_l(X_2) \\ 0 & otherwise \end{cases}
\end{aligned}
\tag{7}
$$

With u_l describing the non-observable direct utility function for agent l for goods x_A and x_B as contained in the two presented bundles X_1 and X_2. Note that in this particular case utility takes on a Cobb-Douglas functional form and also exhibits

a constant elasticity of substitution. This assumption of functional form is purely made for illustrative processes – the estimation strategy described in the following steps is explicitly designed for situations where functional form assumptions are to be avoided. That is, the observation of l, X_1, X_2, and $\pi_l(X_1, X_2)$ is sufficient for outcome identification.

Step 1 – Bundle generation In close analogy to the field experiment, which generated the data I analyse below, I generate 50 bundle comparison questions, which are presented to both agents. The contents of each bundle comparison $\mathcal{C}_j$ are determined randomly, satisfying the following conditions:

$$\begin{aligned} & \forall j \in \{1, ..., 50\} \;\; \mathcal{C}_j = \{X_{j,1}, X_{j,2}\} \\ s.t. \;\; \forall i \in \{1,2\} \;\; & w > x^A_{j,i} \cdot p^A + x^B_{j,i} \cdot p^B \geq w \cdot 0.75 \\ & and \;\; \forall i \in \{1,2\} \;\; x^A_{j,i} > 0 \;\; and \;\; x^B_{j,i} > 0 \end{aligned} \tag{8}$$

Where w is a fixed available budget per bundle and p^A and p^B are the market prices for goods A and B, respectively. That is, all bundles that make up the 50 comparisons are drawn from a channel of the goods space below the budget plane.

Step 2 – Bundle choice elicitation I present all 50 bundle pairs sequentially to both simulated agents – that is, I evaluate the direct utility of both bundles involved in each bundle twice. First, I apply the utility function assumed for agent q, then I apply the one assumed for agent p – both times using the same bundle contents, pre-determined in Step 1. If the calculated direct utility value for bundle $X_{j,1}$ is larger or equal than the one for bundle $X_{j,2}$, I record a value of 1 in the associated output vector – otherwise, I record a value of 0.

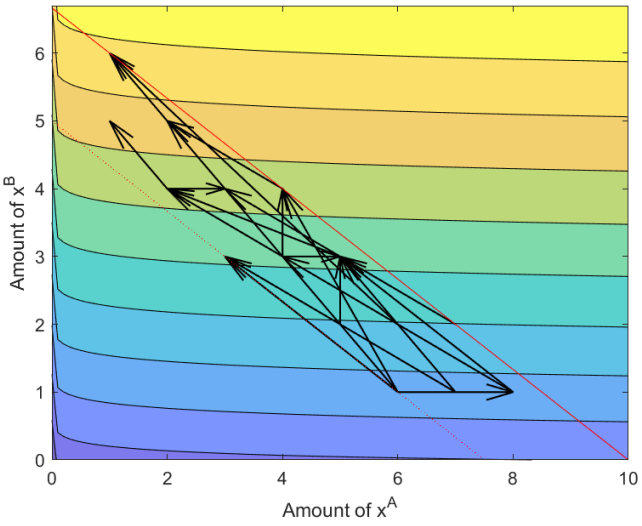
Thus, I construct the following dataset, which is used as the basis for the subsequent Gaussian Process estimation.

$$\mathbf{y} = \begin{pmatrix} y_{1,q} \\ y_{2,q} \\ ... \\ y_{50,q} \\ y_{1,p} \\ ... \\ y_{50,p} \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} x^A_{1,1}, x^B_{1,1}, x^A_{1,2}, x^B_{1,2}, q \\ x^A_{2,1}, x^B_{2,1}, x^A_{2,2}, x^B_{2,2}, q \\ ... \\ x^A_{50,1}, x^B_{50,1}, x^A_{50,2}, x^B_{50,2}, q \\ x^A_{1,1}, x^B_{1,1}, x^A_{1,2}, x^B_{1,2}, p \\ ... \\ x^A_{50,1}, x^B_{50,1}, x^A_{50,2}, x^B_{50,2}, p \end{pmatrix} \tag{9}$$

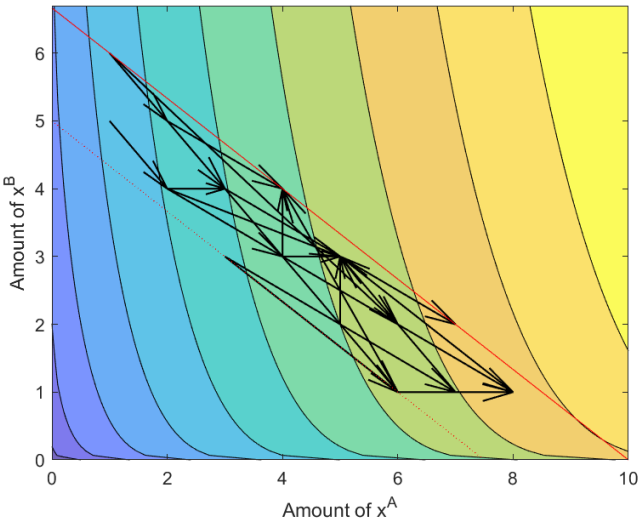
Where $y_{j,l}$ indicates the instance where agent l has expressed a preference for Bundle $X_{j,1}$ in the j^{th} presented bundle comparison – in which case $y_{j,l}$ assumes value 1, or value 0 in case of reverse preferences.

Figure 12 provides an overview of the data aggregated so far. It displays two key elements for each of the two agents over the applicable segment of the goods space created by the two goods x^A and x^B: 1) The indifference curves implied by the utility function assumed for each of the two agents, and 2) the set of 50 randomly created bundle comparisons – each consisting of two bundle-level points connected by an arrow that marks their joint belonging to a mutual comparison.

Figure 12: Preferences expressed by Agents over bundles in sample



(a) Agent P



(b) Agent Q

Note: In foreground, pairs of jointly presented bundles connected by arrows (black) pointing to that bundle that was preferred by respective agent; solid red line marks budget restriction, dotted red line marks lower band of permissible bundle expenditure (fixed at 75% of budget in this case). In background, visualisation of underlying utility function through selected iso-utility / indifference curves (each goods space spanned by two iso-utility lines filled in with distinct colour). For improved figure legibility, only 30 out of the presented 50 bundles are shown.

2.2 Simulated Maximum Likelihood Estimation Approach

2.2.1 Correct specification

One of the most common conventional applied microeconomics strategy for estimating utility-maximising bundles consists of a two-step approach: first, assume a structural model driven by a selection of structural parameters (often also includes an assumption of the distribution of associated error terms). Then, employ a likelihood estimation technique to compare the performance of a range of tested values for the posited parameters based on a relevant set of empirical observations.

A relatively simple implementation of such an approach is the Cobb-Douglas utility function as a structural model and Maximum-Likelihood Estimation (MLE) as an estimation technique.[21] Consider the following specification.

Structural model In analogy to Equation 7, assume that the choice of Agent q between bundle X_1 and X_2 is governed by the following process:

$$\begin{aligned} &\textit{For Agent } q \textit{ with } X_{1,j} = (x_{1,j}^A, x_{1,j}^B) \textit{ and } X_{2,j} = (x_{2,j}^A, x_{2,j}^B) \\ &\qquad\qquad \textit{and } u(X_i) = (x_i^A)^\alpha \cdot (x_i^B)^{1-\alpha} \textit{ and } \epsilon_j \sim N(0,\sigma) \textit{ i.i.d.} \\ &\textit{let } \pi_j(X_{1,j}, X_{2,j}, \epsilon_j, \alpha) = \begin{cases} 1 & u(X_{1,j}) - u(X_{2,j}) \geq \epsilon_j \\ 0 & \textit{otherwise} \end{cases} \end{aligned} \tag{10}$$

With u describing the non-observable direct utility function for Agent q for goods x_A and x_B as contained in the two presented bundles X_1 and X_2; α and σ are the two

[21] Note that the Cobb-Douglas utility function also represents the true data-generating process in this particular example (see above). However, this will not necessarily be the case in the field and thus the performance of this approach is likely overstated in the following.

parameters being estimated by the maximum-likelihood estimation; ϵ is a normally distributed shock or error term; j denotes a specific observation instance within a sequence of bundle comparison prompts.

Estimation technique Accordingly, the likelihood function of a given set of parameters (α, σ) is:

$$\mathcal{L}(\mathbf{X}, \mathbf{y}, \alpha, \sigma) = \prod_{k=1}^{K} \prod_{j=1}^{J} \pi_{j,k}(X_{1,j}, X_{2,j}, \epsilon_{j,k}, \alpha) \cdot y_j \tag{11}$$

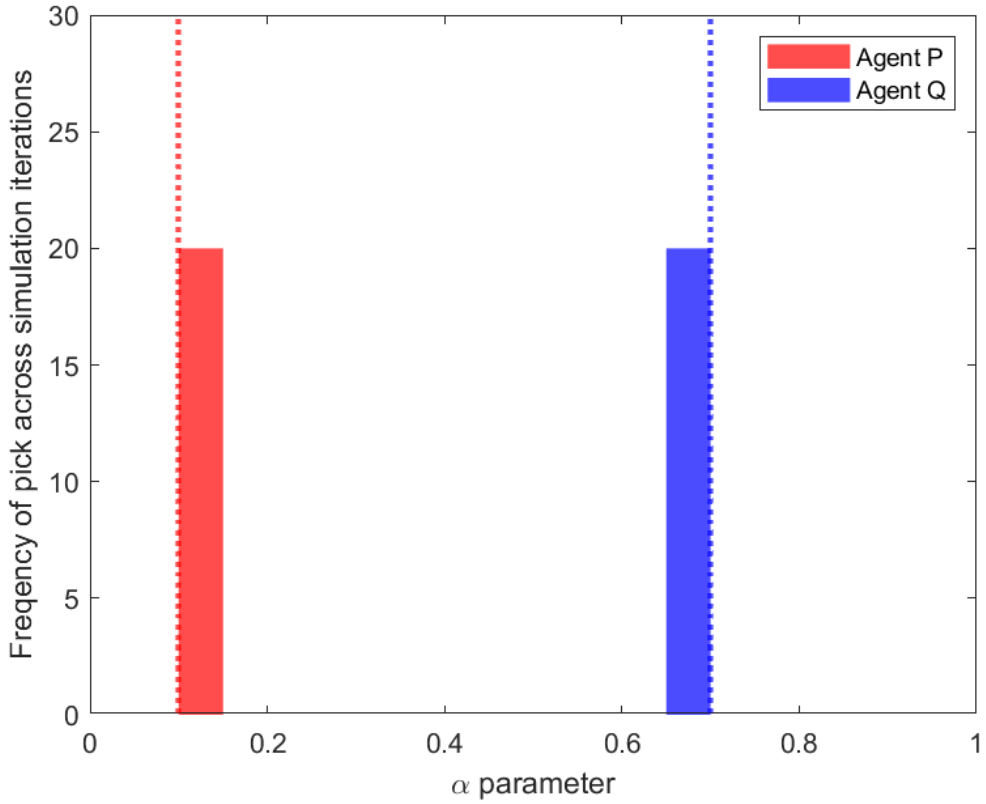
Where $\mathcal{L}$ denotes the joint likelihood of parameters (α, σ) conditional on the observable bundle content inputs, $\mathbf{X}$, and observed bundle choices, $\mathbf{y}$ – both of which carry J observations each for Agent q; k indicates the possibility to iterate the estimation process over the given amount of observation K-times. Note that the influence of σ registers indirectly through the error term $\epsilon_{j,k}$, individual observations are assumed to be i.i.d. Thus, the resulting values for the desired parameters are obtained by numerically solving the following problem:

$$(\hat{\alpha}_{ML}, \hat{\sigma}_{ML}) = \underset{\alpha,\sigma}{\operatorname{argmax}} \ \mathcal{L}(\mathbf{X}, \mathbf{y}, \alpha, \sigma) \tag{12}$$

Applying this estimation procedure to the simulated dataset, I obtain the results shown in Figure 13 for Agents P and Q, respectively. As follows from Equation 7, every estimate of α implies a corresponding optimal bundle allocation (while sigma indicates signal-noise ratio in agents' responses). Figure 14 shows the distribution of these resulting estimates for optimal bundles across the goods space assumed ahead of this simulation. With the estimation technique having assumed the functional form

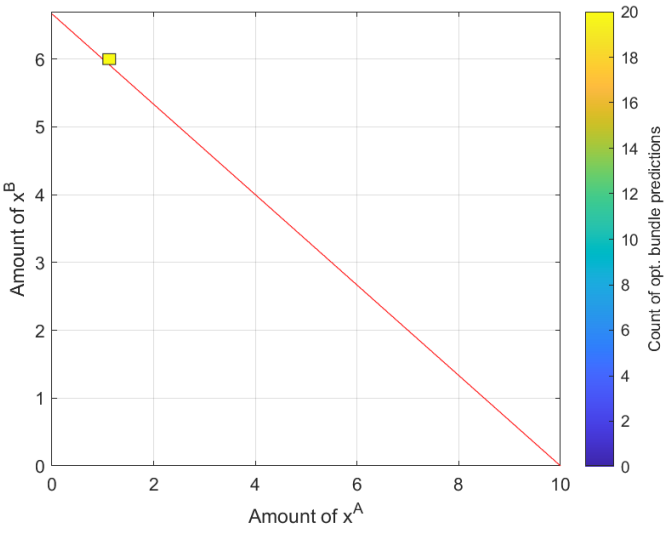
and parameter space of the underlying data-generating process, it follows without surprises that the resulting estimate is: 1) accurate (identifying the correct respective values for α for Agents P and Q; and 2) precise (choosing the lowest permissible values for σ. Subsequently, the implied policy bundles in Figure 14 are highly concentrated on the respective values that one can derive from Equation 7. This is partially due to the robustness of the MLE approach itself, but mostly due to the "lucky" guess of exactly mirroring the data-generating process with the structural model.

Figure 13: Maximum Likelihood Estimation – α-parameter picks

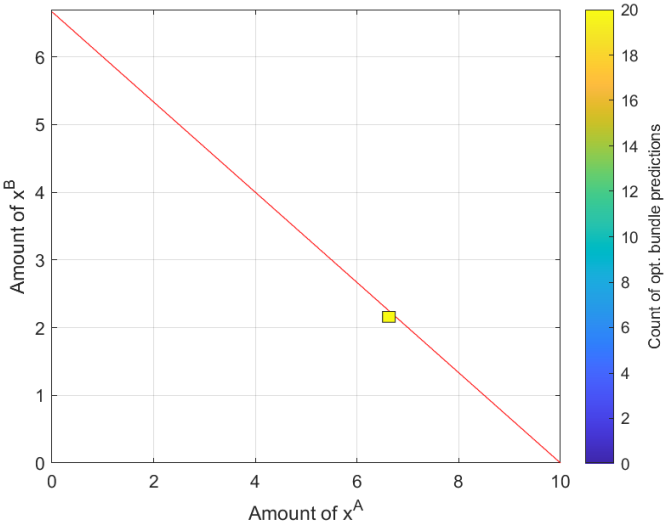


Note: Histogram of frequency parameter picks of α parameter picks across 20 simulation iterations. Red bars denote simulation for Agent P, blue bars denote simulation for Agent Q. Vertical dotted line indicates location of true underlying parameter value.

Figure 14: Maximum Likelihood Estimation – Optimal Bundles under CB



(a) Agent P



(b) Agent Q

Note: Coloured tiles span "heatmap" of optimal bundle allocations indicated by α values yielded by Maximum Likelihood Estimation. Colour of tile corresponds to number (out of 20) of policy choice bundles estimated to be within bounds of respective tile. Please consult colour legend on RHS of respective chart.

2.2.2 Incorrect specification

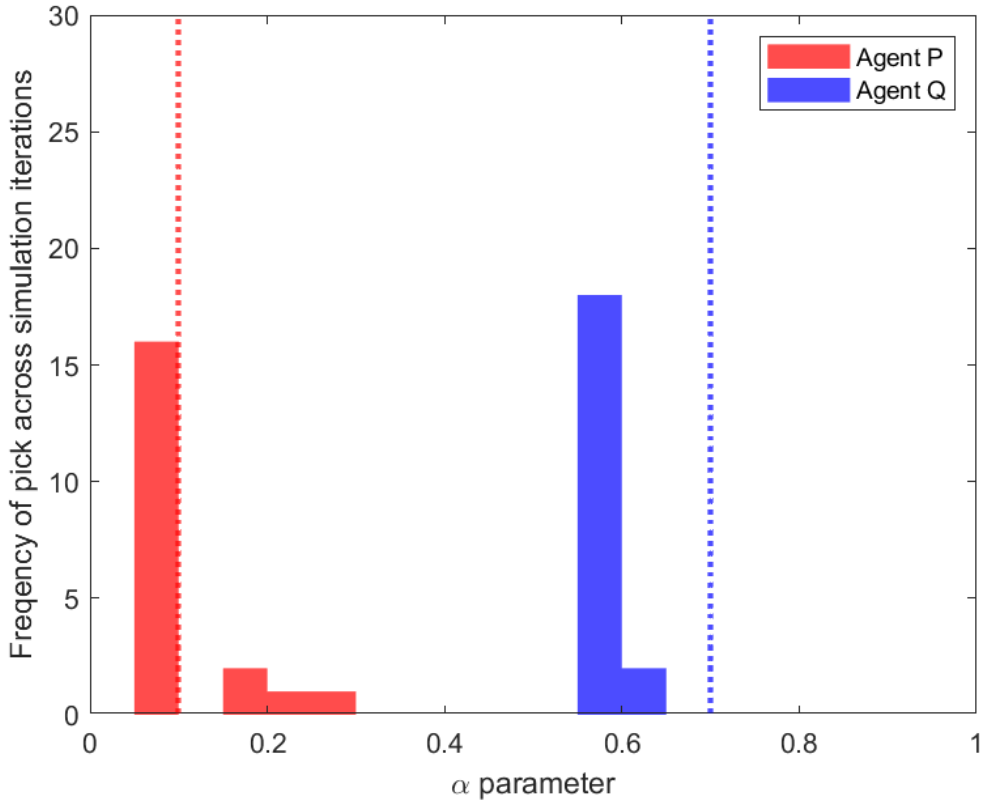
The predictive success of the method outlined in the previous section is of little surprise, given the alignment of actual and modelled preference relation. To illustrate the potential implication of this alignment not being in place – as may be plausibly the case in empirical applications where lack of knowledge about the actual preference relations is precisely the point of the empirical examination – I subsequently illustrate the results using 1) the same simulated bundle choice data; 2) the same MLE-approach; but 3) I assume the following preference relation while estimating.

$$\begin{aligned} &For\ Agent\ q\ with\ X_{1,j} = (x^A_{1,j}, x^B_{1,j})\ and\ X_{2,j} = (x^A_{2,j}, x^B_{2,j}) \\ &\qquad\qquad and\ u(X_i) = [(x^A_i)^\alpha + (x^B_i)^\alpha]^{\frac{1}{\alpha}}\ and\ \epsilon_j \sim N(0,\sigma)\ i.i.d. \\ &let\ \pi_j(X_{1,j}, X_{2,j}, \epsilon_j, \alpha) = \begin{cases} 1 & u(X_{1,j}) - u(X_{2,j}) \geq \epsilon_j \\ 0 & otherwise \end{cases} \end{aligned} \tag{13}$$

With u describing the non-observable direct utility function for Agent q for goods x_A and x_B as contained in the two presented bundles X_1 and X_2; α and σ are the two parameters being estimated by the maximum-likelihood estimation; ϵ is a normally distributed shock or error term; j denotes a specific observation instance within a sequence of bundle comparison prompts. Please note that the functional form of the utility function only represents a minor deviation from the one employed in Equation 10 above.

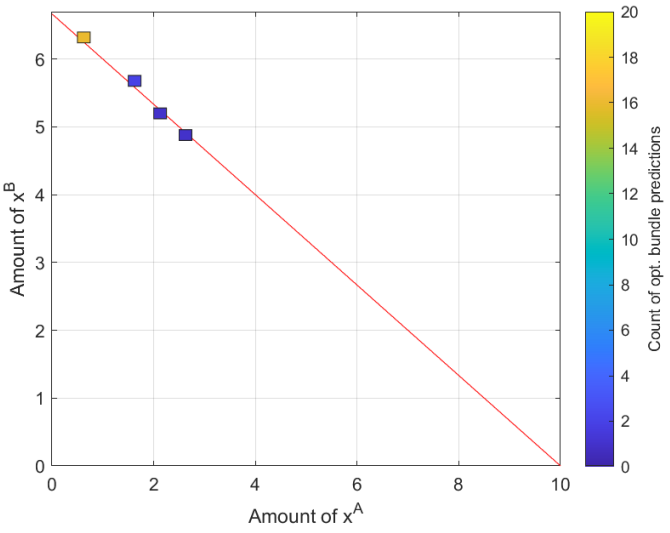
As in the previous section, I show estimation results in Figure 15 for Agents P and Q, respectively. Note the significantly lower level of predictive power compared to the previous section – both in accuracy and precision. As is illustrated by the implied policy bundles in Figure 16, this approach represents a worse performance compared to the previous modelling example. With full context knowledge, it would seem obvious that there is a clear misspecification issue here. However, researchers wanting to shed light on empirical preference relations almost never have the luxury of knowing the true functional form of agents' utility. For example, the CES-utility function used in Equation 13 is a staple of microeconomics literature – its use for this model would not seem misplaced, ex ante. This challenge is further complicated by inaccurate responses and a larger set of goods regarding which preferences are queried.

Figure 15: Maximum Likelihood Estimation – α-parameter picks

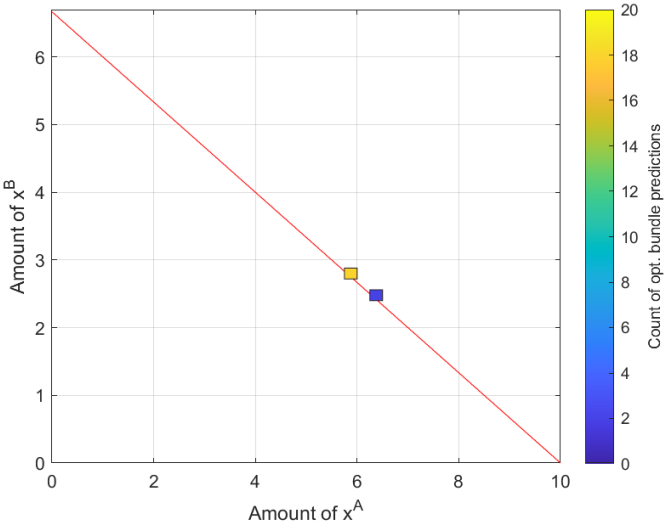


Note: Histogram of frequency parameter picks of α parameter picks across 20 simulation iterations. Red bars denote simulation for Agent P, blue bars denote simulation for Agent Q. Vertical dotted line indicates location of true underlying parameter value.

Figure 16: Maximum Likelihood Estimation – Optimal Bundles under CES



(a) Agent P



(b) Agent Q

Note: Coloured tiles span "heatmap" of optimal bundle allocations indicated by α values yielded by Maximum Likelihood Estimation. Colour of tile corresponds to number (out of 20) of policy choice bundles estimated to be within bounds of respective tile. Please consult colour legend on RHS of respective chart.

2.3 Gaussian Process Estimation Approach

In contrast to the conventional approach described in the previous sections, I present an alternative path to solving the initially stated problem. To do that, I use the same simulated bundle choice dataset (see above) to execute a Gaussian Process training estimation. This technique is well described in the statistics and machine learning literature.

2.3.1 Introduction to Gaussian Processes

Rasmussen and Williams (2006) define a Gaussian Process as a *"collection of random variables, any finite number of which have a joint Gaussian distribution"*. That is, a Gaussian Process describes a stochastic way to draw a given number of functions, all of which are linked to each other through a Gaussian distribution. This process is defined by the Gaussian distribution at its core – which is in turn fully described by two parameters:

$$\begin{aligned} y := f(\mathbf{X}) \sim \mathcal{GP}(m(\mathbf{X}), k(\mathbf{X}, \mathbf{X}')) \\ where\ m(\mathbf{X}) = E[f(\mathbf{X})] \\ and\ k(\mathbf{X}, \mathbf{X}') = E[(f(\mathbf{X}) - m(\mathbf{X}))(f(\mathbf{X}) - m(\mathbf{X}))'] \\ such\ that\ \tilde{f}(\tilde{\mathbf{X}}) \sim \mathcal{N}(m(\tilde{\mathbf{X}}), k(\tilde{\mathbf{X}}, \tilde{\mathbf{X}}')) \end{aligned} \tag{14}$$

Where f is the function yielding output value $\mathbf{y}$ depending on $\mathbf{X}$, $\mathbf{X}$ is a set of input variables of any finite dimension; m is the mean function contained in the Gaussian Process; and k is the kernel function contained in the Gaussian Process. Note that in the case of a realised draw of a specific realisation of the input variables, $\tilde{\mathbf{X}}$, the $\mathcal{GP}$ yields a specific Gaussian function, $\tilde{f}$ that contains the process' 'prediction' of the output variables for this given set of input data.

For the case of this application of Gaussian Processes, the following guiding choices were made with respect to the functional form of the two parameters defining the Gaussian Process:[22]

- **Mean function** The mean function is assumed to yield 0 for all possible input values. Note that this does not imply an expected value of 0 for all posterior functions (such as $\tilde{f}$). Expected values can and will assume non-zero values based on corresponding conditional updating. Assuming a non-zero mean function would be appropriate in cases where a deterministic trend or pattern in output values, driven by input values, is known.

$$m(\mathbf{X}) := 0 \tag{15}$$

- **Kernel** The kernel representing the covariance function of the process is assumed to be of a squared exponential shape:

$$k(\mathbf{X}_1, \mathbf{X}_2) = cov(f(\mathbf{X}_1), f(\mathbf{X}_2)) := \sigma^2 exp[-\frac{1}{2l}\|\mathbf{X}_1 - \mathbf{X}_2\|^2] \tag{16}$$

  Where $\mathbf{X}_1$ and $\mathbf{X}_2$ are any two sets of draws of input variables, σ reflects a signal variance hyper-parameter, and l is a scale hyper-parameter. This very commonly used kernel family implies that the sequence of function draws from the Gaussian Process can be shown to be identical to a Bayesian linear regression.[23] Among other useful properties, this information highly simplifies

[22]Note that while this does, of course, impact the Gaussian Process in the sense of choosing a prior-to-the-prior, it does not limit the shape or nature of the posterior functions – that is, the random functions drawn from the Gaussian Process.

[23]For proof of this proposition, the reader is asked to consult Rasmussen and Williams (2006)

the conditional updating outlined below.

One of the most powerful features of the Gaussian Process set up described so far is that it follows a highly intuitive, analytically tractable, and Bayesian updating path when new observations come to light. This could be a singular data addition (such as in the case of adding a prediction set to a training set) or it could be a continuously growing set of data. Recall that, by definition, any two function draws from a common Gaussian Process are subject to a joint Gaussian distribution:

$$\begin{bmatrix} f_1(\mathbf{X}_1) \\ f_2(\mathbf{X}_2) \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} k(\mathbf{X}_1, \mathbf{X}_1), k(\mathbf{X}_1, \mathbf{X}_2) \\ k(\mathbf{X}_2, \mathbf{X}_1), k(\mathbf{X}_2, \mathbf{X}_2) \end{bmatrix}\right) \quad (17)$$
$$with\ \mathbf{X}_1 \in \mathbf{X}_2$$

Where $\mathbf{X}_1$ and $\mathbf{X}_2$ are two sequentially expanded sets of observations for the considered input variables. f_1 and f_2 are the respective function draws yielded by the underlying Gaussian Process as posterior estimates following exposure to $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. This example illustrates the essence of Gaussian Processes: the (Gaussian) relationship between individual functions – the functional form of which depends on the observed input and output data.

Encoding this process in the terminology of Bayesian statistics, one can therefore call f_1 the 'prior' and f_2 the 'posterior' belief about what the true relationship between input and output values. Any additional observations prompt a further iteration of conditional updating – a new function-level draw from the Gaussian Process. Recall that by general property of Gaussian distributions, this process can be captured analytically as follows:

pp. 97-98.

$$
\begin{aligned}
f_2|_{\mathbf{X}_2,\mathbf{X}_1,f_1} \sim \\
\mathcal{N}\Big(k(\mathbf{X}_2,\mathbf{X}_1)\cdot k(\mathbf{X}_1,\mathbf{X}_1)^{-1}\cdot f_1, \\
k(\mathbf{X}_2,\mathbf{X}_2)-k(\mathbf{X}_2,\mathbf{X}_1)\cdot k(\mathbf{X}_1,\mathbf{X}_1)^{-1}\cdot k(\mathbf{X}_1,\mathbf{X}_2)\Big)
\end{aligned}
\tag{18}
$$

2.3.2 Applying Gaussian Processes to Bundle Choice Data

Circling back to the task at hand – the processing of bundle-choice data for the purpose of deriving insights on underlying preference relations – recall that the sections above focused on an MLE-based estimation of a preference-probability function π that would yield the probability of a given agent preferring one bundle to another (see Equations 10 and 13). Re-stating this function in the context of Gaussian Processes thus yields the following estimation target:

$$
\begin{aligned}
\pi_l(\mathbf{X}_1,\mathbf{X}_2) = P(y_l(\mathbf{X}_1,\mathbf{X}_2)=1) \\
s.t.\ y_l(\mathbf{X}_1,\mathbf{X}_2) = \begin{cases} 1 & u_l(\mathbf{X}_1)-u_l(\mathbf{X}_2)\geq 0 \\ 0 & otherwise \end{cases}
\end{aligned}
\tag{19}
$$

where $\mathbf{X}$ and y are structured as described in Equation 9 for a given agent l. With the π function taking the place of the more general f function used in Equation 18. I use the simulated bundle-choices data used in the previous section to train such a Gaussian Process model.[24]

[24]Training obs. count: 100, In-sample accuracy: 1.00, σ hyper-parameter: 0.45, l hyper-parameter: 0.00

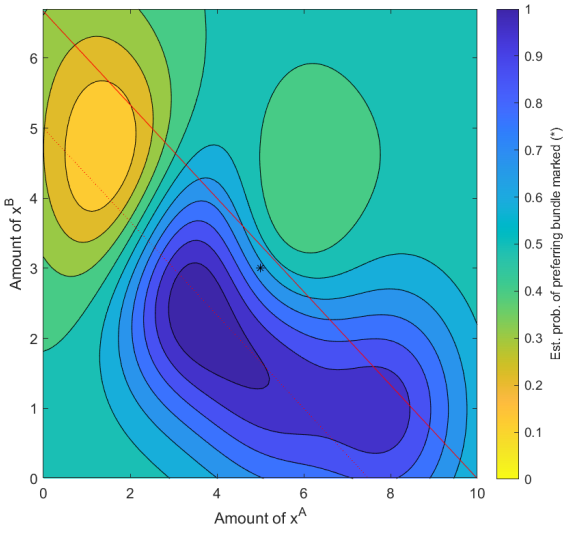
The process of "training" a Gaussian Process specifically implies estimating its defining parameters from a given data set. Of primary importance is the kernel function $k(\mathbf{X}_i, \mathbf{X}_j)$, which subsequently permits the formulation of a Gaussian Process prediction: a conditional updating of the Bayesian prior according to Equation 20. After choice of i) a kernel function shape (squared exponential kernel, in this case); and ii) appropriate kernel function hyper-parameters, the estimation of the kernel function takes place by calculating the following:[25]

$$\begin{gathered} \forall\ \mathbf{X}_{i,j}\ s.t.\ i \neq j \\ \hat{k_{i,j}}(\mathbf{X}_i, \mathbf{X}_j | \sigma_f, \sigma_l) = \sigma_f^2 \exp\left(-\frac{(\mathbf{X}_i - \mathbf{X}_j)^T (\mathbf{X}_j - \mathbf{X}_i)}{2\,\sigma_\ell^2} \right) \end{gathered} \tag{20}$$

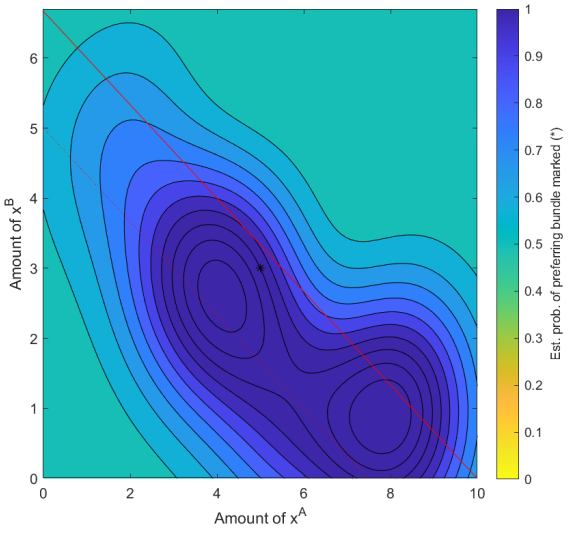
Where $\hat{k_{i,j}}$ is the element of the estimated kernel function matrix with index (i, j), σ_f and σ_l are the two kernel function hyper-parameters representing a scale factor and a signal standard deviation / smoothness factor, respectively. Other kernel function forms – such as the Matern or the Rational Quadratic Kernel – are also described in the literature but less commonly used in applications. The resulting posterior for the predicted bundle-preference probability is illustrated in Figure 17.

[25]There is a range of comprehensive software solutions available – for the purpose of this chapter, I used the *fitrgp* function embedded in Matlab.

Figure 17: Gaussian Process – Posterior (No priors)



(a) Agent P



(b) Agent Q

Note: Illustration of Gaussian Process posterior estimate of bundle-comparison preference probability. Colour encodes predicted probability of preferring the certain but arbitrarily chosen reference bundle – (5,3) – to the bundle corresponding to the respective coordinates.

Note that bundle preference probabilities shown here are anchored to a reference bundle (probability to prefer one bundle *to which other bundle*?) – which was chosen arbitrarily for illustrative purposes in Figure 17. This does not imply that the predictive capabilities of Gaussian Processes is limited to use cases where some reference bundle is already known or otherwise chosen. Much rather, this is simply a specific cut through the high-dimensional Gaussian Process structure which most closely resembles the two-by-two charts presented above. Comparing Figure 17 to the underlying 'true' shape of preferences in Figure 12, the shapes are not identical. What is more important for this application: there is a clear difference in the gradient of the probability field between estimates for Agent q and Agent p in proximity of the reference bundle. This difference in gradient corresponds qualitatively exactly to what one might expect based on the underlying true preferences: around the indicated reference point, Agent q's iso-probability curves are less steep than Agent p's. While there is no suggestion that iso-probability curves in Figure 17 are an approximation of indifference curves (such as those shown in Figure 12), the difference in gradients between agents reflects exactly the same concept that is presented by the difference in steepness of indifference curves. That is, at the indicated point, Agent q is less willing to exchange units of x^B for x^A than Agent p.

To further leverage the predictive power of the Gaussian Process, I utilise its Bayesian properties (see Equation 18) and introduce a prior belief without relying on any additional strong economic assumptions. This prior belief is generated by training the Gaussian Process with 100 randomly generated bundle comparison observations which are "economically empty". That is, both of the bundles presented in them are exactly the same, except for one randomly chosen good in one randomly chosen bundle being altered by a random shock. This prior training data could have

the following shape, for example:

$$\tilde{y} = \begin{pmatrix} 0 \\ 1 \\ ... \end{pmatrix} \tilde{X} = \begin{pmatrix} x^A_{1,1}, x^B_{1,1}, x^A_{1,1}, x^B_{1,1} + \epsilon_{1,q}, q \\ x^A_{2,1}, x^B_{2,1} + \epsilon_{2,q}, x^A_{2,2}, x^B_{2,2}, q \\ ... \end{pmatrix} with\ \forall i, l: \ \epsilon_{i,l} \sim N(0, 2)\ i.i.d. \quad (21)$$

where values in $\tilde{y}$ indicate whether the respective agent has expressed a preference for Bundle 1 in the corresponding randomly generated row of the $\tilde{X}$ matrix. These observations are "economically empty" in the sense that the only meaningful information they contain is that agents choice behaviour is consistent with the axiom of local non-satiation. As weak an assumption as that may be, this has an important effect on the posterior estimate of the Gaussian Process, which I run mixing these priors with the previous simulated dataset.[26] Just like Figure 17, Figure 18 shows probabilistic indifference curves resulting from the estimation based on such a prior belief. The Gaussian Process – accessing exactly the same actual bundle comparison observation as before – is now much more able to capture and quantify the differences in the agents' willingness to trade the two goods against each other, at the margin.

I then generate policy bundles – representing suspected utility-maximising allocations – from the posterior distribution obtained from the Gaussian Process derived above. I do this through randomised grid searches of the eligible goods space by iteratively comparing candidate bundles as follows:

- An initial anchor bundle $X_{anchor,0}$ is chosen at random from the queried goods space (that is, limited from above by the budget constraint and from below by

[26] Training obs. count: 200, In-sample accuracy: 1.00, σ hyper-parameter: 0.45, l hyper-parameter: 0.00

the lower bound of total bundle monetary value described above).

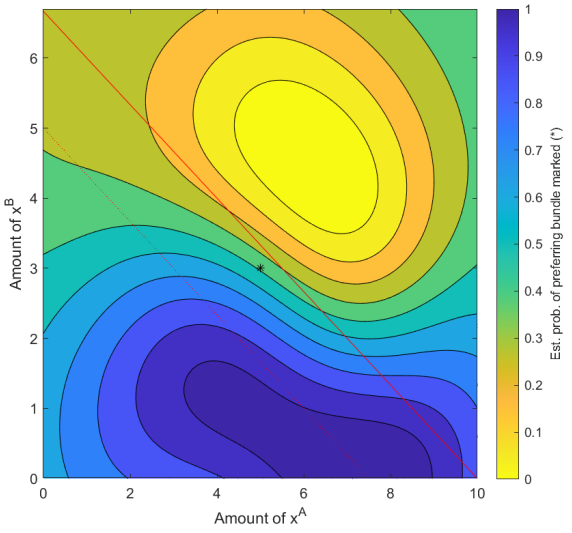
- Another ten candidate bundles, $X_{cand,l}$ with $l \in \{1, 10\}$, are drawn at random from the queried goods space. The Gaussian Process is queried to estimate the probability of each $X_{cand,l}$ being preferred to $X_{anchor,0}$, yielding $\hat{y}_l$ with $l \in \{1, 10\}$.

- The candidate bundle with the highest chance of being preferred to $X_{anchor,0}$ is nominated as the new anchor bundle, $X_{anchor,1}$.

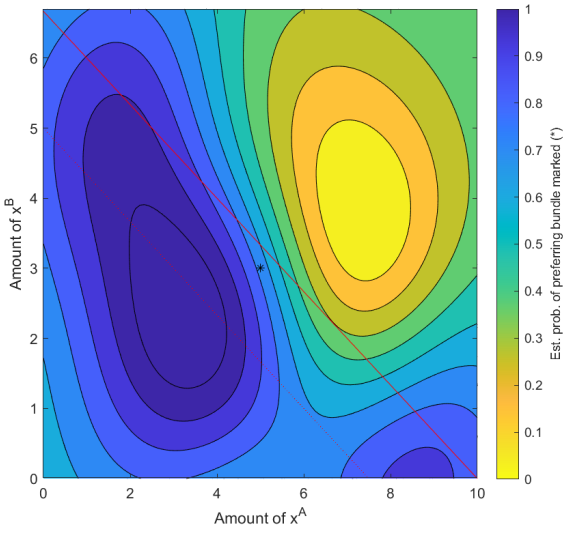
- The process is repeated for a maximum of 10 iterations or until there is no more substantial change between the anchor bundles of each iteration. The final iteration's anchor bundle is designated the policy bundle.

The resulting distribution of policy bundles (from 100 independent randomised grid searches) are projected onto the goods space in Figure 19. While there is a range in estimates, the center of probabilistic mass shows a clear distinction between the two simulated agents – exactly in the way one expects based on the underlying preference order. The Bayesian nature of the estimates introduces two idiosyncrasies: 1) Inferring smoothly from frequently queried parts of the goods space, the Gaussian Process algorithm hesitates to suggest extreme bundles as policy bundles – even if as in the case of Agent P, the "correct" policy bundle is located even further in the corner of the diagram than the Gaussian Process dares to suggest. 2) The probabilistic nature of the Gaussian Process' predictions leads to a share of predicted policy bundles being scattered at low density throughout the goods space – however, their distance from the Budget hyperplane very clearly rules these bundles out as policy bundles.

Figure 18: Gaussian Process – Posterior (With priors)



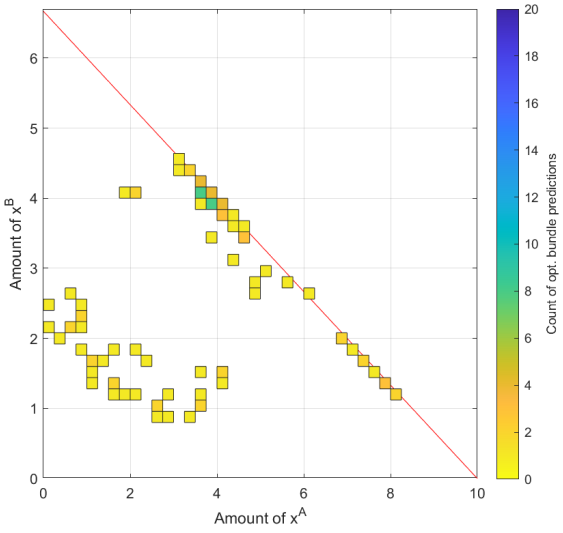
(a) Agent P



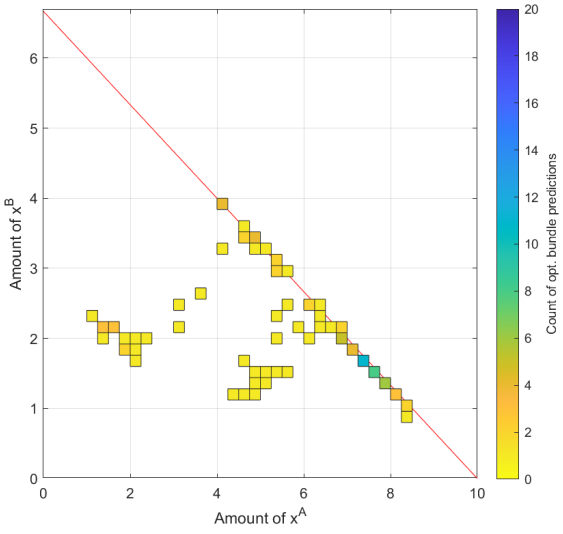
(b) Agent Q

Note: Illustration of Gaussian Process posterior estimate of bundle-comparison preference probability. Colour encodes predicted probability of preferring the certain but arbitrarily chosen reference bundle – (5,3) – to the bundle corresponding to the respective coordinates. Addition of 100 prior observations with no economic content.

Figure 19: Policy Bundle Estimates – Gaussian Process with Priors



(a) Agent P



(b) Agent Q

Note: Illustration of Gaussian Process posterior estimate of bundle-comparison preference probability. Colour encodes predicted probability of preferring the certain but arbitrarily chosen reference bundle – (5,3) – to the bundle corresponding to the respective coordinates. Addition of 100 prior observations with no economic content.

3 Empirical Results

In the previous section, I proposed a method through which Gaussian Processes can be used to generated insights about consumer preferences from otherwise difficult to process but easy to solicit bundle choice data. I now turn to apply the same methodology with empirical data gathered in rural Bangladesh.

3.1 Estimating policy bundles from empirical data

With the support of local partners, I gathered a comprehensive novel dataset of bundle-comparison data suitable for Gaussian Process estimation as outline above. The resulting data features 30 bundle comparison observations from 500 respondents each – adding to a total of 15,000 observations – with bundles constructed from a universe of 51 types of goods.[27] Using this dataset, I train Gaussian Processes with whom I then estimate a *policy bundle* for each agent – that is, an allocation of a fixed budget across all available 51 types of goods, which is most likely to represent the utility-maximising bundle of goods for this particular agent. I execute this estimation using the following iterative simulation algorithm for each agent a:

- **Step 1** Recall that the Gaussian Process always calls for two bundles at a time as input in order to provide as an output the probability of agent a preferring the first bundle to the second one. In each simulation resets r (out of a maximum of 10 resets), I initiate the search for the policy bundle by generating an initial *anchor bundle* $B_{a,r,anchor}$, which randomly allocates the entire available budget (BDT 50,000) across 2 randomly chosen goods within the 51-good universe.

[27] For additional information on the surveying and sample construction process, see chapter I.

- **Step 2** I then generate 10 *candidate bundles* – also constituting a random allocation of the entire budget to two randomly chosen types of goods. For every candidate bundle $B_{a,r,i}^{Cand}$, I query the Gaussian Process and I note the probability point estimate and associated estimated standard deviation for the probability of agent a preferring $B_{a,r,i}^{Cand}$ to $B_{a,r,anchor}$.

- **Step 3** Comparing the estimates for all candidate bundles, I select the bundle with the highest probability of being preferred to the anchor bundle and I nominate it the *suspected policy bundle*. I also compute a *confidence score* for the suspected policy bundle, which encodes the degree of conviction expressed by the Gaussian Process in the bundle comparison which led to the nomination of the suspected policy bundle.[28] The suspected policy bundle is placed on a shortlist and another simulation reset begins (back to Step 1) with the current suspected policy bundle as its new anchor bundle unless one of two conditions is met: i) The current suspected policy bundle is not materially different from that of the previous iteration; or ii) a total of 10 simulation resets has already been executed.

- **Step 4** Concluding the estimation for agent a, I calculate the arithmetic mean

[28] I calculate the confidence score for any given bundle comparison between two bundles B_1 and B_2 as follows:

$$\widehat{CF}_{B_1,B_2} = \frac{|\hat{y}_{B_1,B_2} - 0.5|}{\widehat{\sigma_y}_{B_1,B_2}} \tag{22}$$

where $\hat{y}_{B_1,B_2}$ is the point estimate for the probability of B_1 being preferred to B_2, and $\widehat{\sigma_y}_{B_1,B_2}$ is the associated standard deviation.

of up to 10 suspected policy bundles (and respective confidence scores) calculated in the sequence of simulation resets to generate the definitive policy bundle estimate for this agent. Since the averaging process can lead to slight deviations from the budgetary constraint, I re-scale every policy bundle proportionally back to exactly the budget constraint.

Therefore, the resulting dataset constitutes a matrix with a row for every agent in the sample – where every row contains: 1) average confidence score; 2) for each of the 51 types of goods in the universe: the allocated amount of that good in the policy bundle.

3.2 Estimation outcomes

The output generated by the previously described algorithm is highly actionable and contains concrete recommendations for bundles of goods that likely will be seen as utility-maximising by the respondents. Nevertheless, the output is also high dimensional and therefore not easy to represent in its entirety on a sheet of paper – below, I provide a set of illustrative perspectives. Tables 12 and 13 show the concrete allocations that represent the proposed policy bundles for a subset of 10 respondents. They show the allocations for those respondents whose policy bundle estimation scored the highest and lowest confidence scores, respectively. It is important to observe the diversity of the goods which are included in this subset of policy bundles: the fact that every respondent has been assigned a unique allocation as a policy bundle is a strong indication that the bundle comparison / Gaussian Process preference elicitation methods proposed above is able to finely distinguish and highlight different individual preferences by different respondents.

In order to further highlight the deep underlying preference relations that the Gaussian Process was able to capture based on the available data, Figure 20 shows the structure of the Gaussian Process for two respondents featuring in Tables 12 and 13 – delving into the posited preference relationship between the two types of goods that feature most prominently in each respondent's policy bundle. That is, this visualisation is the empirical data equivalent to what Figures 17 and 18 are showing for the case of simulated data. As this illustration highlights, the Gaussian Process is not only able to capture individual-level nuances in preference structure, it also appears to (independently) replicate a functional form that is remarkably similar to a high-dimensional Cobb-Douglas utility function. Note that there was no input of microeconomic theory whatsoever at any stage during the creation of Figure 20 – the underlying Gaussian Process has never been exposed to any specific microeconomic concept (such as Cobb-Douglas utility) or even utility functions in general (not even a set of "empty" priors to encode local nonsatiation of preferences). Moreover, different to the Gaussian Processes estimated in the theoretical illustrations above (which were trained on artificial data with an imposed underlying Cobb-Douglas utility function), the Gaussian Process that informs Figure 20 was strictly only trained on empirical response data gathered from survey respondents in the field.

Table 12: Policy bundle allocation of top 5 agents

Good Desc.	Top 5 Agents (by est. confidence)				
	(1) *Agent 3*	(2) *Agent 5*	(3) *Agent 34*	(4) *Agent 64*	(5) *Agent 120*
Cash	0%	0%	0%	0%	0%
Boat rental voucher	.700%	0%	2.8%	2.1%	3%
Building brick	0%	0%	.8%	.3%	1.7%
Agricultural fertilizer	0%	.4%	2.2%	.6%	0%
Irrigation motor (diesel-fuelled water pump)	0%	0%	0%	89.4%	0%
Construction (unskilled) labour voucher	0%	0%	0%	.5%	0%
Chicken feed	0%	.7%	1.4%	1.2%	0%
Goat / Cow feed	.4%	0%	0%	0%	0%
Young chicken / duck (female)	0%	.5%	0%	0%	0%
Young goat (female)	0%	0%	0%	0%	29.3%
Young cow (female)	0%	0%	0%	0%	2.6%
School uniform	0%	0%	0%	0%	0%
School tuition fee voucher	.8%	0%	0%	0%	3%
CNG fuel	0%	0%	0%	1%	0%
Downpayment voucher for CNG vehicle	0%	5.5%	0%	0%	0%
Downpayment voucher for Easybike	0%	20.7%	0%	0%	0%
Vehicle annual repair voucher	0%	0%	0%	.3%	0%
Downpayment voucher for Rickshaw	0%	0%	0%	.3%	0%
Laptop computer	0%	0%	0%	0%	0%
Debt payment voucher	0%	0%	0%	1.6%	0%
Diesel fuel	0%	.9%	1%	0%	2.1%
Fish feed	0%	0%	0%	0%	0%
Fishing rod	0%	0%	0%	0%	0%
Fruit tree seedling	.4%	1.4%	2%	0%	0%
Fertile soil	0%	.7%	4.80%	1.1%	0%
Dowry / wedding cost payment voucher	4.9%	1.6%	0%	0%	1.3%
New household fridge	52.1%	0%	0%	0%	0%
Voucher for household gas bill	0%	0%	0%	0%	0%
Downpayment for purchase / lease of arable land	0%	57.1%	0%	0%	0%
Voucher for construction of toilet	0%	0%	0%	0%	.7%
Mobile phone	0%	0%	0%	.1%	0%
Mobile phone data plan	0%	0%	2.5%	0%	2.9%
Motorbike downpayment	0%	2.4%	8.4%	0%	21.1%
Motorbike helmet	.2%	0%	0%	1.4%	0%
Electric air compressor	0%	0%	0%	0%	0%
Car battery	0%	.1%	0%	0%	.1%
Rickshaw battery	0%	0%	0%	0%	31.6%
Bike	0%	0%	0%	0%	0%
TV	17.4%	0%	69.9%	0%	0%
Radio	1.8%	.3%	.2%	0%	0%
Wifi router	0%	.4%	0%	0%	0%
Cement	0%	0%	0%	0%	0%
Bed / Wardrobe / Table	.1%	0%	.2%	0%	0%
Handloom machine	.4%	0%	0%	0%	0%
Paddy threshing machine downpayment	19.8%	0%	0%	0%	0%
Sewing machine	0%	6.5%	1.5%	0%	0%
Petrol fuel	0%	0%	0%	0%	.5%
Firing wood	.5%	0%	2.4%	0%	0%
Doctor visit voucher	.5%	.8%	0%	0%	0%
Tubes for well construction	0%	0%	0%	0%	0%

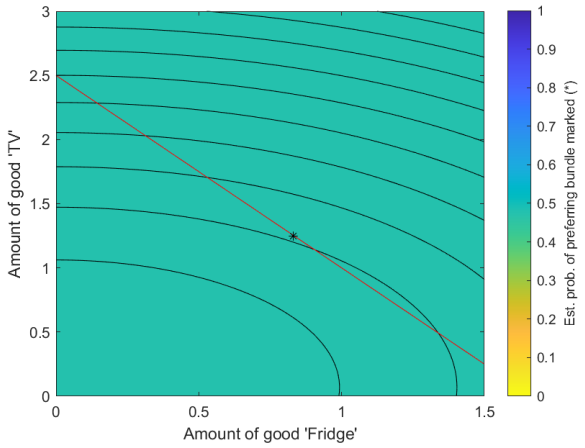
Note: Allocation of the policy bundles produced for the top 5 agents – ranked by descending estimation confidence score. Values indicate proposed percentage of available budget to be spent on the respective good type in the policy bundle as per Gaussian Process estimation.

Table 13: Policy bundle allocation of bottom 5 agents

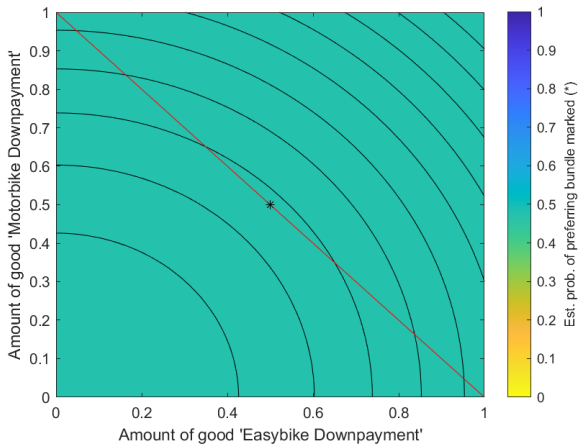
Good Desc.	Bottom 5 Agents (by est. confidence)				
	(1) *Agent 2394*	(2) *Agent 2627*	(3) *Agent 2650*	(4) *Agent 2671*	(5) *Agent 2675*
Cash	0%	0%	.8%	0%	.5%
Boat rental voucher	0%	0%	.2%	0%	0%
Building brick	0%	0%	.3%	0%	1%
Agricultural fertilizer	0%	0%	0%	0%	.7%
Irrigation motor (diesel-fuelled water pump)	0%	0%	0%	0%	0%
Construction (unskilled) labour voucher	0%	0%	0%	0%	0%
Chicken feed	0%	0%	0%	0%	0%
Goat / Cow feed	0%	0%	0%	0%	.6%
Young chicken / duck (female)	0%	0%	0%	0%	.5%
Young goat (female)	0%	0%	1.5%	0%	0%
Young cow (female)	0%	0%	0%	0%	0%
School uniform	0%	0%	0%	0%	0%
School tuition fee voucher	0%	0%	0%	0%	0%
CNG fuel	0%	0%	0%	0%	0%
Downpayment voucher for CNG vehicle	0%	.3%	0%	0%	0%
Downpayment voucher for Easybike	0%	0%	0%	0%	23.5%
Vehicle annual repair voucher	0%	0%	0%	0%	0%
Downpayment voucher for Rickshaw	0%	0%	0%	0%	0%
Laptop computer	0%	0%	0%	0%	0%
Debt payment voucher	0%	0%	0%	0%	0%
Diesel fuel	0%	.1%	0%	0%	0%
Fish feed	0%	0%	0%	0%	.5%
Fishing rod	0%	0%	0%	0%	0%
Fruit tree seedling	0%	0%	.4%	0%	1.3%
Fertile soil	0%	0%	0%	0%	.3%
Dowry / wedding cost payment voucher	99.8%	0%	67.3%	99.9%	0%
New household fridge	.1%	0%	0%	.1%	.4%
Voucher for household gas bill	0%	0%	0%	0%	.7%
Downpayment for purchase / lease of arable land	0%	0%	.2%	0%	0%
Voucher for construction of toilet	0%	0%	0%	0%	0%
Mobile phone	0%	0%	0%	0%	0%
Mobile phone data plan	0%	0%	0%	0%	0%
Motorbike downpayment	0%	0%	0%	0%	65.3%
Motorbike helmet	0%	7.2%	.1%	0%	0%
Electric air compressor	0%	92.4%	0%	0%	0%
Car battery	0%	0%	0%	0%	0%
Rickshaw battery	0%	0%	0%	0%	0%
Bike	0%	0%	0%	0%	0%
TV	0%	0%	23.8%	0%	0%
Radio	0%	0%	.2%	0%	0%
Wifi router	0%	0%	.2%	0%	.9%
Cement	0%	0%	0%	0%	0%
Bed / Wardrobe / Table	0%	0%	4.9%	0%	0%
Handloom machine	0%	0%	0%	0%	0%
Paddy threshing machine downpayment	0%	0%	0%	0%	.1%
Sewing machine	0%	0%	0%	0%	0%
Petrol fuel	0%	0%	0%	0%	0%
Firing wood	0%	0%	0%	0%	0%
Doctor visit voucher	0%	0%	0%	0%	0%
Tubes for well construction	0%	0%	0%	0%	0%

Note: Allocation of the policy bundles produced for the bottom 5 agents – ranked by descending estimation confidence score. Values indicate proposed percentage of available budget to be spent on the respective good type in the policy bundle as per Gaussian Process estimation.

Figure 20: Gaussian Process posteriors – empirical estimates



(a) Agent 3



(b) Agent 2675

Note: Illustration of Gaussian Process posterior estimates for Agent 3 and Agent 2675 (highest and lowest confidence score associated with policy bundle, respectively). Showcasing the two types of good each agent's policy bundle featured most prominently, compared to a reference bundle of a 50:50 allocation of those goods. Black lines show iso-probability curves, red lines show respective budget line.

The estimations shown above focus on a small subset of the total population of respondents in the empirical sample in order to facilitate a deep understanding of the layers making up the preference elicitation method I am proposing. Nevertheless, equivalent outcomes are available for the entire population of respondents. Table 14 below shows a summary of policy bundles stemming from this estimation using empirical data. For every one of the 500 respondents, an individual policy bundle was estimated – reflecting the most probable utility-maximising bundle according to the Gaussian Process estimation. Part B Table 14 shows the average and standard deviation, respectively, of the amount of each good contained in respondents' policy bundles. Part C shows the share of the total (hypothetical) available expenditure that was spent on any given good across all participants. As was the case for the more focused Tables 12 and 13, there is a wide range of the shares of global spend across good types – reflecting varying degrees of aggregate popularity of different goods. Secondly, the respondent-level spend share is also subject to very strong variance across respondents' policy bundles – with min-max ranges frequently ranging from 0 to 1. Both of these observations are further robust evidence that this method is able to pick up on differences in individual preferences. Figure 21 further provides a more focused view on goods popularity by showing expenditure share in policy bundles by type of good – but only reflecting those policy bundles in which the respective good has a non-zero allocation.

There appear to be certain types of goods which feature more prominently than others, on average, in policy bundles. As is particularly shown in Figure 21, participants' preferences place an emphasis on relatively chunky investments that are challenging to procure (for example advanced means of transportation). This result confirms prior findings of the microfinance literature, which imply that access to

microfinance helps alleviate underserved demand for chunky capital goods. Such an apparent difference in popularity of certain types of goods represents an insight which could have been generated by conventional reduced-form methods.[29] However, the added insights brought about by the utilisation of Gaussian Process is that a view of goods' popularity (or rather demand), can be generated on the level of the individual – not just at the aggregate level. Moreover, the granular good-by-good preference structure allows for the application of many intermediate analysis vectors, which are not straight-forward to translate from the realm of simplified theoretical applications using conventional methods (a direct estimation of the elasticity of substitution, for example). As outlined by these results, there is significant potential of employing Gaussian Processes for applied microeconomic analysis. The estimation of optimal goods allocations based on bundle comparison data is one proximate application that I propose in this chapter – however, many further paths of analysis open themselves up to further study.

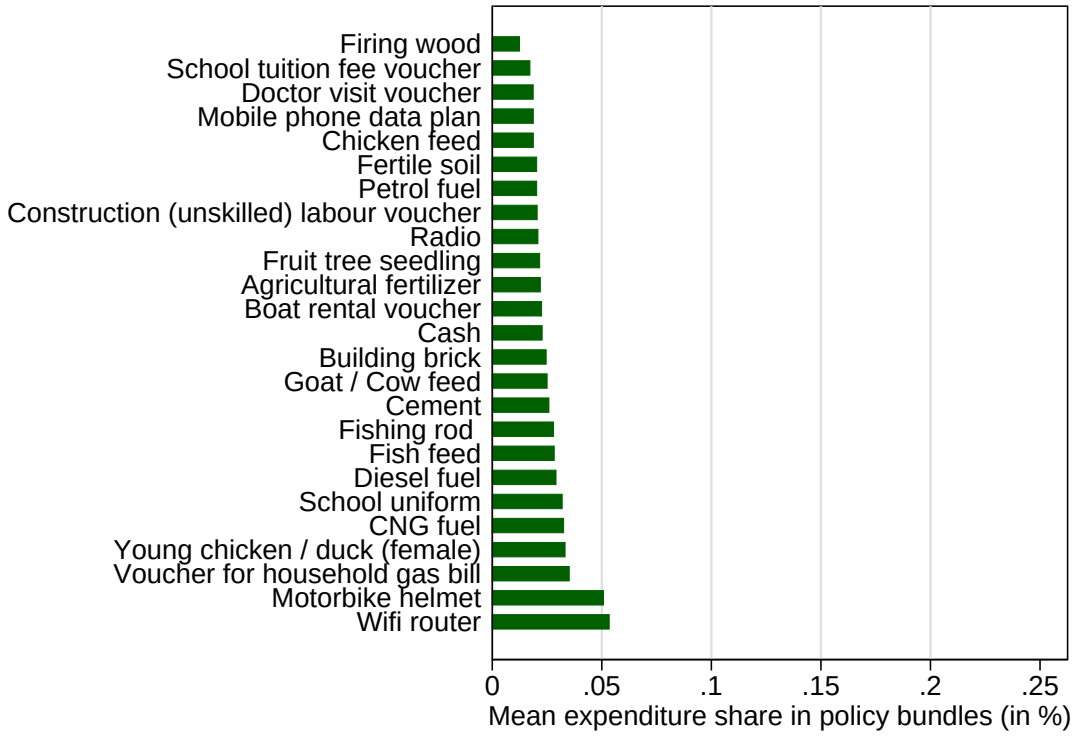
[29]For a summary of such analysis using the same empirical dataset, see chapter I.

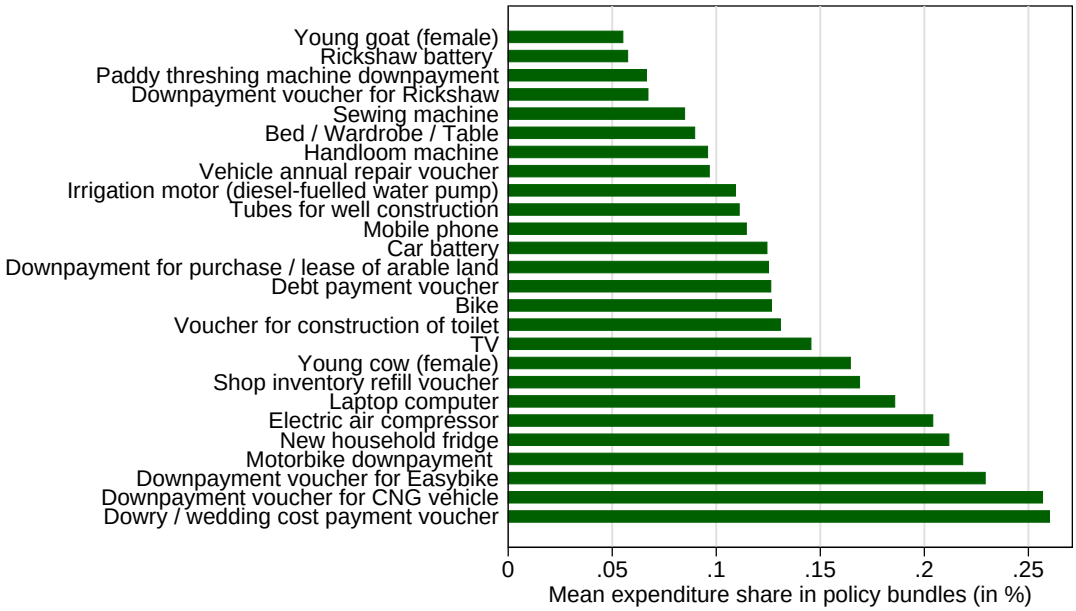
Table 14: Distribution of policy bundles across good types

Part A: Good Desc.		**Part B:** Bundle Content		**Part C:** Global Spend	**Part D:** Agent Spend Share			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Type	Price	Avg. (abs. unit)	Std. (abs. unit)	% of total	Avg. %	Med. %	Min. %	Max. %
Cash	1	1024.33	1754.93	.0205	.02	.01	0	.26
Boat rental voucher	150	5	9.58	.015	.01	0	0	.21
Building brick	100	10.62	18.15	.0212	.02	.01	0	.27
Agricultural fertilizer	300	.18	1.71	.0011	0	0	0	.2
Irrigation motor (diesel-fuelled water pump)	15000	.13	.48	.0384	.04	0	0	.97
Construction (unskilled) labour voucher	700	.65	4.86	.0091	.01	0	0	.82
Chicken feed	500	1.89	4.7	.0189	.02	0	0	.5
Goat / Cow feed	450	.36	4.87	.0032	0	0	0	.96
Young chicken / duck (female)	1000	.87	3.24	.0174	.02	0	0	.83
Young goat (female)	4500	.17	.99	.0154	.02	0	0	1
Young cow (female)	25000	.05	.25	.0259	.03	0	0	1
School uniform	800	.79	3.52	.0126	.01	0	0	.82
School tuition fee voucher	450	.74	2.76	.0066	.01	0	0	.36
CNG fuel	650	.28	2.16	.0036	0	0	0	.57
Downpayment voucher for CNG vehicle	100000	.02	.09	.0447	.04	0	0	.98
Downpayment voucher for Easybike	50000	.04	.16	.0421	.04	0	0	1
Vehicle annual repair voucher	10000	.09	.48	.0175	.02	0	0	1
Downpayment voucher for Rickshaw	5000	.11	.61	.0114	.01	0	0	.62
Laptop computer	35000	.05	.19	.0329	.03	0	0	.99
Debt payment voucher	12000	.11	.47	.0256	.03	0	0	1
Diesel fuel	800	.11	.66	.0017	0	0	0	.13
Fish feed	550	.46	3.94	.005	.01	0	0	.81
Fishing rod	1000	.36	2.57	.0071	.01	0	0	.75
Fruit tree seedling	200	3.77	12.31	.0151	.02	0	0	.82
Fertile soil	200	.31	3.9	.0012	0	0	0	.34
Dowry / wedding cost payment voucher	100000	.03	.09	.0551	.06	0	0	1
New household fridge	30000	.04	.17	.0215	.02	0	0	.94
Voucher for household gas bill	1350	.31	2.26	.0085	.01	0	0	.85
Downpayment for purchase / lease of arable land	12000	.04	.25	.0105	.01	0	0	1
Voucher for construction of toilet	15000	.11	.48	.0329	.03	0	0	1
Mobile phone	8000	.1	.51	.0161	.02	0	0	.99
Mobile phone data plan	450	.93	3.6	.0084	.01	0	0	.59
Motorbike downpayment	50000	.04	.17	.0441	.04	0	0	1
Motorbike helmet	1200	.28	2.2	.0067	.01	0	0	.98
Electric air compressor	40000	.04	.17	.0326	.03	0	0	.96
Car battery	15000	.13	.48	.0404	.04	0	0	1
Rickshaw battery	7000	.16	.72	.023	.02	0	0	1
Bike	12000	.13	.53	.0311	.03	0	0	.99
TV	20000	.05	.21	.0184	.02	0	0	.75
Radio	600	.84	4.61	.0101	.01	0	0	.81
Wifi router	2000	.27	1.55	.0108	.01	0	0	.81
Cement	1200	1.8	3.51	.0433	.04	.01	0	.89
Bed / Wardrobe / Table	10000	.08	.42	.0163	.02	0	0	1
Handloom machine	13000	.09	.38	.0227	.02	0	0	1
Paddy threshing machine downpayment	10000	.09	.49	.019	.02	0	0	1
Sewing machine	7000	.12	.52	.0166	.02	0	0	.99
Petrol fuel	900	1.32	4.72	.0238	.02	0	0	.97
Firing wood	600	2.85	5.7	.0342	.03	.01	0	.70
Doctor visit voucher	500	.61	3.03	.0061	.01	0	0	.55
Tubes for well construction	12000	.04	.26	.0103	.01	0	0	.68
Shop inventory refill voucher	15000	.08	.37	.024	.02	0	0	1

Note: Output of Gaussian Process estimation applied to empirical data of 550 Bangladeshi respondents. For every of the 51 good types, this table shows summary statistics on the content of the respective good in respondents' policy bundles (as per Gaussian Process estimation).

Figure 21: Expenditure shares in policy bundles
(non-zero goods content only)





Note: Mean expenditure share by good type across all policy bundles with non-zero content of respective good.

4 Conclusion

In this chapter, I have presented the derivation of a novel method based on Bayesian machine-learning and its empirical application in consumer preference elicitation. Using a simplified example, I have highlighted its comparative strength in contrast to simulated Maximum Likelihood Estimation – a widely utilised techniques in the pursuit of such elicitation. I have illustrated the dependence of MLE on correct functional form assumptions – the verification of which is costly in an empirical setting. Gaussian Processes, on the other hand are able to achieve similar results with substantially lower degree of ex-ante rigidity, while also allowing for simultaneous high-dimensional preference queries. In a nutshell, I demonstrate that Gaussian Processes espouse three core advantages over comparable existing methods: i) Non-parametric estimation – Gaussian Processes are able to generate empirical predictions with a very low statistical / economic assumption load; ii) Nuanced approach to uncertainty – Gaussian Processes represent a natively Bayesian evaluation of data with a "built-in" measure of uncertainty through the associated Gaussian distribution; iii) Gaussian Processes maintain excellent robustness to high-dimensional applications as evidenced by the queried 51-good universe.

Subsequently, I apply this elicitation technique based on Gaussian Processes to a novel empirical dataset on consumer preference of microfinance recipients in Bangladesh. This dataset contained sparse information of 30 instances of binary bundle comparisons (each bundle consisting of two types of goods) per individual, querying over a universe of 51 types of goods. This stated-preference data was cheap and fast to collect. With currently employed techniques, it would be highly challenging to model an underlying preference relation for 500 individuals – even with the use of very rigid and likely inaccurate assumptions. With the use of Gaussian Process

estimation, I was able to construct a model of every participants' preferences over the goods universe – and I propose a policy bundle recommendation based on this model. This bundle encompasses the set of goods (subject to a budget constraint) which most likely represent a utility-maximising allocation to the participant. There is a clear path to applications in applied research as well as policy-making for this output.

Illustrating the concrete nature of the use case, my empirical estimates support previous findings in the development economics literature. For example, I show that this population of microfinance participants exhibits preferences which, on the intensive margin, lean into chunky capital goods as part of their respective policy bundles. This is in line with the core economic mechanism underlying to microfinance interventions: a large set of would-be entrepreneurs have an underserved demand for goods that may require some access to capital markets, limiting potentially profitable investments in such goods. Consumer preference elicitation will continue to be an important component of much of the applied microeconomics literature. With the method featured in this chapter, I make the case that economists ought to leverage the support of a hitherto underutilised part of the statistics literature – in doing so, they not only unlock additional datasets but also generate new quantitative insights. However, there is still potential beyond that – as I have highlighted, Gaussian Processes show distinct promise to provide an easily trainable "goldfish glass" of economic decision-making. Much like having a marine ecosystem in a compact and versatile container, Gaussian Processes have demonstrated the potential to capture complex preference relations of real economic agents and make them accessible to detailed economic study. This contribution is hopefully a first step in that direction.

References

Alemu, M. H., Morkbak, M. R., Olsen, S. B., & Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and resource economics*, *54*, 333–359.

Angelucci, M., De Giorgi, G., & Rasul, I. (2018). Consumption and investment in resource pooling family networks. *The Economic Journal*, *128*(615), 2613–2651.

Arrow, K. J., Sen, A., & Suzumura, K. (2010). *Handbook of social choice and welfare* (Vol. 2). Elsevier.

Athey, S., Imbens, G., & Wager, S. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

Baiardi, A., & Naghi, A. A. (2024). The value added of machine learning to causal inference: Evidence from revisited studies. *The Econometrics Journal*, *27*(2), 213–234.

Banerjee, A., & Duflo, E. (2020). *Good economics for hard times: Better answers to our biggest problems*. PublicAffairs.

Borrellas, P., & Unceta, I. (2021). The challenges of machine learning and their economic implications. *Entropy*, *23*(3), 275.

Bujosa, A., Riera, A., & Hicks, R. L. (2010). Combining discrete and continuous representations of preference heterogeneity: A latent class approach. *Environmental and Resource Economics*, *47*, 477–493.

Carlsson, F. (2010). Design of stated preference surveys: Is there more to learn from behavioral economics? *Environmental and Resource Economics*, *46*, 167–177.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and resource economics*, *37*, 181–210.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Clithero, J. A., Lee, J. J., & Tasoff, J. (2023). Supervised machine learning for eliciting individual demand. *American Economic Journal: Microeconomics*, *15*(4), 146–182.

Cunningham, C. E., Deal, K., & Chen, Y. (2010). Adaptive choice-based conjoint analysis: A new patient-centered approach to the assessment of health service preferences. *The Patient: Patient-Centered Outcomes Research*, *3*, 257–273.

Gal, Y., Van Der Wilk, M., & Rasmussen, C. E. (2014). Distributed variational inference in sparse gaussian process regression and latent variable models. *Advances in neural information processing systems*, *27*.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation research part B: methodological*, *44*(6), 735–752.

Hensher, D. A., & Johnson, L. W. (2018). *Applied discrete-choice modelling*. Routledge.

Hossain, T., & List, J. A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, *58*(12), 2151–2167.

Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management science*, *62*(12), 3393–3411.

Khan, A. Q., Khwaja, A. I., Olken, B. A., & Shaukat, M. (2022). *Rebuilding the social compact: Urban service delivery and property taxes in pakistan* (tech. rep.).

Ledyard, J. O., et al. (1994). *Public goods: A survey of experimental research*. Division of the Humanities; Social Sciences, California Inst. of Technology.

List, J. A. (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *The Journal of Political Economy*, *114*(1), 1–37.

Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys*, *25*(2), 363–370.

McFadden, D., & Train, K. E. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, *15*(5), 447–470.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning* [Available at rlhttps://gaussianprocess.org/gpml/chapters/]. MIT Press.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Varian, H. R. (2014). *Intermediate microeconomics with calculus: A modern approach*. W.W. Norton & Company.

Viceisza, A. C. (2016). Creating a lab in the field: Economics experiments for policymaking. *Journal of Economic Surveys*, *30*(5), 835–854.

Chapter III
Heterogeneous Agent Matching and Microfinance Contract Structure in Bangladesh

Abstract

I compare debt-based and equity-based microfinance contracts in an incentivised lab-in-the-field experiment with Bangladeshi freelancers. Building on the literature that stresses the need for understanding agent heterogeneity to predict microfinance performance, I show that equity-based contracts overperform debt-based instruments in an environment that requires partner matching with noisy ability signals. Using structural estimation, I demonstrate that microequity contracts enabled participants to respond to exogenously added signal noise by adjusting the rules of thumb they used towards behaviour more in line with a rational expectations framework – while conventional microfinance contracts did not.

1 Introduction

A substantial academic literature on the potential and limitations of microfinance has grown over the last two decades. As is frequently cited as a reason for the sometimes moderate effects of classic microfinance, relaxing the rigid and structured repayment schedules of microcredit contracts can contribute to the development of economic complexity in microenterprises (Field et al. (2013), Barboni (2017), Barboni (2024)). This literature has shown that added flexibility significantly increases microenterprises' bulky investments (for example in new machines) – which are associated with entrepreneurial activity and long-term growth. Broadening the idea of repayment flexibility – and the possibility that "the structure of existing microfinance contracts themselves may discourage risky but high expected-return investments" – Fischer (2013) proposes an equity-like microfinance contract. That is, a borrower and a lender agree to share output and risk of an investment dynamically.[30] A microequity contract is then a financing instrument that encapsulates this through mandating performance-contingent payments from the party receiving the funds. Even though such microequity contracts bring with themselves a range of potential issues (such as free-riding or overly risky choices), Fischer (2013) notes that participants with microequity contracts were particularly likely to increase risky, high expected-return investment. More recent work on microequity – proving the feasibility of such contracts in the field – includes De Mel et al. (2019) and Meki (2024).[31]

[30]Early theoretical considerations of the importance of microequity models include Pretes (2002) and Ayayi (2012). However, these do not dedicate themselves to empirical tests of hypotheses related to microequity interventions.

[31]Special consideration has also been granted to non-interest-based microfinance contracts in the light of religious restrictions. Randomised control trials examining microfinance instruments addressing such preferences – for example bank-insured rotating savings and credit associations (RoSCA) in El-Gamal et al. (2014) – find broadly positive effects, comparable with the results of classical microcredit elsewhere.

In this chapter, I build on the emerging microequity literature by studying an incentivised lab-in-the-field experiment, which I designed and conducted with a population of online graphic-design freelancers in Bangladesh. I test participants' willingness to engage in collaborative division of labour under both a classic microcredit and a microequity regime. In this experiment, I simulate a partner-matching setting. Participants were asked to complete an incentivised task whose success depends in a complementary way on both English and graphic-design skills. In particular, I aim to elicit participants' preferences for matching with other participants while their (joint) output is distributed based on either microequity or microcredit contracts, respectively. Participants could submit preferences on whom they would like to match up with based on test scores. Test scores were a result of an ability elicitation exercise: a series of multiple-choice questions designed to provide a measure of each participant's English and graphic-design skills. This partner-matching structure introduces the risk of drawing a partner whose ability would lower one's profits. Microequity contracts could be an instrument for participants to mitigate this risk – leading to an increase in the willingness to commit to partners compared to microcredit. I identify separately the effects of (i) equity-based instead of debt-based contracts and of (ii) increasing participants' uncertainty about the ability of potential partners. I elicit and analyse a rich set of outcome variables capturing many dimensions of participants' behaviour when matching with potential partners and their subsequent performance. In order to uncover the economic mechanisms behind participant choices, I further employ Maximum Simulated Likelihood estimation of a structural mixture model. I find the underlying prevalence of two archetypes of decision-making patterns among experiment participants. The relative share of archetypes in each sub-sample has an important effect on aggregate production efficiency and is impacted by the type

of collaboration mode put in place by the financial regime. These findings help understand how equity-based microfinance interventions can contribute to persistently positive outcomes for its participants.

As is mentioned by Meki (2024), microequity contracts have the potential to not only increase the average agent's treatment outcomes but they may also be a particularly well-suited response to agent heterogeneity. The incentive structure of microequity could directly address the shortcomings of microcredit in stimulating entrepreneurship and therefore address an important limitation of conventional microfinance. Microfinance has been a prominent topic of the development economics literature during the last three decades and its promise has been primarily based (for example by Morduch (2013) or Kraay and McKenzie (2014)) on enabling very small firms in low-income countries to access capital. This type of firm – often only employing its owner – is particularly key to the economic output in many developing countries (UNSD (2015)). One barrier to economic development could then be an inefficiently low supply of capital to profitable firms, limiting the economic growth perspective of up to 1.4 billion adults who are still operating without formal access to a bank account (CGAP VII Strategy (2023)). This view has motivated an optimistic perspective on microfinance: Connecting such microenterprises to an alternative capital supply could empower them to grow beyond a size limited by capital market frictions. However, recent evidence draws more ambivalent conclusions about the transformative power of conventional microfinance interventions. This paper contributes new, empirical, and causal evidence on how microfinance can foster the emergence of complex economic structures – such as partnerships between the self-employed. A significant obstacle to microentrepreneurs engaging in more productive but also more complex modes of production is uncertainty about the ability

of potential partners. As showcased by an emerging literature (Meki and Quinn (2024a), Bari et al. (2024)), equity-based contracts – as opposed to the debt-based contracts most commonly used in microfinance – could help firms overcome this obstacle by providing insurance against said risk.

Beyond limiting the success of individual entrepreneurs or policy interventions, frictions in partner matching could lead to a structural impediment of economic growth. To explain why, consider a single-person firm faced with two production functions: a simpler one which the owner can operate in isolation, and a more sophisticated one that draws on two complementary skills. Aggregating the labour of two workers into one production process then has benefits beyond market exchange and the utilisation of comparative advantages. Such aggregation of complementary skills could be crucial for explaining long-term variation in growth as was shown by Kremer (1993). The reliance on the sequential completion of many tasks when producing complex economic goods could help explain the persistence of poverty. Thus, even a country of highly-skilled upholsterer will find it hard to produce furniture if, for some reason, it only has access to clumsy carpenters. While it would be technically possible for this country to produce furniture, each upholsterer might be better off producing advanced furniture instead of exposing herself to the low-quality input of her carpenter colleagues. On a microeconomic level, matching agents with different skills would then be a critical step in enabling complex production functions in firms. This matching could be a necessary condition for the kind of "transformative" economic development microcredit was expected to engender. Conversely, if for example imperfect observability of agent ability interferes with the matching process, Kremer (1993) shows that multiple possible equilibria can exist. On the level of economies, overcoming matching difficulties could be necessary for moving from the low-growth

to the high-growth equilibrium. Accordingly, if microenterprises are to be supported through microfinance, the encouragement of matching even in noisy environments ought to be part of a policy's goals. However, a classically rigid microcredit contract might then be counter-productive for achieving this. The prospect of being locked into a fixed repayment schedule may leave agents unwilling to engage in the risk of matching up with a partner of uncertain ability. If in addition highly skilled agents perform particularly well on their own, a partner matching market with noisy ability signals could face the threat of a breakdown in loose allegory to Akerlof (1970)'s market for "lemons". An equity-based contract, on the other hand, provides built-in insurance against drawing a "lemon" as a partner. A repayment that depends flexibly on both partners' performance absorbs a share of ability shocks and smooths the bottom line. This could keep enough participants in a matching market to prevent it from breaking down. Besides alleviating frictions related to the matching process itself, microequity could also plausibly facilitate match-making by positively impacting the group-dynamics that most directly align economic incentives. This represents another key insight provided by Field et al. (2013). Their findings suggest that one core driver of entrepreneurial activity in a setting of group-based microfinance is that such groups are formed based on equal contributions. They find that members under these circumstances exert a higher level of effort and willingness to share. The presence of free-riding members, however, discourages behaviour associated with entrepreneurship, lowering productivity and output. Similar points are made in Fischer (2013): Under contracts that include equity treatments, participants were willing to contribute more effort and resources, increasing both risk and reward, and decreasing free-riding behaviour.

The remainder of this chapter is structured as follows. First, I detail the set-up of the experiment I designed and implemented. Then, I provide reduced-form analysis of key outcomes under different treatment settings – followed by Maximum Simulated Likelihood estimation of deep structural parameters guiding participants' matching behaviour. I conclude by outlining the implications for policymakers, especially with regards to empowering small firms through microfinance.

2 Data

I constructed a sample of 90 young adults – resident in or around the city of Gaibandha, Bangladesh – who had previously enrolled in a series of vocational graphic-design courses, consisting of three lectures and three supervised practice sessions each week over a duration of three months.[32] Both the lectures and the practice sessions focused on developing participants' technical skills in computer-aided graphic design. No English language training was provided. This population is highly suited for studying the research questions outlined above due to 1) its place within a developing-country context; and 2) its exposure to participants who had been pre-screened with regard to their entrepreneurial preferences. [33]

2.1 Experimental Design: Principles

Most of the participants have marketed their services on online platforms after participating in the aforementioned training courses. Many commit substantial efforts towards attracting business online and can be identified as microentrepreneurs.[34] While some participants have been successful at this and manage to generate a considerable income, a large share cannot rely on their online activities alone. Anecdotal evidence suggests that particularly a lack of English communications skills could be the reason why certain participants had had limited success. Nevertheless, at least some participants seem to possess all the necessary skills to successfully interact online and trade with international clients. In this regard, the allocation of skills among

[32] See Bakhtiar et al. (2019) for a preliminary project brief on the website of PEDL at CEPR.

[33] This population – with some degree of fluctuation – the same as the one featured in Mazumdar (2019), but with a different set of research questions.

[34] A brief survey conducted with every participant as part of the experiment revealed that, on average, participants generated 27 % of their household income through online graphic-design services.

this population could be seen as the type of gridlock described in Kremer (1993). Participants can offer their graphic-design skill on lucrative international markets and move to a different production function – but only if they have access to English skills. In absence of this complementary input factor, they may have to find scarce local employment or they may not even be able to utilise their graphic-design skills at all (for example by working in agriculture). One way of escaping this would be to acquire English skills, but this is usually very costly in absence of well-funded educational facilities. Another way would be to partner up with someone whose skill set complements one's own. To examine partner-matching behaviour, I implement a mechanism within the experiment through which participants could submit preferences on whom they would like to match up with based on test scores. (Test scores were a result of an ability elicitation exercise: a series of multiple-choice questions designed to give an impression of each participant's English and graphic-design skills.) After an algorithm[35] had formed matches, where possible, based on the manifested preferences, participants then worked on an incentivised task. This task consisted of the same types of questions already used in the ability elicitation. Participants were remunerated based on the number of questions answered correctly in this task. Crucially, for two participants in a match both partners' results entered into their pay-off function in a complementary way. If participants stayed unmatched, only their own results were relevant for pay-off calculation, but they did incur a penalty compared to matched-up participants. That way, in general matching up with a high-ability participant was preferable to staying unmatched which was in turn preferable to matching with a low-skill participant. The financial frame for these partnerships

[35]I used a random serial dictatorship (RSD) algorithm to aggregate preferences into matches as part of this experiment. Its make-up and implications are discussed in the section below in detail. Most importantly, under RSD participants are incentivised to reveal their true preferences.

was built around different microfinance contracts: At the beginning of each round, participants would borrow an investment from a stylised banking party (representing cost such as market access expenditure, equipment rent or financing of additional training). One treatment arm was given a microcredit contract – meaning that both participants owed a fixed amount regardless of their and their partner's performance. Another treatment arm was administered a microequity contract – meaning that participants repaid a share of their earnings. For both microequity and microcredit contracts, I ran a "clean" arm each – where ability signals are presented unchanged – and a noisy arm each – where ability signals are "jammed" through addition of random shocks. In the following, I detail how exactly I designed each part in each treatment arm.

I implemented the structure described above in an automated, standardised, and computer-driven game which can be split into four main stages. In chronological order, each experimental session would progress through the following set of stages:

Stage 0 – Ability Elicitation After reading a framing message presenting daily-life examples of comparative advantage participants might be familiar with, a written explanation of the game's rules followed. Participants then answered a brief personal questionnaire covering relevant fields of potential heterogeneity.[36] Next, I asked each participant to answer five questions related to graphic design as well as five questions related to English business communication skills. Graphic-design questions were pooled from a set of two types of questions based on two online games avail-

[36]Topics of questions asked included: age, gender, main occupation, mean household income, past success rate as a graphic designer, risk aversion preferences. The latter were elicited via an incentivised lottery as introduced by Gneezy and Potters (1997). That is, participants were given an endowment of 100 points and subsequently asked which share of it they would like to invest in a project that could yield either no return or 2.5 times the investment with equal probabilities.

able for free to test users' skills as graphic designers.[37] While these games were designed primarily with entertainment in mind, similar tools are used in recruitment for graphic designer positions. To elicit English communication ability, two types of questions were used. Here, participants were asked to evaluate real online messages between clients and graphic-design students affiliated with the training program through which participants were recruited.[38] Each of the ten questions in Stage 0 were mandatory for all participants. Participants were given up to 30 seconds to answer each question.[39] From the answers submitted by participants, ability scores were generated for each of the two skill categories by calculating the share of questions answered correctly. Ability scores were assigned to participants for the duration of an entire session.

Stage 1 – Matchmaking First, participants were presented with another framing page informing them of: (i) the fact that a new period had started; (ii) the details of the financial contract valid under the current treatment; (iii) depending on the treatment a reminder that ability signals may be affected by random shocks in this period. Participants were then presented with a matching table with an entry for all nine other participants in the current session. Each entry contained:

- A four-digit code that was re-randomised in each period, preventing any reputation effects.
- The two ability scores that the respective participant had scored in Stage 0 of the current session.

[37] For details about the graphic-design questions, see Appendix B.1.

[38] For details about the English questions, see Appendix B.2.

[39] A blank response would automatically be submitted if this threshold was reached. In total, fewer than 1% of all presented questions were submitted as blanks (exclusively because of timing reasons). Blanks were counted as incorrect.

- An input field to select a preference rank to be assigned to this entry. Alternatively, participants could also select that they do not wish to match up with this entry (an option labelled "No match").

- An input field to choose the role the inputting participant would want this entry to take on if they should end up matched. That is, whether they would want this potential partner to work on activities related to graphic design or English communications (with the implication that the choosing participant would work on the type of task not assigned to the potential match).[40]

Simultaneously, participants were asked which role they would want to take on in case they ended up unmatched.[41] Aggregating the preferences of all participants in the current session, a random serial dictatorship algorithm (RSD)[42] then assigned matches where it was possible. In order to do so, the algorithm drew a random permutation of participants and iteratively matched any participant i to his or her

[40]Participants could also indicate that they were indifferent with regards to which task a potential partner would work on if matched – if this particular match would end up being created, tasks were assigned randomly.

[41]With the options being graphic-design task, English task or indifference between the former and the latter. In case a participant that signalled indifference would end up unmatched, they would be assigned to a graphic-design or English task with equal probabilities.

[42]This class of algorithm was popularized by Abdulkadiroglu and Sönmez (1998) for allocating indivisible goods among a set of agents. They show that an RSD algorithm is functionally equivalent to a random endowment allocation in the sense that their cores are identical. Thus, RSD inherits a number of desirable properties including being ex-post pareto efficient, anonymous, symmetric, and strategy-proof. In this regard, it carries properties comparable to a Gale-Shapley algorithm (as outlined in Gale and Shapley (1962)) with the exception that RSD does not guarantee stable matches. In the context of the game presented in this paper the latter is rather inconsequential as matches would be redrawn in every period anyway. RSD is, however, less computationally demanding and arguably a more plausible description of the formation of partnerships without any binding legal framework. Crucially, it should be noted that the RSD makes the revelation of one's true preferences a dominant strategy. There is no leeway for tactical manipulation since participants cannot possibly anticipate on which part of their submitted chain of preferences the mechanism will actually act. Note that RSD implies that one's preferences need not necessarily be respected when forming matches – as they do not rely on mutual expression of said preferences, but instead give way to those of the participant higher up in the random order of dictatorship.

most preferred pick j still in the pool of available matches. Upon being matched together, both i and j were removed from the pool. When a match was built by serving i's preferences, j's role in the current period corresponded to the role specified by i simultaneously with submitting his or her matching preferences. Participant i took on whatever task j would not work on.[43]

Stage 2 – Task Execution Next, participants were informed about whether they had been assigned a match or not and – if matched – with whom (recall that identifiers were re-randomised in each period) and who would work on which task. Each participant would then go on to answer five questions related to the type of task they were assigned to in Stage 1 (English or graphic design). These questions were drawn from the same respective pools as the ability elicitation questions in Stage 0 and presented in an identical way. A maximum of 30 seconds was allotted to each question.

Stage 3 – Profit Calculation Finally, participants were presented with a tabular pay-off calculation, outlining their own and their partner's performance (the latter only if a participant was matched in the current period) as well as the implied repayment under the relevant contract, and their own net pay-off for this period. Following Kremer (1993), the skill inputs would then be aggregated in an "o-Ring"-type multiplicative production function for matched pairs. This introduces a complementarity of skills when advancing to the more complex production function. If participants remained unmatched – voluntarily or not – they could produce while using only one skill at the cost of a potentially lower output.

[43] If the assigning participant i expressed indifference between the two tasks for potential match j, tasks would be distributed randomly with equal probabilities.

- Under the microcredit regime:

$$\pi_{i,t} = N_{i,t} B \prod_{j=1}^{N_{i,t}} b_{j,t} \; - D(1+r) \tag{23}$$

- Under the microequity regime:

$$\pi_{i,t} = N_{i,t} B (1-s) \prod_{j=1}^{N_{i,t}} b_{j,t} \tag{24}$$

where $\pi_{i,t}$ is the profit of participant i in period t; $N_{i,t}$ is the complexity of the production chain – in this case assuming value 1 if unmatched and 2 if participant i was matched in period t; B is the added value generated through completion of each task; $b_{k,t}$ is the share of questions answered correctly by participant k in period t – being the principal performance measure of task execution; D and r are the debt incurred and interest rate under the microcredit regime; s is the share of revenue used for capital compensation under the microequity regime.[44]

A brief example shall illustrate the gridlock that could ensue from these multiplicative production functions. Assume participant i has b-values of 0.2 and 0.8 for English and graphic design respectively. Under microequity, participant i could expect to achieve a profit of 560 experimental currency units (ECU) when producing

[44]Experimental parameters were calibrated as follows: $B = 1000$, $D = 100$, $r = 0.1$, $s = 0.3$. This was done with the intent of (i) balancing average bank earnings across contract types based on average ability levels achieved in a pilot run of this experiment and; (ii) scaled to ensure an easy translation of experimental currency units (ECU) into a cash payout in Bangladeshi Taka (BDT). The latter exchange rate was: $1000\ ECU = 50\ BDT$

unmatched. Should i, however, be able to compensate for his or her own low English ability by partnering up with j, who has an English skill of 0.8, then they could jointly generate profits of 896 ECU each. There is one complication: If participant i fails to find a match with an English skill of more than 0.5, he or she will always be better off staying unmatched. Thus, even if the population features many highly skilled designers – as an ability of 0.8 would suggest – participants will have to forgo large profits if too few have English skills above a certain threshold, too. Furthermore, if the ability signals foreshadowing b-values should be noisy, one might think twice before marking down as a match a participant whose desired skill is close to the threshold. Matching based on noisy ability signals could therefore exacerbate the economic lock-in that is possible under multiplicative production functions. Thus, matching with high-ability partners is generally preferable to having no partner, but that is preferable to matching with low-ability partners. To test these predictions empirically, I implemented a total of four different treatment arms with variations along two axes. As outlined in Table 15, the profit function – and especially the contract type – employed would either follow the microcredit or microequity scheme outlined in Equations 22 or 23, respectively. Second, treatments varied by whether ability signals presented in Stage 1 were unchanged ("clean") or subject to random shocks ("jammed").[45] Accordingly, I term the four treatment arms: DEBT-CLEAN, EQUITY-CLEAN, DEBT-JAMMED, and EQUITY-JAMMED.

[45]These random shocks were i.i.d. and drawn from a uniform distribution spanning from -10 to +10 percentage points. Recall that ability signals were presented in the form of the share of correctly answered questions. Thus, signals could assume values from 0 to 100 percent.

Table 15: Overview of treatment arms

		Capital Regime	
		Microcredit (debt-based)	Microequity (equity-based)
Ability Signals	Clean	*DEBT-CLEAN* N_1=70	*EQUITY-CLEAN* N_2=70
	Jammed	*DEBT-JAMMED* N_3=70	*EQUITY-JAMMED* N_4=70
N_{Total}**=280**			

Note: Debt-based / microcredit contracts required repayment of a fixed amount of experimental currency units per period; equity-based / microequity contracts required the repayment of a constant fraction of generated revenue in each period. Under "jammed" treatments, ability signals displayed during matching in Stage 1 were subjected to a random, uniformly distributed, i.i.d. shock drawn from a support ranging from -10 to +10 percentage points. Under "clean" treatments, ability signals were presneted unchanged. Each treatment was weighted equally and contains 70 participant-level observations.

Note: These 280 session-level observations were generated by the 60 (physical) individual participants.

2.2 Study Sample

I use the software package *oTree* as a technical platform for this experiment. *oTree* was introduced by D. L. Chen et al. (2016) and allows an automated and standardised execution of experiments in the field. Appendix B contains screenshots of the most important stages of the *oTree* application I wrote for this experiment.[46] This software enables the participants to interact with each other via a standard web browser interface – which in the case of this study was running on handheld tablet computers with internet access.[47] In order to make an economic interpretation of this experiment plausible, I directly incentivised participants' decisions. That is, participants collected experimental currency units (ECU) which corresponded exactly to the accumulated output of the respective profit function (accounting for the repayment implied by the respective microfinance contract). At the end of each session, participants would receive a cash pay-off in Bangladeshi Taka (BDT), the local currency. The amount paid out to participants was a linear transformation of the ECU generated throughout the game such that 1000 ECU would correspond to 50 BDT.[48] In addition, a fixed show-up fee of 300 BDT was paid in any case to every participant. Consequently, participants earned an average payment of 457 BDT[49] per day (compared to a median self-reported monthly income of 10,000 BDT).

I conducted a total of 28 sessions, in person, over the span of three weeks in September and October 2019. Each day of experimenting followed the following structure, starting at around 9 am:

[46]The version actually used in the field contained text directly translated from English into Bengali.

[47]In the case of a power outage – which would occur multiple times a day on a regular basis in Gaibandha – the game had to be briefly paused to switch the tablet computers into mobile internet mode. This was only necessary during one session.

[48]At the time of experimentation, this corresponded to approximately 0.48 GBP or 0.59 USD.

[49]At the time of experimentation, this corresponded to approximately 4.35 GBP or 5.38 USD.

15 min Welcome message, recording of participants' consent, explanation of experiment mechanics. See Appendix C for translated details of conveyed communications.

30 min Test run: Game was played with one period only. Any points collected during this test run did not affect participants' pay-off (as communicated to participants). The specifications of the game corresponded to the treatment assignment of the subsequent main run.

45 min Session 1: Game was played starting with Stage 0 and then with Stages 1, 2, 3 being repeated three times in series. Treatment assignment corresponded to the one of the first session scheduled for this particular day.

20 min Break

45 min Session 2: Game was played starting with Stage 0 and then with Stages 1, 2, 3 being repeated three times in series. Treatment assignment corresponded to the one of the second session scheduled for this particular day.

15 min Pay-off distribution and farewell message.

Running two sessions back-to-back on each day of experimenting allowed the evasion of most power outages, which would frequently occur in the early afternoon. Both sessions on any given day featured the same participants with no attrition between sessions. Furthermore, to enable participants to gain a familiarity with experiment mechanics, every session consisted of a single instance of Stage 0 followed by three repetitions of the sequence of Stage 1 through Stage 3. Treatment assignment – and thus the contract type – was constant for all three repetitions of any given session.

In order to construct a subject pool for this study, all former participants of the graphic-design courses mentioned above were contacted in an order which would be re-randomized daily. New pool members were contacted until a total of ten commitments were secured for the day. In order to limit travel costs, the same set of participants were used for two consecutive sessions (taking place on the same day). Therefore, this paper implements a *within-subject* design, meaning that depending on the respective random draws one participant could be exposed to multiple treatments at different times. Indeed, each member of the initial subject pool participated on 0 to 6 days.[50] A total of 88 participants were in the pool of potential subjects. Out of that, 60 participants participated at any point – that is, 25 % of the contacted sample declined to participate at all. A total of 280 session-level observations were generated by these participants, indicating that every individual participant generated 4.7 observations, on average. Before any experimentation started, each of the 28 planned sessions were randomly pre-assigned to one of the four treatment arms. Treatment assignment of sessions was stratified by the number of treatment arms: Each set of four consecutive sessions were assigned a treatment by drawing without replacing a random order of the four treatment arms. This implies that each treatment received an equal weight in terms of generated observations (in line with the absence of a prior regarding treatment effects).

While within-subject experimental designs constitute a more efficient way of gathering data, it incurs potential threats of the internal validity to the insights it produces. Below, I state common bias sources for within-subject experiment designs and I summarise its potential impact on this specific experiment.

[50]The median number of days attended being 1.5 and the average 1.6 days.

- **Order effects**: the sequence in which participants experience treatments can affect their decisions – for example through framing effects. The impact of order effects is unlikely to be material in this experiment since the sequence of treatment arms that multiple-attenders were exposed to is successfully randomised. Therefore, the sequence effect of being exposed first to Treatment 1 and then Treatment 2 is as likely to occur as the inverse case of first experiencing Treatment 2 and then Treatment 1, counterbalancing resulting order effects.

- **Learning / carryover effects**: participants' choices are affected by an increasingly proficient understanding of experiment dynamics. Such an improvement in proficiency is a likely outcome in this experiment – with participants becoming better at solving the "mini-games" at the heart of the design. Nevertheless, the variable most likely affected by this (the ability metrics) is fully measured in this experiment, its outcomes balanced by treatment, and subsequent regression analysis does control for them, singifnicantly reducing the potential for resulting bias.

- **Fatigue effects**: As participants progress through a treatment course, they may become tired, which could reduce their attention, effort, and therefore affect treatment outcomes over time. It is likely that such fatigue did materialise – however, there is no evidence for the degree of this varying by treatment. Furthermore, subsequent regression analysis controls for decreasing attention by including round-fixed effects. It should also be noted that participants are heavily disincentivised by the experimental design from exhibiting fatigue effects.

- **Attrition effects**: estimates may be biased by participants dropping out at a rate that is correlated with treatment outcomes. The impact of this is limited by the absence of any within-day drop-outs – that is, every participated that showed up for an experimenting day stayed for both sessions that took place that day. It is possible that over time, for example less successful participants were less likely to show up a second time. Nevertheless, this would likely imply a deteriorating balance of participant characteristic across treatments over time – which is not what Table 16 suggests.

To test whether this treatment assignment resulted in successful randomisation, I examine the balance of survey variables across treatments. Recall that participants filled out a short questionnaire on basic socio-economic variables at the beginning of each session. This data is displayed in Table 16. The lower panel of Table 16 displays p-values from mean-comparison tests between treatments, accounting for simultaneous testing with a Bonferroni correction.[51] The only variable for which the null-hypothesis of equal means across treatments is rejected is English ability. This is mostly due to the comparatively low scores under EQUITY-CLEAN. While this could be a possible source of bias, it is unlikely that this is compromising my interpretation of treatment effects. In Appendix A Table 21, I show a PDS-LASSO estimate to gauge the impact any potential imbalances might have had, following Belloni et al. (2014). Comparing the lasso estimates to the corresponding plain OLS result from Column (3) in Table 18, there is little divergence. Plain OLS seems to be understating treatment effect sizes for the affected EQUITY-CLEAN treatment. Bearing this in mind, it is therefore safe to proceed with a causal treatment effect analysis.[52]

[51]The same calculations without a Bonferroni correction are displayed in Appendix A Table 20.
[52]Compare also Appendiy A Figure 26 where a Kolmogorov-Smirnov rejects the null-hypothesis of

3 Results

In the following, I use reduced-form analysis to analyse matching behaviour, treatment impact, and participant outcomes.

3.1 Ranking behaviour

Valuation of ability The challenge of finding the best possible match remains unchanged across all contract types, and so do the strategic tools available to participants. What could change depending on treatment assignment, however, is the way participants react to a given set of ability signals. For example, the provision of implied insurance through microequity might lead to a higher propensity to match with participants whose ability signals are close to the threshold of profitability – whereas one might *ceteris paribus* expect a less favourable ranking when ability signals become noisier. To examine these patterns systematically, I estimate the relationship between the ability signals that were presented to others by any participant and the preference rank assigned to that participant by others. I formalise this relationship using a rank-ordered logit model contained in Equation 25. This type of model – also known as Plackett-Luce model – takes an assigned ranking of options as the dependent variable and processes the ordinal information contained within in relation to a set of covariates. In this case these explanatory variables were the visible ability signals.[53] Note that a positive coefficient estimate stands for an increase in relative

equal distributions of English ability at the 1% level for EQUITY-CLEAN vs. DEBT-CLEAN, and at the 5% level for EQUITY-CLEAN versus EQUITY-JAMMED. For graphic-design ability, the null of equal distribution is rejected at the 5% level for DEBT-CLEAN versus EQUITY-JAMMED as well as for DEBT-JAMMED versus EQUITY-JAMMED.

[53] All no-match preferences submitted by a player were assigned jointly the lowest rank numeral not assigned to any positive matching preference. For example, if a participant only wanted to match with four other participants and assigned ranks 1 through 4 to them, the other five participants were treated as occupying rank five together.

Table 16: Balance of covariates

			(1)	(2)	(3)	(4)
		All	**DEBT-CLEAN**	**EQUITY-CLEAN**	**DEBT-JAMMED**	**EQUITY-JAMMED**
English Ability		.44	.5	.36	.43	.47
		[.4]	[.4]	[.4]	[.4]	[.4]
		(.22)	(.22)	(.21)	(.24)	(.21)
Graphic-Design Ability		.66	.71	.62	.69	.62
		[.8]	[.8]	[.6]	[.8]	[.6]
		(.26)	(.25)	(.25)	(.31)	(.23)
Age of Participant		21.68	21.54	22.22	21.2	21.77
		[21]	[21]	[22]	[21]	[21]
		(3.96)	(3.93)	(4.75)	(2.63)	(4.23)
Female		.25	.26	.2	.26	.27
		[0]	[0]	[0]	[0]	[0]
		(.43)	(.44)	(.4)	(.44)	(.45)
Monthly Income (BDT)		11190.66	12329.79	11607.93	9173.14	11651.79
		[10000]	[10000]	[10000]	[9750]	[9250]
		(10195.38)	(10195.8)	(9038.70)	(5441.72)	(14113.81)
Monthly Freelancing Income (BDT)		2397.76	2289.57	3124.33	2458.57	1718.57
		[0]	[0]	[0]	[0]	[0]
		(6467.40)	(6225.19)	(7144.01)	(7401.26)	(4851.78)
Freelancing Jobs Applied		3.55	3.57	4.06	3.19	3.37
		[0]	[0]	[0]	[0]	[0]
		(7.11)	(7.65)	(7.7)	(6.40)	(6.5)
Freelancing Jobs Completed		1.92	2.39	1.72	1.33	2.22
		[0]	[0]	[0]	[0]	[0]
		(4.3)	(5.72)	(3.47)	(2.9)	(4.59)
Risk Aversion		54.22	54.21	57	52.07	53.6
		[50]	[50]	[50]	[50]	[50]
		(29.59)	(30.61)	(28.65)	(28.81)	(30.67)
Observations		280	70	70	70	70

P-value testing	**(1)=(2)**	**(1)=(3)**	**(1)=(4)**	**(2)=(3)**	**(2)=(4)**	**(3)=(4)**	**(1)=(2)=(3)=(4)**
English Ability	0***	.063	.528	.07	.001***	.199	.001**
Graphic-Design Ability	.032	.721	.029	.123	.944	.123	.062
Age of Participant	.363	.549	.741	.122	.559	.341	.434
Female	.424	1	.849	.424	.322	.849	.748
Monthly Income (BDT)	.658	.023	.745	.055	.983	.171	.051
Monthly Freelancing Income (BDT)	.462	.884	.545	.589	.174	.485	.584
Freelancing Jobs Applied	.707	.76	.875	.482	.576	.872	.911
Freelancing Jobs Completed	.408	.172	.847	.473	.476	.178	.396
Risk Aversion	.579	.67	.906	.311	.498	.761	.783

Note: Main figure shown for each variable is sample mean; sample median is in square brackets and sample standard error is in parentheses. P-values in the lower panel refer to the confidence in the null-hypothesis of equal distributions between sub-samples. For the joint tests in the lower panel, significance was calculated using a Bonferroni correction, accounting for the simultaneity of 9 or 27 (in the last column) hypotheses being tested, respectively. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

"logistic valuation" and thus indicates a more favourable ranking:

$$\begin{aligned} & v_{i,t,r} = \beta_0 + \beta_1 ability_{1,r,t} + \beta_2 ability_{2,r,t} + \epsilon_{i,t,r} \\ & \text{s.t.} \quad P(v_{i,t,1} > v_{i,t,2} > ... > v_{i,t,L}) = \prod_{l}^{L} [\frac{\exp(v_{i,t,l})}{\sum_{m=l}^{L} \exp(v_{i,t,m})}] \end{aligned} \tag{25}$$

where $v_{i,t,r}$ is the valuation of participant i in period t towards his or her r^{th}-most preferred match option. $R_{i,t}$ stands for the ranking submitted by participant i in period t. $\epsilon_{i,t,r}$ are assumed to be independent and drawn from an extreme value type I distribution. The β-parameters are obtained through a Maximum Likelihood Estimation, given all ability signals and all submitted rankings.

Corresponding estimation results in Column (1) of Table 17 confirm that ability signals are predictive of participants' ranking choices. Throughout all estimated models coefficients for English and graphic-design abilities are significantly positive – representing a higher ranking valuation on average. This is in line with participants treating ability signals as proxies of input factors for their pay-off. By allowing abilities to interact with a dummy variable capturing whether abilities were above or below the respective median, Column (4) shows that the resulting "ability premium" was furthermore non-linear. There is limited variation of valuation by treatment, as Columns (2) and (3) indicate, with both representing models that interact abilities with treatment assignment.

The imperfect balance discussed in Section 2.2 could, for example, mean that differential availability of ability affected matching market equilibria. The ensuing changes in behaviour could then simply be artifacts of this. This seems not to have been the case: As can be gleaned from Table 17 there is no significant change in how participants valued English ability. With the relative shortage of English ability under EQUITY-CLEAN one might expect that participants adjusted their valuation of

that scarce input upward, if market equilibria were indeed skewed by imbalance. Further, qualitative evidence suggested a more general shortage of English skills among the study population (beyond the outlier of EQUITY-CLEAN). Table 16 confirmed that English ability was at least more scarce than graphic-design ability. Despite this, design ability was valued more strongly on average than English ability, as suggested by the first two rows in Table 17. Allowing for a differential valuation of abilities above median, Column (4) shows that English ability was valued at approximately the same rate as design ability, but only for above-median English ability. This is in line with a good understanding of the multiplicative production function, which implies a minimum ability threshold for participants with whom it makes sense to match up. If participants perceive a general scarcity of English ability, they might accordingly proceed with caution: English ability that is below said threshold was indeed significantly discounted. The much more abundant graphic-design ability was much less likely to fall short of that threshold and was therefore subjected to less scrutiny. These results confirm that participants understood the basic mechanics of a production function modelled after Kremer (1993).

Table 17: Rank-ordered logit models – valuation of ability

	(1)	(2)	(3)	(4)
English ability	1.20***	1.39***	1.32***	1.00***
	(0.10)	(0.21)	(0.23)	(0.12)
Design ability	1.61***	1.48***	1.42***	1.22***
	(0.11)	(0.20)	(0.20)	(0.11)
English ability X EQUITY-CLEAN		-0.06	-0.01	
		(0.31)	(0.33)	
English ability X DEBT-JAMMED		-0.26	-0.17	
		(0.28)	(0.30)	
English ability X EQUITY-JAMMED		-0.48*	-0.53*	
		(0.27)	(0.28)	
Design ability X EQUITY-CLEAN		0.58*	0.60*	
		(0.33)	(0.33)	
Design ability X DEBT-JAMMED		-0.06	0.07	
		(0.27)	(0.28)	
Design ability X EQUITY-JAMMED		0.22	0.37	
		(0.33)	(0.33)	
I{English > median}				0.23**
				(0.09)
I{English > median} X EQUITY-CLEAN				0.02
				(0.14)
I{English > median} X DEBT-JAMMED				-0.21*
				(0.11)
I{English > median} X EQUITY-JAMMED				-0.19*
				(0.11)
I{Design > median}				0.11
				(0.12)
I{Design > median} X EQUITY-CLEAN				0.48**
				(0.19)
I{Design > median} X DEBT-JAMMED				0.09
				(0.14)
I{Design > median} X EQUITY-JAMMED				0.13
				(0.14)
N	7380	7380	4860	7380
Log-Likelihood	-10056.3	-10045.3	-6623.5	-10012.6

Note: The observations in this table are on the level of ranking assignments of the nine other group members each participant submitted in each period. Dependent variable is the ranking of each of those nine possible matches by a participant. A positive coefficient means that the respective covariate, on average, led to a more favourable ranking (represented by a lower preference rank number). Column (3) mirrors the specification in Column (2) but omits observations from each session's first round. $I\{...\}$-expressions are dummy indicators which assume value one if a participant's respective ability was above the global median value of that ability, and zero otherwise. Standard errors in parentheses are clustered at the level of persons. No fixed effects and no control variables beyond those specified were included. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

The correlation between the ability signals and the assigned matching preference rank observed above is a pre-condition for rational participant behaviour but the key question is whether participants' preferences are more or less consistent with profit-maximising decisions. I analyse measures of profit-maximisation for ranking decision. Figure 22 (a) shows the share of all those match requests (match non-rejections) which would have led to higher profits if the participant had stayed unmatched. The calculation of these counterfactual profit metrics uses direct ability signals as presented to participants. That is, a perfectly profit-maximising participant would score a value close to zero on this metric. Comparing EQUITY-CLEAN to EQUITY-JAMMED in Figure 22 (a), one can indeed observe a remarkable shift towards a rejection pattern more in line with optimal participant profits. Note also that an island of participants with rejections almost orthogonal to ability signals remains even under EQUITY-JAMMED. On the other hand, comparing DEBT-CLEAN to DEBT-JAMMED, participants seem to depart from strictly signal-based matching preferences. A more general way of checking participants' signal-to-preference processing is to look at rank correlations of submitted rankings and the rank of expected profits (projected using direct ability signals).[54] The results, displayed in Figure 22 (b), mirror the patterns described above. Participants shift towards preferences in line with optimised expected profits. Meanwhile under a microcredit contract the distribution of correlations seems stable at best with a slight tendency towards orthogonal matching behaviour.[55] While these estimates suggest that participants sub-

[54]Match rejections were treated jointly as part of the next-worse rank underneath the lowest-ranked non-negative matching preference. For example, if a participant only submitted preference ranks one, two and three the rest of their nine possible matches were treated as if they were assigned rank four.

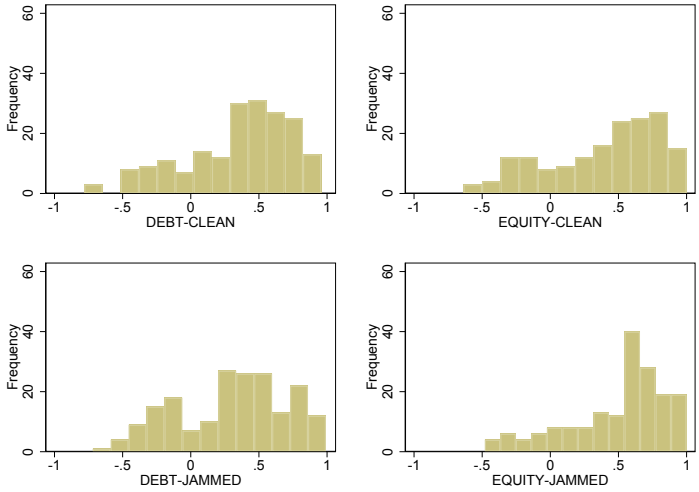
[55]Expressing this same analysis numerically, Appendix A Table 22 quantifies these tendencies as a robustness check. I also display results for the inverse metric of match requests that should not have been made – that is the ratio of match rejections that should not have been rejected. The same differences manifest themselves: When signals are jammed matching behaviour becomes

mitted matching preferences in a way at least partially based on profit maximisation, there appears to be a role for non-rational behaviour as well.[56] One possible explanation could be the usage of rules of thumb, which have been a long-standing concept in the behavioural literature (Kahneman and Tversky (1974), Kahneman and Tversky (1979)). More recently, the use of rules of thumb has been identified in empirical settings (List (2006)) and also specifically in a development economics setting (Banerjee and Duflo (2007), Drexler et al. (2014)). There has been no comprehensive study of such rules in the context of matching markets for microentrepreneurs, but a change in the frequency with which different rules are applied could explain an increase of rational behaviour. I will quantify the prevalence of rule archetypes below.

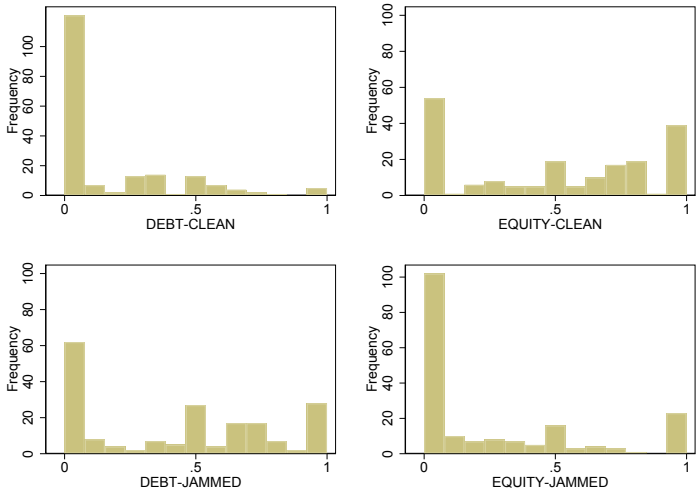
more in line with rational expectations under microequity contracts, while the opposite is true for microcredit contracts. Finally, in the bottom panel I show these effects for the rank correlation of preferences and predicted profits (corresponding to Figure 22 (b)). Here, the increase of rational behaviour under microequity is significant as well, while there is no significant change when jamming under microcredit. These results are robust and significant across a range of different specifications.

[56]For example, it is hard to explain why participants would be more willing to extend match requests when jamming is introduced under the microcredit contract

Figure 22: Matching behaviour – preference-signal consistency



(a) Rank-correlation – signal-predicted profits and preference rank



(b) Ratio of non-profitable match requests

Note: Sub-figure (a) describes the rank-correlation between the ranking of possible matches assigned by participants and the ranking of the maximal expected profits achievable with each possible match if they performed exactly as indicated by ability signals (this includes jamming where applicable). Non-profitable match requests in sub-figure (b) represent share of all match requests where participants declared they wanted to match with another participant, even though they would have been better off on their own.

3.2 Participant outcomes

Previously, I have introduced evidence that participants processed ability noise differently in microequity and microcredit contracts. In the following, I move to analyse how these changes translated into possible policy outcomes – such as participant profits. Recall from Section 2.1 that at the heart of the partner-matching problem is the risk of drawing a partner whose ability would lower one's profits. Through providing a way to mitigate this risk, microequity contracts could lead to an increase in the willingness to commit to partners compared to microcredit, which would also increase participants' average profits. Furthermore, these differences ought to become bigger if ability signals become noisier.

$$y_{i,t} = \beta_{EQUITY-CLEAN} D_{i,t}^{EQUITY-CLEAN} + \beta_{DEBT-JAMMED} D_{i,t}^{DEBT-JAMMED} + \beta_{EQUITY-JAMMED} D_{i,t}^{EQUITY-JAMMED} + X_{i,t} + \sum_{j=1}^{N} \delta_j D_i^{i=j} + \epsilon_{i,t} \tag{26}$$

Table 18 shows the results of the reduced-form regression described in Equation 26. $y_{i,t}$ refers to the respective outcome of participant i in session t; $X_{i,t}$ refers to a set of covariates[57], included where applicable; $D_i^{treatment}$ represent dummy variables that have a value of 1 if participant i was assigned the tagged treatment in session t and 0 otherwise; β_k coefficients represent treatment effects expressed as deviations from the baseline treatment – DEBT-CLEAN; the summation sign contains individual fixed effects using a set of dummy variables $D_i^{i=j}$ which take on value 1 when an iteration over all participants yields participant i and 0 otherwise. As show in Column (4) in the first panel: Compared to the DEBT-CLEAN jammed,

[57]Consisting of: English and graphic-design ability, round fixed effects, participant age, participant gender, income of participant.

every treatment significantly increased average profits per round. While the increase under EQUITY-CLEAN and to a lesser degree EQUITY-JAMMED are well-defined by the proposed mechanism, explaining the effect for DEBT-JAMMED requires additional explanation. Estimates from microequity treatments appear to be robust to smoothing outliers (as Column (6) indicates), while the DEBT-JAMMED treatment is not. When taking the probability of surpassing median profits as an outcome, treatment effects vanish under microequity while remaining marginally significant under DEBT-JAMMED. This suggests that: (i) estimates of average treatment effects hide substantial heterogeneity; (ii) different mechanisms could be at play, which may generate similar average treatment effect estimates in this regression analysis. The treatment effects on profits are more likely a consequence of economic decision-making rather than a possible mechanical consequences of contract design – as is illustrated by there not being a significant effect on bank repayments (especially not between microequity and microcredit contracts). Note that average matching rates appear not to be impacted by any treatment (with the putative exception of EQUITY-CLEAN), as illustrated by the bottom panel. The presence of a robust average treatment effect of microequity regimes on matching rates would have represented clear support for the core initial hypothesis (that is, implicit counter-party insurance through microequity leading to lower matching avoidance). Nevertheless, the absence of such an average effect does not imperil this hypothesis: the quality of matches could have been positively affected without an average effect on the overall frequency of matches. Contrasting the frequency results shown in Table 18 with earlier analysis of matching quality metrics (Figure 22), there is clear support for microequity treatments leading to more desirable matches – even though participants did not necessarily match more frequently.

Table 18: Treatment effects – reduced form

Outcome: Profits per round	Lin.	Lin.	Lin.	Lin.	P(> Med.)	IHS
	(1)	(2)	(3)	(4)	(5)	(6)
Equity-Clean	46.11	88.79*	87.33*	112.96**	0.09	1.30***
	(44.98)	(45.49)	(44.61)	(49.67)	(0.07)	(0.40)
Debt-Jammed	75.20	118.90**	95.87*	124.44**	0.12*	0.64
	(53.59)	(57.62)	(52.90)	(58.49)	(0.06)	(0.45)
Equity-Jammed	66.25*	80.91*	95.07**	98.17**	0.05	1.68***
	(37.70)	(42.71)	(41.66)	(46.06)	(0.06)	(0.36)
Control Mean	419.62	419.62	419.62	419.62	.46	5.06
Outcome: Bank repayment per round	Lin.	Lin.	Lin.	Lin.	P(> Med.)	IHS
	(1)	(2)	(3)	(4)	(5)	(6)
Equity-Clean	-1.54	-6.07	8.87	-0.62	0.46***	-0.70***
	(12.19)	(13.29)	(11.36)	(13.53)	(0.05)	(0.18)
Debt-Jammed	-0.32	3.49	4.42	5.97	0.05*	0.04
	(5.04)	(8.38)	(4.90)	(8.14)	(0.02)	(0.08)
Equity-Jammed	7.09	4.59	12.33	7.50	0.44***	-0.37***
	(11.48)	(11.88)	(11.44)	(12.03)	(0.04)	(0.13)
Control Mean	201.14	201.14	201.14	201.14	0	5.97
Outcome: I{Matched in round}	Lin.	Lin.	Lin.	Lin.	Probit	Logit
	(1)	(2)	(3)	(4)	(5)	(6)
Equity-Clean	-0.00	-0.02	0.09**	0.06	0.36*	0.63*
	(0.05)	(0.05)	(0.04)	(0.05)	(0.19)	(0.34)
Debt-Jammed	0.00	-0.02	0.03	0.02	0.14	0.22
	(0.05)	(0.05)	(0.04)	(0.04)	(0.17)	(0.32)
Equity-Jammed	-0.06	-0.03	-0.01	0.00	-0.02	-0.05
	(0.06)	(0.05)	(0.05)	(0.05)	(0.18)	(0.32)
Control Mean	.81	.81	.81	.81	.81	.81
N	4860	4860	4824	4824	4824	4824
FE	No	Yes	No	Yes	No	No
Controls	No	No	Yes	Yes	Yes	Yes

Note: Profits are defined as production minus bank repayment and refer to each round's change in incentivised score. I{...} represents a dummy variable assuming a value of 1 if a participant was in a match in a given period, and 0 otherwise. Lin. stands for linear regression, $P(> Med.)$ for probability of outcome being above resp. sample median. IHS stands for regression with inverse hyperbolic sine of outcome variable. Control mean refers to the DEBT-CLEAN treatment. Control variables include: English and graphic-design ability, round fixed effects, participant age, participant gender, and income of participant. Table displays coefficients from OLS regressions. Standard errors in parentheses are clustered at the level of persons. Fixed effects are on the level of persons. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

Moving to possible explanations of why jamming ability signals under microcredit increases profits, I move to analyse potential effort effects. Recall that participants were reminded that ability signals were being jammed, whenever that was the case. It is possible that this impacted the framing of the experiment and lead to differential effort provision. It is well-established within the experimental economics literature that small changes in the verbal presentation of an economic incentive can have strong effects on participant behaviour.[58] Effort is challenging to quantify as an input, but can be affected by more elusive mechanics such as framing.[59] In the context of this experiment, effort could enter the profit function during the incentivised tasks in Stage 2. This was styled as a quiz consisting of five questions which participants had to answer under time pressure. A potential proxy for the effort participants exerted during this stage could be the average time spent answering per question. Regression analysis of this outcome is displayed in Table 19, revealing that it is mostly the difference between jammed and clean treatments that drive differences in average time spent on task.[60] Therefore, effort effects did very likely impact participant outcomes for the treatments with jammed ability signals offering an explanation for the increase of average participant profits under DEBT-JAMMED.

[58]See Dreber et al. (2013) looking at a dictator game or Cookson (2000) looking at a public goods game.

[59]Abeler et al. (2011) demonstrates the dependency of effort provision on the shape of expectations, Hossain and List (2012) lists a range of framing techniques known to impact productivity of factory workers.

[60]Appendix A Figure 27 shows this measure split by treatment. Jamming indeed leads to an increase in time spent working on the task on average both under the microcredit and the microequity contract. Furthermore, the microequity contracts compared to the microcredit contracts seem to have attracted slightly more effort overall.

Table 19: Effort – time spent on task

Outcome: Time spent per question (seconds)	Lin. (1)	Lin. (2)	Lin. (3)	Lin. (4)	P(> Med.) (5)	IHS (6)
Equity-Clean	1.32**	1.02*	0.64	0.21	0.06	0.07
	(0.62)	(0.53)	(0.65)	(0.71)	(0.05)	(0.06)
Debt-Jammed	1.97***	1.31*	1.51**	0.99*	0.06	0.11*
	(0.65)	(0.69)	(0.59)	(0.56)	(0.04)	(0.06)
Equity-Jammed	1.96***	1.80***	1.37***	1.33***	0.10***	0.13***
	(0.57)	(0.59)	(0.43)	(0.44)	(0.03)	(0.04)
Control Mean	9.93	9.93	9.93	9.93	.37	2.85
N	2540	2540	2510	2510	2510	2510
FE	No	Yes	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes

Note: Regression of time (in seconds) spent on each of the five incentivised questions (during the task completion in Stage 2). Lin. stands for linear regression, $P(> Med.)$ for probability of outcome being above resp. sample median. IHS stands for regression with inverse hyperbolic sine of outcome variable. Control mean refers to the DEBT-CLEAN treatment. Control variables include: English and graphic-design ability, round fixed effects, participant age, participant gender, and income of participant. Table displays coefficients from OLS regressions. Standard errors in parentheses are clustered at the level of persons. Fixed effects are on the level of persons. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

The increase of average profits under all treatments in the absence of changes in matching rates suggests that treatment effects are subject to heterogeneous patterns. I analyse such patterns by performing quantile regressions.[61] Figure 23 shows the estimates obtained from such a quantile regression of profits per period. As before, effects are expressed as relative deviations from the baseline of the DEBT-CLEAN treatment.

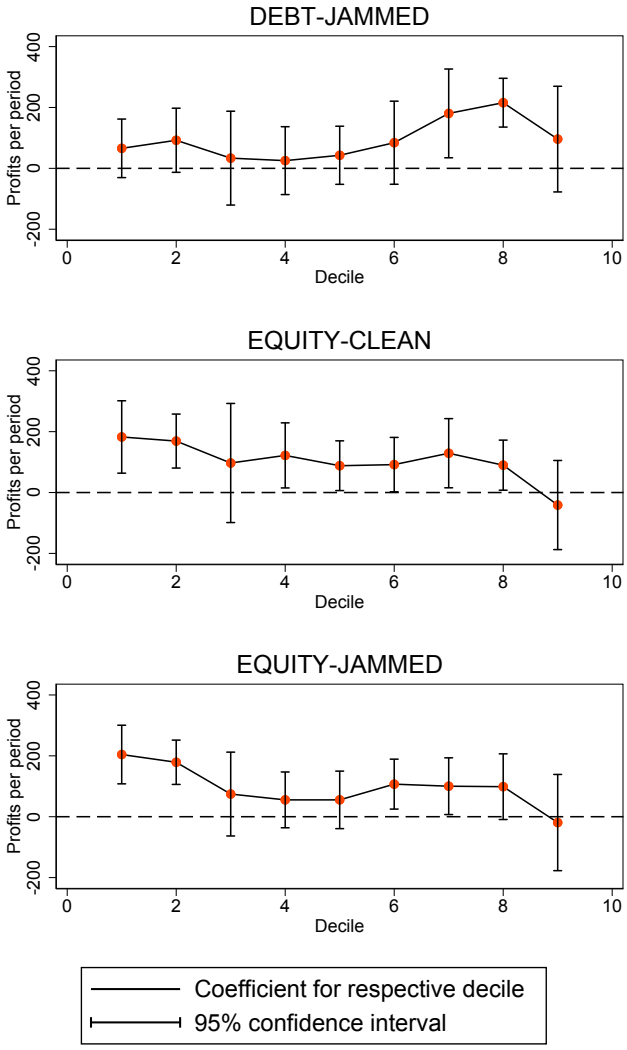
Under the DEBT-JAMMED treatment, average profits are largely flat across deciles and only increase significantly towards the higher deciles. Thus, this treatment seems to mainly support participants with already comparatively high (predicted) profits. Meanwhile, for both microequity treatments the increase in average profits rests on a broad support: Significantly positive effects are registered at most deciles. Notably, some especially strong increases in profits can be observed for participants with low predicted profits. The outcome profiles for microequity contracts both with and without jammed signals is furthermore very similar. The limited change in outcome profile when signal-jamming is introduced under a microequity contract (compare EQUITY-CLEAN to EQUITY-JAMMED), serves as evidence in support of microequity's resilience in the face of ability signal noise. Microcredit contracts, on the other hand, experience a non-trivial change in profits distribution when signals are jammed. Appendix A Table 23 confirms that this heterogeneity was driven by ability: the increase in profits under DEBT-JAMMED is concentrated on the sub-group of participants with low English ability, while treatment effects under both microequity contracts were driven by an increase in profits for participants with high English or high graphic-design ability. Furthermore, matching rates did

[61] The specifications of the regressions estimated are same as in Column (4) in Table 18. Note that quantile regression cannot by applied to discrete or quasi-discrete outcome variables – that is why analogous estimations could not have been conducted neither for the matching rate nor the bank repayment amount.

in fact increase under all treatments – but only for high-ability participants.[62] Despite superficial similarities at the level of average treatment effects, the underlying behavioural shifts appear to be substantially different depending on contract type. Under microequity treatments, there is a shift to matching preferences based on rational expectations, increasing profits broadly but with a bias toward high-ability participants. Under microcredit, a niche of matches driven by a few, not necessarily high-ability participants achieved higher profits – potentially driven by an increase of effort expenditure. A tentative policy implication emerges: while ostensibly both contract types can generate an increase in profits, the DEBT-JAMMED treatment has revealed two unattractive properties. First, a large share of its profits seem to stem from an increase in effort expenditure – which does not solve the fundamental problem of finding business partners in a noisy environment. Second, participants seem to be shifting towards behavioural patterns that do not favour the productivity of high-ability participants. The latter is especially a problem under a regime of multiplicative production functions. Furthermore, the equity contracts' impact on profits is more robust to jammed ability signals. This suggests that microequity contracts could be more suited than microcredit at facilitating welfare-enhancing structural changes.

[62] Appendix A Table 23 presents an overview of regressions that are analogous to Column (4) in Table 18, but where treatment variables are interacted with a subgroup indicator variable. For the first three panels, the subgroups are separated by being above or below median in terms of ability or household income. In the last panel, subgroups are separated by whether participants identified as female. In the lower half of each panel, results for the probability of being in a match instead of profits are shown.

Figure 23: Quantile regression



Note: Treatment effects from quantile regressions for various estimation deciles. The dependent variable is profits (production minus bank repayment, refers to each round's change in incentivised score) – note that quantile regression is not feasible for discrete outcome variables. Control variables include: English and graphic-design ability, round fixed effects, participant age, participant gender, and income of participant. Standard errors are clustered at the level of persons.

4 Structural Estimation

The analysis presented above showed that microequity and microcredit contracts exhibited a range of systematic changes in participants' behaviour when ability signals were jammed. To assess the degree to which these changes are desirable, it is necessary to understand participants' matching behaviour with a microeconomic foundation. The reduced-form analysis that produced the insights above possesses limited ability to provide an explanation of underlying channels of economic decision making beyond analysing correlates. Structural estimation, on the other hand, builds explicit models of these relationships, allowing for the identification of specific economic mechanisms characterised by structural parameters. Therefore, I subsequently move to incorporating a structural estimation to complement the previous section with a brief model of rules guiding matching preferences, followed by a Maximum Simulated Likelihood estimation of key parameters.

4.1 Modelling Framework

I model participant behaviour as follows – for each period in each session and for all treatments:

- At the beginning of each period, any player i draws a jointly Normal 2x1 vector of true abilities for the respective two tasks:

$$x_i = \begin{bmatrix} x_{1,i} \\ x_{2,i} \end{bmatrix} \sim N \left[\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \begin{pmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{pmatrix} \right] \tag{27}$$

Assumption 1 The x_i-vector are normally distributed following Equation 27. In particular, the components of each x_i-vector are independent.[63]

- Based on their x_i, each player is assigned two more 2x1 vectors: a_i represents ability scores (those were the scores generated in the Stage 0 ability elicitation exercise); b_i^* represents the productivity during the subsequent incentivised task in Stage 2. Note that by construction, both of these measures must be $\in [0,1]^2$.

Assumption 2 A monotonic mapping exists between the "deep" ability a_i that Player i wields when solving the ability elicitation exercise and the thereby generated ability score a_i^* (which all other Players -i can observe for Player i when submitting their ranking preferences). An analogous mapping exists between the "deep" productivity b_i and the score achieved answering questions in Stage 2, b_i^*. This mapping takes the form:

$$\begin{aligned} &a_i^* = \Lambda(a_i) := \frac{1}{1+\exp(-a_i)} \;\; and \;\; b_i^* = \Lambda(b_i) := \frac{1}{1+\exp(-b_i)} \\ &\text{with} \;\; \Lambda(x) \in (0,1) \; \forall x \end{aligned} \tag{28}$$

- That is:

$$\begin{aligned} &a_i = \begin{bmatrix} x_{1,i} + \epsilon_{1,i} \\ x_{2,i} + \epsilon_{2,i} \end{bmatrix} \;\text{and}\; b_i = \begin{bmatrix} x_{1,i} + \epsilon'_{1,i} \\ x_{2,i} + \epsilon'_{2,i} \end{bmatrix} \\ &\text{with} \;\; \epsilon_i = \begin{bmatrix} \epsilon_{1,i} \\ \epsilon_{2,i} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix} \right] \; i.i.d \\ &\text{in particular:} \;\; Cov(\epsilon_i, \epsilon'_i) = 0 \end{aligned} \tag{29}$$

[63] The empirical correlation between English and graphic-design ability scores was -0.08 across all data points.

- Recall that the structure of profit functions for players i and j matched together under an equity contract[64] are common knowledge and are as follows:

$$\begin{aligned} \text{Matched} \quad & \pi_{i,j,1,2} = 2b^*_{i,1}b^*_{j,2}(1-s) \\ \text{Unmatched} \quad & \pi_{i,unmatched,1} = 1b^*_{i,1}(1-s) \end{aligned} \tag{30}$$

  where $b^*_{i,1}$ is Player i's score when answering questions related to task 1.

Assumption 3 Participants are risk-neutral – they therefore rank possible matches according to expected profits.

Then each paired observation $a_{k,i}, b_{k,i}$ of corresponding signal and productivity of participant i toward task k is drawn from a jointly Normal distribution:

$$\begin{aligned} &\begin{bmatrix} a_{k,i} \\ b_{k,i} \end{bmatrix} \sim N\left[\begin{pmatrix} \nu_k \\ \nu_k \end{pmatrix}, \begin{pmatrix} \psi_k + \phi_k & \psi_k \\ \psi_k & \psi_k + \phi_k \end{pmatrix} \right] \\ &\text{with} \quad \rho := Corr(a_{k,i}, b_{k,i}) = \frac{\psi_k}{\psi_k + \phi_k} \end{aligned} \tag{31}$$

Assumption 4 Participants know their own ability.

Thus, the only stochastic component to be quantified when calculating expected profits is the productivity of potential partners – for which relevant signals are available. Using the general properties of the bivariate normal distribution, the conditional expectation can be expressed as follows:

[64] The subsequent derivations can be made equally for a debt-based contract. Only the micro-equity case is shown here for illustrative purposes.

$$
\begin{aligned}
E[b_{k,i}|a_{k,i}] &= E[b_{k,i}] - \rho\frac{\sigma_b}{\sigma_a}(a_{k,i} - E[a_{k,i}]) \\
&= \nu_k - \rho((a_{k,i} - \nu_k) \\
&= \nu_k(1-\rho) + \rho a_{k,i}
\end{aligned}
\tag{32}
$$

Thus, the prediction of profits associated with a potential match critically hinges on participants' beliefs about the signal-to-productivity correlation, ρ. Given the complexity and multitude of parameters participants were asked to estimate, it seems reasonable that not every guess got it just right. The degree to which a participant's predictions correspond to rational expectations is captured by the "correctness" of ρ. If a given participant underestimated the true value of ρ, for example, this participant's ranking preferences would not follow ability signals as closely as one may expect. Particularly the DEBT-JAMMED treatment seemed to have a high frequency of such cases. Another value of ρ that merits mentioning due to its empirical importance is the corner case under the belief $\widehat{\phi_k} = 0$. Then the rational prediction becomes identical to the signal itself, since: $E[b_{k,i}|a_{k,i}] = a_{k,i}$. Due to the simplicity of this rule of thumb it is not hard to see why this shortcut might have seemed appealing to participants even if they did not strictly believe in perfect coincidence of signal and performance. This makes the strategy of direct prediction an interesting baseline case when moving to estimating the importance of different rules of thumb. To express analytically the measurement error incurred when utilising this heuristic, let $\widehat{b_{k,i}} = a_{k,i}$ be such a naive approach to predicting performance. Then one may rewrite Equation (32) as:

$$
\begin{aligned}
E[b_{k,i}|a_{k,i}] &= \hat{b}_{k,i} + \nu_k(1-\rho) + \rho a_{k,i} - a_{k,i} \\
&= \widehat{b_{k,i}} + \tau_{k,i}
\end{aligned}
\tag{33}
$$

In other words, the "correct" conditional expectation can be written as the naive prediction plus a normal, mean-zero shock $\tau_{k,i}$. Further, one can state that:

$$\begin{aligned} &Var(\tau_{k,i}) = (1-\rho)^2\psi_k^2 \\ &\text{Let } \ \sigma := (1-\rho)\psi_k = \frac{\psi_k\phi_k}{\psi_k+\phi_k} \end{aligned} \tag{34}$$

This σ parameter is of interest because it contains a structural summary of participants' strategy for translating signals into preferences. It is an easily captured measure of the deviation of participants' beliefs from the baseline case of a one-to-one translation of signals to performance prediction. To illustrate the implications of different σ values consider two cases that can be encountered frequently in the data. First, suppose σ is equal to 0 for participants using the naive performance prediction rule outlined above, with the implied ϕ_k equal to 0 as well.[65] This would then lead to a highly signal-oriented ranking behaviour as described in Section 3.1. Second, suppose that participants engage in a non-rational or non-signal-based ranking process (which was also frequent). One would then expect the data generated to look like rankings implied by a σ that is substantially higher than 0. Recall that experimental literature (Banerjee and Duflo (2007), Drexler et al. (2014)) suggests that rules of thumb play a significant role for complex decisions of microentrepreneurs. While σ might exist on a continuous spectrum, consider the approximation of two broad archetypes of σ values. These could represent two families of rules of thumb. The case of very low sigma values corresponds to taking signals at face values. Meanwhile the high-σ cases could stem from rules that focus on different factors entirely (for example, many participants had strong preferences for working on a specific task) and, therefore, produce preferences that seem almost orthogonal to signals.

[65]Note that estimating the "true" σ using full ex-post knowledge of the data yields results very close to 0. It is therefore a remarkably well-approximated guess to treat signals at face-value.

4.2 Empirical Estimation (Maximum Simulated Likelihood)

In order to parametrise the model outlined above and quantify the significance of some of the types of participants and the parameters implied by their action, I use Maximum Simulated Likelihood estimation. Suppose all other parameters are known, a maximum-likelihood pick for σ will generally satisfy the following condition:

$$\widehat{\sigma}_{ML} = \arg\max_{\sigma} \prod_{i} \prod_{j} \prod_{l} \widehat{P}_{i,j,l}(\sigma, A)\widehat{d}_{i,j,l} \tag{35}$$

where $\widehat{P}_{i,j,l}$ describes the likelihood that a rational agent i employing parameter σ and observing the ability matrix A will prefer matching with agent j over matching with agent l; and $\widehat{d}_{i,j,l}$ describes a dummy assuming value 1 if agent i has actually submitted a preference of j over l, and 0 otherwise. Note that the matching decision in this experiment has three related steps to it: (i) Decide whether to match with any participant at all; (ii) Decide which of the two tasks should be assigned to which participant; (iii) Assign a preference ordering to all participants with which one would like to match. Note that the inclusion of all possible ranking permutations for hundreds of ranking observations over thousands of simulation iterations necessitates infeasibly high computational loads. Thus, in a first effort toward easing computational requirements, I assume that σ impacts any participant i's ranking of each of the nine possible partners in the same way. I shall henceforth focus on the following criterion for σ:

$$\widehat{\sigma}_{ML} = \arg\max_{\sigma} \prod_{i} \prod_{j} \widehat{P}_{i,j}(\sigma, A)\widehat{d}_{i,j} \tag{36}$$

where $\widehat{P}_{i,j}$ describes the likelihood that any participant i will want to match up with participant j.

In a Maximum Simulated Likelihood context, one may further simplify (using the example of an equity-treatment production function):

$$\begin{aligned} &\iff \arg\max_{\sigma} \sum_{i}\sum_{j} log(\widehat{P}_{i,j}(\sigma, A))\widehat{d}_{i,j} \\ &\iff \arg\max_{\sigma} \sum_{i}\sum_{j} \widehat{P}_{i,j}(\sigma, A)\widehat{d}_{i,j} \\ &\iff \arg\max_{\sigma} \sum_{i}\sum_{j}\sum_{k} \mathbb{1}\{\pi^{e}_{i,j} > \pi^{e}_{i,unmatched}\}_{k}\widehat{d}_{i,j} \end{aligned} \tag{37}$$

where $\pi^{e}_{i,j}i$ is the maximiser of expected profits over both tasks given (σ, A). Due to the sheer number of possible combinations with nine possible matches already, a further approximation was introduced to save computational power and only shift the levels of likelihood upward jointly:

$$\iff \arg\max_{\sigma} \sum_{k}\sum_{i} \mathbb{1}\{(\sum_{j} \mathbb{1}\{\pi^{e}_{i,j} > \pi^{e}_{i,unmatched}\}_{k}\widehat{d}_{i,j}) > 8\} \tag{38}$$

That is, to pick the most likely value for σ, I count the number of simulation draws where eight out of nine possible matches were marked as either "want to match" or "do not want to match" in accordance with predicted profits. In reflection of the anecdotal fact that many participants seem to have expressed a preference for one task or the other, I introduce a second case-level parameter κ to account for such preferences. κ assumes values between 0.7 and 1.3 and is multiplied with the expected profit $\pi^{e}_{i,j,t}$ in case that predicted scenario includes participant i working on the English task. Thus, values of κ below 1 capture an aversion to the English task, values above 1 a preference for the English task and a value of exactly 1 implies indifference. Furthermore, since the designation of tasks was a part of the submitted preferences, I maximise over the two tasks in order to compare the best possible

outcomes under both scenarios. Thus, the expression to be maximised looks as follows:

$$\arg\max_{\sigma,\kappa} \sum_k \sum_i \mathbb{1}\Big\{(\sum_j \mathbb{1}\{\max_{t=1,2}\{\pi^e_{i,j,t}\kappa^{\mathbb{1}\{t=1\}}\} > \max_{t=1,2}\{\pi^e_{i,unmatched,t}\kappa^{\mathbb{1}\{t=1\}}\}\}_k \widehat{d}_{i,j}) > 8\Big\} \tag{39}$$

4.3 Estimation Results

In order to recover the structural parameters through Maximum Simulated Likelihood, I iterate over each of the 840 instances of ranking stages ("cases"), of which there is one for each period for each participant in each session. For each of these cases, I redraw 5,000 iterations of i.i.d. normal shocks with mean 0 and standard deviation σ and add these shocks to the matrix[66] of ability signals A visible to the participant in the respective case in each subsequent iteration. Using the shocked ability signals to predict expected profits, I evaluate which element in the (σ, κ)-space[67] maximises the number of simulation interactions where the predicted match/no-match classification is identical to the one submitted by the participant for at least eight out of nine options.[68] Assigning the values for σ and κ respectively which maximise the simulated likelihood for each case, I arrive at the distributions shown in Figure 24. Note that this represents 91.3% of observed cases to which any set of parameters

[66] With individuals signals being transformed using Λ described above, added to the respective shocks, and then transformed back into the interval of real numbers between 0 and 1.

[67] Admissible σ-values were multiples of 0.5 in the interval of [0, 20]; κ-values were either 0.7,1, or 1.3 (preliminary estimations with smaller ranges and step-sizes hinted at a relative preponderance of extreme solutions among the maximum-likelihood picks.).

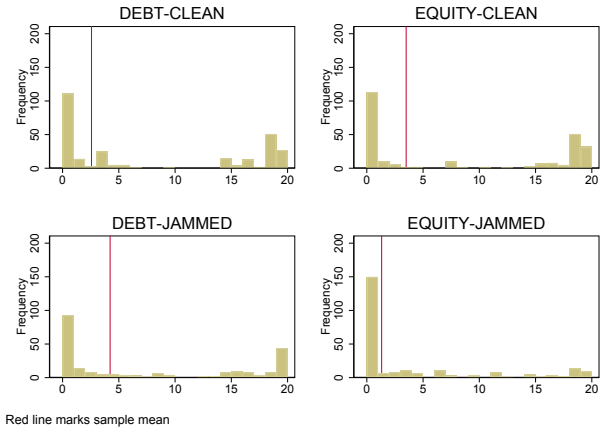
[68] Appendix A Figure 28 displays the resulting relationship between σ and the simulated log-likelihood, split by treatment and κ-value.

could be fitted.[69] There seem to be two principal approaches through which participants have tried to solve the matching problem. A predominant share of participants – those with $\sigma = 0$ – take signals at face value and seem to be operating a rule of thumb that predicts profits by simply plugging in the ability signals into the production function. I shall call participants employing this rule of thumb the "mice" – these participants take ability signals literally and present ranking preferences mostly collinear to the resulting predictions of profits. Another share of participants seem to employ different rules of thumb that contain a more volatile element of decision-making. Based on the fact that few participants express indifference between the two tasks (as suggested by the lower half of Figure 24) and anecdotal inspection of the data, these participants seem to be operating using niche strategies. For example, certain participants would only ever be willing to match with potential partners with exceptional ability but strictly no other participants. I call the participants guided by this family of rules, acting almost orthogonally to ability signals, the "mavericks". There is no implication that high-σ rule are necessarily erroneous behaviour: There could be valid or even rational reasons why participants have chosen to use these approaches. As the results presented above suggest, some low-ability participants were indeed very successful under DEBT-JAMMED – a treatment under which the degree of orthogonal matching increased overall. This archetype of rule of thumb seems not to be in fully line with rational expectations, but it does not represent a departure economic considerations.

[69]This does not include those 71 cases for which no fit was possible. These observations include, among others, cases in which only one preference was submitted by participants – making it computationally extremely challenging to fit to an ability-based prediction. Furthermore, this includes many seemingly random rankings for which even 5'000 simulation rounds with extremely high σ-values did not achieve a positive likelihood. This type of behaviour could potentially be attributed to the high-σ cluster introduced below, even though it is different from observations regularly classified as such.

To clarify the split into archetypes, I perform a k-means classification on the cases using the κ and σ values picked by maximum likelihood. The resulting distribution of σ values is shown in Figure 25 (a). Two sub-distributions of best-fit σ values are unveiled: One centred around the lower bound (value of 0) and one bunched against the upper bound (value of 20). The relative share of these clusters varies by treatment. As is shown in Figure 25 (b), the share of the high-σ cluster number 2 (corresponding to the "mavericks") is high under DEBT-CLEAN and EQUITY-CLEAN with a slight drop under DEBT-JAMMED and a stronger drop under EQUITY-JAMMED. This indicates that jamming ability signals did lead to a significant change in matching behaviour. Rather than through the extensive margin (that is, fewer match requests per participant per session), participants adjusted the rules of thumb according to which they formed matching preferences. The nature of that adjustment was significantly impacted by the contract regime in place in the respective treatment arms. When signals were noisier and participants did have access to implicit insurance through microequity, they nevertheless displayed a high willingness to rank based on ability signals. This was not the case under a jammed microcredit treatment.

Figure 24: Maximum Simulated Likelihood – parameter picks



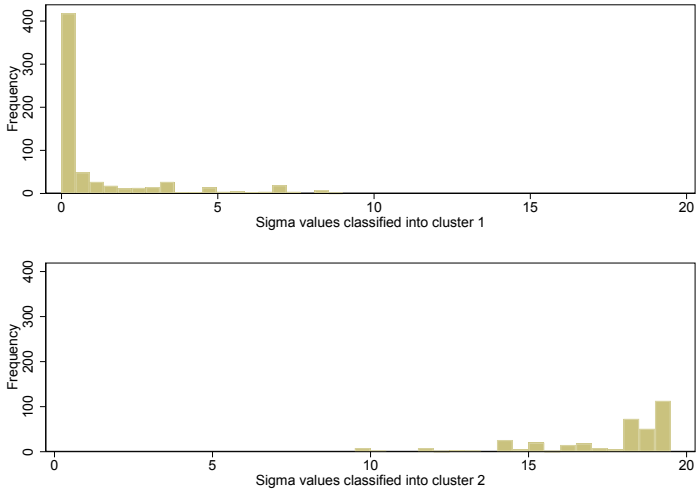
(a) Sigma-parameter



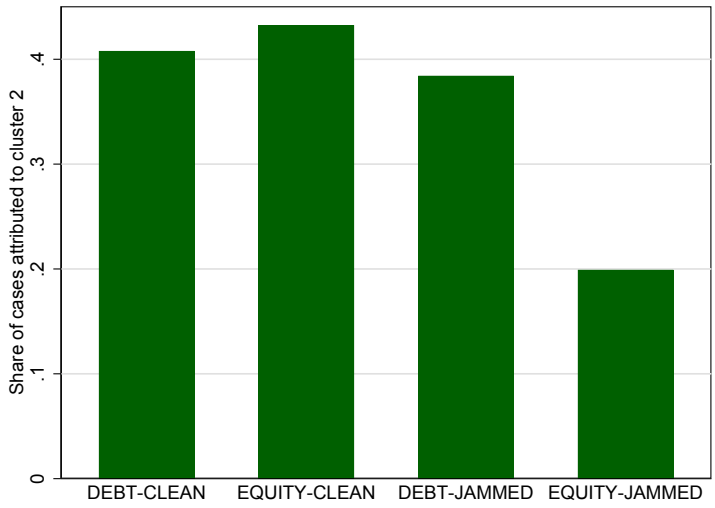
(b) Kappa-parameter

Note: Maximum Simulated Likelihood parameter estimation results by treatment with 5000 independent simulation draws for each case observation. If maximum likelihood was achieved under multiple sets of values, the average of these values is displayed. Values shown for 91.3% of case-observations for which any parameters achieved positive likelihood. σ refers to the standard deviation which additional hypothetical shocks to ability signals ought to have had to justify participants' ranking. κ captures a relative preference (expressed in expected profits) for the English task over the graphic-design task. σ was allowed to assume values of 0.5 increments between 0 and 20, while κ assumed values from the set of $\{0.7, 1, 1.3\}$.

Figure 25: Maximum Simulated Likelihood – case clusters



(a) Sigma distribution by cluster



(b) Share of cluster 2 / "maverick"-cluster by treatment

Note: Results of a k-means clustering (with two clusters) based on σ and κ values of all case-observations for which parameters could be estimated. Sub-figure (a) shows the distribution of σ values for the two resulting clusters. Sub-figure (b) shows the share of cases attributed to the high-σ cluster 2 or "maverick"-cluster (shown in lower half of Sub-figure (a)) by treatment. Values shown for 91.3% of case-observations for which any parameters achieved positive likelihood. σ refers to the standard deviation which additional hypothetical shocks to ability signals ought to have had to justify participants' ranking. κ captures a relative preference (expressed in expected profits) for the English task over the graphic-design task. σ was allowed to assume values of 0.5 increments between 0 and 20, while κ assumed values from the set of {0.7, 1, 1.3}.

5 Conclusion

I have reported the results of an incentivised lab-in-the-field experiment I designed and conducted with a population of graphic-design freelancers in Bangladesh. I explore how microequity contracts perform compared to conventional, debt-based microcredit contracts – particularly when it comes to fostering economic complexity (such as partnerships). I apply this to a setting where production requires complementary skills as inputs. Theoretical considerations such as Kremer (1993) suggest that connecting agents with different skill sets is key for long-term growth. In my relevant experimental model of a service marketplace, I investigate how well suited microequity and microcredit contracts are for facilitating matches based on variously noisy signals.

On average, all examined treatments caused an increase in profits, no treatment increased average matching rates. A part of this pattern is attributable to effort effects for those treatments where matching-ability signals were jammed. However, these average effects conceal an important distinction: As is supported by much of the recent microfinance literature, treatment effects are shaped by considerable heterogeneity. While microcredit increased profits for a narrow segment of participants, not necessarily corresponding to the most highly-skilled participants, microequity's increase in profits took place across a broad spectrum of participants with a focus on high-ability participants. This represents evidence that microequity contracts have desirable properties compared to conventional microcredit. I build on reduced-form analysis with structural modelling of the participant matching behaviour. Through a Maximum Simulated Likelihood estimation, I show that participants' behaviour can be approximated by the use of heterogeneous rules of thumb to cut through the complexity associated with finding partners based on noisy ability signals. Two

archetypal strategies are prevalent in the sample, with participants choosing between them in response to treatment. One rule follows signals closely – the approach of the "mice" – and another family of rules is mostly orthogonal to signals and niche-focused – exercised by the "mavericks". More signal noise has a differential effect on the relative importance of these rules – pushing towards or away from efficient ranking equilibrium. The share attributable to these rule clusters differs by treatment. While the share of "mice" is lowest for microcredit and microequity under clean ability signals, it increases substantially under microequity when signals are jammed. I interpret this as causal evidence of a shift in rules of thumb due to a strategic adaptation to noise that was permissible under microequity but not microcredit.

Combining these individual pieces of evidence into a key insight, a significant advantage of microequity treatments solidifies: robustness to ability signal jamming. Diving from average participant profit profiles into matching quality metrics and underlying structural parameters, the introduction of jamming to signals has caused consistent outcome deterioration under the microcredit treatment – while the microequity treatment was able to absorb the jamming or even improve on the baseline. This is an important consideration both for policymakers who aim to set up markets where collaboration is encouraged (such as online service marketplaces) and for researchers seeking to understand the limitations of classical microfinance instruments. Microequity seems to enable microentrepreneurs to switch to behavioural heuristics that permit them to address matching noise more successfully. These effects – once registered by the microentrepreneurs themselves – would contribute to the self-propelling nature of a marketplace and prevent a possible break-down.

References

Abdulkadiroglu, A., & Sönmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, *66*(3), 689–701. http://www.jstor.org/stable/2998580

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, *101*(2), 470–92.

Ahrens, A., Hansen, C., & Schaffer, M. (2018). *Pdslasso and ivlasso: Programs for post-selection and post-regularization ols or iv estimation and inference.* [Accessed on Dec. 3, 2024]. http://ideas.repec.org/c/boc/bocode/s458459.html

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism*. *The Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Ayayi, A. G. (2012). Micro-credit and micro-equity: The david and the goliath of micro-enterprise financing. *Economic Papers: A journal of applied economics and policy*, *31*(2), 244–254.

Bakhtiar, M., Meki, M., Shonchoy, A., & Quinn, S. (2019). *Skin in the game: Microequity and mentorship for online freelancing based micro-entrepreneurs in bangladesh* [Accessed on Dec. 3, 2024]. https://pedl.cepr.org/node/5445

Banerjee, A., & Duflo, E. (2007). Behavioral development economics: Lessons from field labs in the developing world. In *The handbook of development economics* (pp. 1–54, Vol. 4). Elsevier.

Barboni, G. (2017). Repayment flexibility in microfinance contracts: Theory and experimental evidence on take-up and selection. *Journal of Economic Behavior & Organization*, *142*, 425–450. https://doi.org/10.1016/j.jebo.2017.07.003

Barboni, G. (2024). Innovations in the repayment structure of microcredit contracts. *Oxford Economic Papers*, *40*(1), 129–138. https://doi.org/10.1093/oep/gpab011

Bari, F., Malik, K., Meki, M., Quinn, S., Morduch, J., Ogden, T., & Said, F. (2024). Asset-based microfinance for microenterprises: Evidence from pakistan. *American Economic Review*, *114*(2), 534–574.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, *81*(2), 608–650.

CGAP VII Strategy, F.-F. (2023). *Harnessing inclusive finance: A path toward thriving and sustainable futures* [Accessed on Dec. 3, 2024]. https://www.cgap.org/about/key-documents/cgap-vii-strategy

Chen, D. L., Schonger, M., & Wickens, C. (2016). Otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics*, *3*(1), 55–79.

De Mel, S., McKenzie, D. J., & Woodruff, C. M. (2019). *Micro-equity for microenterprises* (Policy Research Working Paper No. 8799). The World Bank.

Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. G. (2013). Do people care about social context? framing effects in dictator games. *Experimental Economics*, *16*(3), 349–371.

Drexler, A., Fischer, G., & Schoar, A. (2014). Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, *6*(2), 1–31.

El-Gamal, M., El-Komi, M., Karlan, D., & Osman, A. (2014). Bank-insured rosca for microfinance: Experimental evidence in poor egyptian villages. *Journal of Economic Behavior & Organization*, *103*, S56–S73.

Field, E., Pande, R., Papp, J., & Rigol, N. (2013). Does the classic microfinance model discourage entrepreneurship among the poor? experimental evidence from india. *American Economic Review*, *103*(6), 2196–2226.

Fischer, G. (2013). Contract structure, risk-sharing, and investment choice. *Econometrica*, *81*(3), 883–939.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, *69*(1), 9–15. http://www.jstor.org/stable/2312726

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, *112*(2), 631–645.

Hossain, T., & List, J. A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, *58*(12), 2151–2167.

Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Kraay, A., & McKenzie, D. (2014). Do poverty traps exist? assessing the evidence. *Journal of Economic Perspectives*, *28*(3), 127–48. https://doi.org/10.1257/jep.28.3.127

Kremer, M. (1993). The o-ring theory of economic development. *The Quarterly Journal of Economics*, *108*(3), 551–575. http://www.jstor.org/stable/2118400

List, J. A. (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *The Journal of Political Economy*, *114*(1), 1–37.

Mazumdar, R. (2019). *Income sharing contracts for bangladeshi freelancers: Evidence from a lab-in-field experiment*. Oxford University, Department of Economics.

Meki, M. (2024). *Small firm investment under uncertainty: The role of equity finance* (tech. rep.). Oxford Department of International Development (ODID) – Working Paper Series.

Meki, M., & Quinn, S. (2024a). Microequity: Some thoughts for an emerging research agenda. *Oxford Review of Economic Policy*, *40*, 176–188.

Meki, M., & Quinn, S. (2024b). Microfinance: An overview. *Oxford Review of Economic Policy*, *40*(1), 1–7.

Morduch, J. (2013). *How microfinance really works* (Milken Institute Review No. Second Quarter). Milken Institute.

Pretes, M. (2002). Microequity and microfinance. *World development*, *30*(8), 1341–1353.

UNSD. (2015). *The world's women report* (World's Women Series). United Nations, Department for Economic and Social Affairs, Statistics Division.

Appendices

A Appendix – Auxiliary Results

Table 20: Balance of covariates - no Bonferroni correction

				(1)	(2)	(3)	(4)
Group			**All**	**DEBT-CLEAN**	**EQUITY-CLEAN**	**DEBT-JAMMED**	**EQUITY-JAMMED**
English Ability			.44	.5	.36	.43	.47
			[.4]	[.4]	[.4]	[.4]	[.4]
			(.22)	(.22)	(.21)	(.24)	(.21)
Graphic-Design Ability			.66	.71	.62	.69	.62
			[.8]	[.8]	[.6]	[.8]	[.6]
			(.26)	(.25)	(.25)	(.31)	(.23)
Age of Participant			21.68	21.54	22.22	21.2	21.77
			[21]	[21]	[22]	[21]	[21]
			(3.96)	(3.93)	(4.75)	(2.63)	(4.23)
Female			.25	.26	.2	.26	.27
			[0]	[0]	[0]	[0]	[0]
			(.43)	(.44)	(.4)	(.44)	(.45)
Monthly Income (in BDT)			11190.66	12329.79	11607.93	9173.14	11651.79
			[10000]	[10000]	[10000]	[9750]	[9250]
			(10195.38)	(10195.8)	(9038.70)	(5441.72)	(14113.81)
Monthly Income from Freelancing (in BDT)			2397.76	2289.57	3124.33	2458.57	1718.57
			[0]	[0]	[0]	[0]	[0]
			(6467.40)	(6225.19)	(7144.01)	(7401.26)	(4851.78)
Freelancing Jobs Applied			3.55	3.57	4.06	3.19	3.37
			[0]	[0]	[0]	[0]	[0]
			(7.11)	(7.65)	(7.7)	(6.64)	(6.5)
Freelancing Jobs Completed			1.92	2.39	1.72	1.33	2.22
			[0]	[0]	[0]	[0]	[0]
			(4.3)	(5.72)	(3.47)	(2.9)	(4.59)
Risk Aversion			54.22	54.21	57	52.07	53.6
			[50]	[50]	[50]	[50]	[50]
			(29.59)	(30.61)	(28.65)	(28.81)	(30.67)
Observations			280	70	70	70	70
P-value testing	**(1)=(2)**	**(1)=(3)**	**(1)=(4)**	**(2)=(3)**	**(2)=(4)**	**(3)=(4)**	**(1)=(2)=(3)=(4)**
English Ability	0***	.063*	.528	.07*	.001***	.199	.001***
Graphic-Design Ability	.032**	.721	.029**	.123	.944	.123	.062*
Age of Participant	.363	.549	.741	.122	.559	.341	.434
Female	.424	1	.849	.424	.322	.849	.748
Monthly Income (in BDT)	.658	.023**	.745	.055*	.983	.171	.051*
Monthly Income from Freelancing (in BDT)	.462	.884	.545	.589	.174	.485	.584
Freelancing Jobs Applied	.707	.76	.875	.482	.576	.872	.911
Freelancing Jobs Completed	.408	.172	.847	.473	.476	.178	.396
Risk Aversion	.579	.67	.906	.311	.498	.761	.783

Note: Main figure shown for each variable is sample mean; sample median is in square brackets and sample standard error is in parentheses. P-values in the lower panel refer to the confidence in the null-hypothesis of equal distributions between sub-samples. For the joint tests in the lower panel, significance was calculated *without* using a Bonferroni correction, accounting for the simultaneity of 9 or 27 (in the last column) hypotheses being tested, respectively. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

Figure 26: Balance – cumulative distributions of covariates



Note: Empirical cumulative distribution functions (CDF) for key covariates by treatment. A Kolmogorov-Smirnov test rejects the null-hypothesis of equal distributions of English ability at the 1% level for EQUITY-CLEAN vs. DEBT-CLEAN, and at the 5% level for EQUITY-CLEAN versus EQUITY-JAMMED. For graphic-design ability, the null of equal distribution is rejected at the 5% level for DEBT-CLEAN versus EQUITY-JAMMED as well as for DEBT-JAMMED versus EQUITY-JAMMED.

Table 21: Treatment effects - PDS-LASSO estimation

Outcome: Profits per period	(1)
Equity-Clean	91.80***
	(15.12)
Debt-Jammed	95.56***
	(14.68)
Equity-Jammed	96.94***
	(14.41)
N	4662
Control Mean	419.62

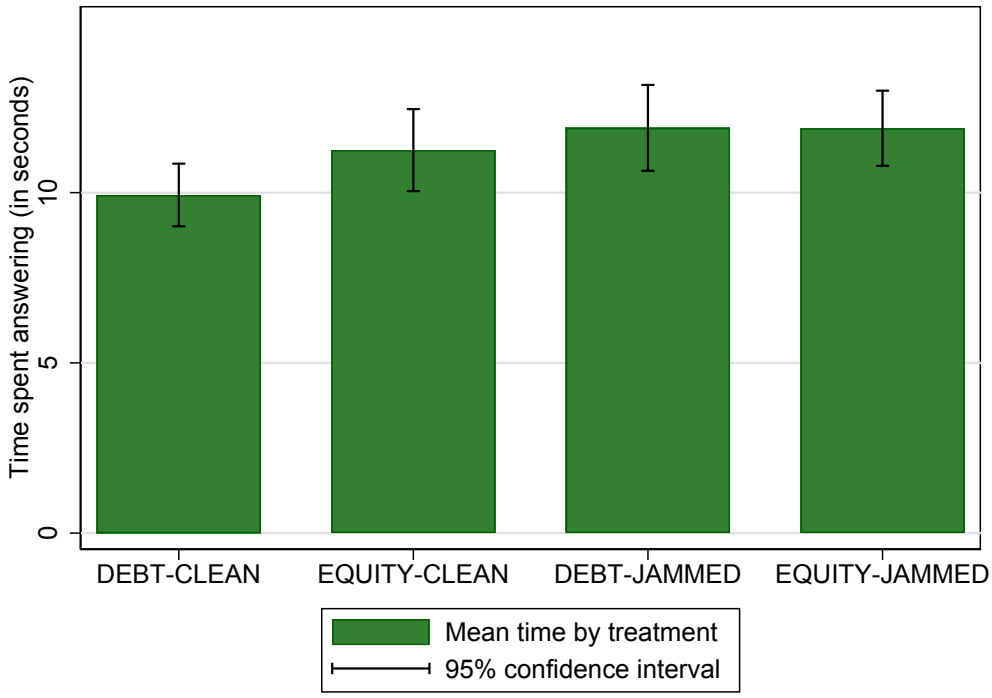
Note: PDS-LASSO estimation of treatment effects on profits per period. See Ahrens et al. (2018) for details on implementation. Covariates included in estimation: English ability, graphic-design ability, age, gender, dummy for third round, risk preference, number of jobs successfully completed by participant. Standard errors in parentheses are only valid for treatment variables.

Table 22: Ranking behaviour – preference-signal consistency

Outcome: Rank-correl. signal-predicted profits and pref. rank	Lin. (1)	Lin. (2)	Lin. (3)	Lin. (4)	P(> Med.) (5)	IHS (6)
Equity-Clean	0.05	0.01	0.08	0.01	0.07	-0.00
	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)	(0.07)
Debt-Jammed	-0.05	-0.05	-0.02	-0.05	-0.12	-0.05
	(0.05)	(0.06)	(0.05)	(0.05)	(0.08)	(0.05)
Equity-Jammed	0.14***	0.14***	0.15***	0.13***	0.16**	0.11**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.04)
Control Mean	.35	.35	.35	.35	.51	.33
Outcome: Ratio of profitable match rejections	Lin. (1)	Lin. (2)	Lin. (3)	Lin. (4)	P(> Med.) (5)	IHS (6)
Equity-Clean	0.36***	0.34***	0.31***	0.28***	0.35***	0.24***
	(0.05)	(0.06)	(0.05)	(0.06)	(0.08)	(0.05)
Debt-Jammed	0.19***	0.18***	0.17***	0.15**	0.17**	0.13**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.05)
Equity-Jammed	0.13**	0.10**	0.13***	0.10**	0.08	0.09**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.07)	(0.04)
Control Mean	.43	.43	.43	.43	.33	.4
Outcome: Ratio of non-profitable match requests	Lin. (1)	Lin. (2)	Lin. (3)	Lin. (4)	P(> Med.) (5)	IHS (6)
Equity-Clean	0.34***	0.34***	0.28***	0.28***	0.22***	0.25***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.05)
Debt-Jammed	0.27***	0.21***	0.24***	0.19***	0.17**	0.17***
	(0.05)	(0.06)	(0.05)	(0.05)	(0.08)	(0.05)
Equity-Jammed	0.09*	0.06	0.09**	0.06	0.05	0.06
	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)	(0.04)
Control Mean	.16	.16	.16	.16	.38	.15
N	782	782	776	776	814	776
FE	No	Yes	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes

Note: First outcome is the rank-correlation between the ranking of possible matches assigned by participants and the ranking of the maximal expected profits achievable with each possible match if they performed exactly as indicated by ability signals. Non-profitable match requests represents share of all match requests where participants declared they wanted to match with another participant, even though they would have been better off on their own. Profitable match rejections represents share of all rejections where participants did not want to match with another participant, and that was in line with profit-maximisation. Lin. stands for linear regression, $P(> Med.)$ for probability of outcome being above resp. sample median. IHS stands for regression with inverse hyperbolic sine of outcome variable. Control mean refers to the DEBT-CLEAN treatment. Control variables include: English and graphic-design ability, round fixed effects, participant age, participant gender, and income of participant. Table displays coefficients from OLS regressions. Standard errors in parentheses are clustered at the level of persons. Fixed effects are on the level of persons. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

Figure 27: Effort – time spent on task



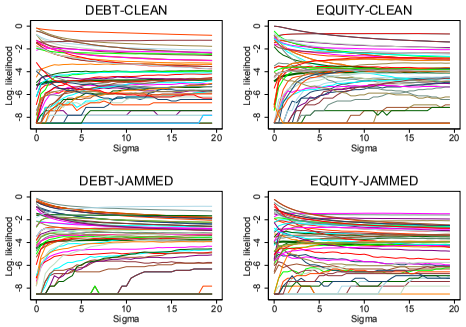
Note: Bars depict average time (in seconds) participants spent on each of the five incentivised questions (during the task completion in Stage 2), split by treatment. Standard errors are clustered at the level of persons.

Table 23: Treatment effects – heterogeneity

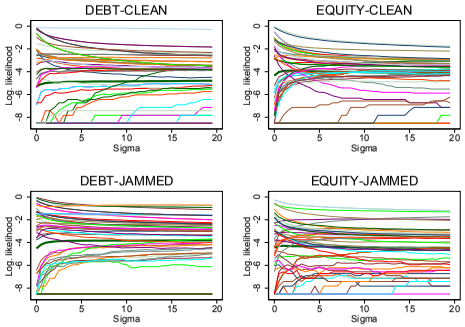
	Low			High		
	EQUITY-CLEAN	DEBT-JAMMED	EQUITY-JAMMED	EQUITY-CLEAN	DEBT-JAMMED	EQUITY-JAMMED
Sep. by English ability						
Profits	87.91	162.77 **	73.71	176.48 ***	91.16	189.89 ***
	(63.1)	(79.36)	(57.22)	(59.99)	(89.35)	(64.59)
Profits	105.13 *	161.35 **	108.59 **	222.76 ***	132.06	207.33 ***
(with Controls)	(58.86)	(79.38)	(53.2)	(58.7)	(83.79)	(65.41)
I{Matched in round}	.07	.07	-.06	.21 ***	.18 **	.18 ***
	(.07)	(.06)	(.09)	(.06)	(.07)	(.06)
I{Matched in round}	.09	.03	-.01	.25 ***	.23 ***	.18 ***
(with Controls)	(.06)	(.06)	(.08)	(.06)	(.07)	(.06)
Sep. by graphic-design ability						
Profits	55.33	57.37	58.8	104.56	170.26 *	241.45 ***
	(53.04)	(68.51)	(47.3)	(94.01)	(88.16)	(89.57)
Profits	63.51	72.07	63.19	128.92	183.07 *	240.41 **
(with Controls)	(54.77)	(68.64)	(48.36)	(97.39)	(92.23)	(95)
I{Matched in round}	.01	-.02	-.05	.23 ***	.19 ***	.2 ***
	(.06)	(.06)	(.07)	(.04)	(.04)	(.05)
I{Matched in round}	.06	0	-.03	.32 ***	.25 ***	.18 ***
(with Controls)	(.05)	(.05)	(.05)	(.05)	(.04)	(.07)
Sep. by income						
Profits	58.08	114.21 *	59.73	15.79	-91.17	74.07
	(61.26)	(62.05)	(47.35)	(55.5)	(81.23)	(55.65)
Profits	83.84	128.32 **	83.58	58.53	-69.68	80.35
(with Controls)	(60.37)	(63.1)	(52.14)	(57.71)	(70.90)	(58.34)
I{Matched in round}	-.01	-.02	-.08	0	.07	.02
	(.07)	(.05)	(.06)	(.07)	(.07)	(.11)
I{Matched in round}	.08	0	-.03	.07	.09	-.01
(with Controls)	(.06)	(.05)	(.05)	(.06)	(.07)	(.1)
Sep. by gender						
Profits	13.31	74.23	49.94	84.71	5.09	41.79
	(43.69)	(61.67)	(41.79)	(63.99)	(95.35)	(55.64)
Profits	13.31	74.23	49.94	84.71	5.09	41.79
(with Controls)	(43.69)	(61.67)	(41.79)	(63.99)	(95.35)	(55.64)
I{Matched in round}	-.03	.02	-.09	-.07	-.17 *	-.07
	(.05)	(.05)	(.07)	(.1)	(.1)	(.06)
I{Matched in round}	-.03	.02	-.09	-.07	-.17 *	-.07
(with Controls)	(.05)	(.05)	(.07)	(.1)	(.1)	(.06)

Note: Table shows treatment assignment interacted with a dummy indicator for whether the separating variable is above the sample median. In the case of separating by gender, "high" corresponds to all female, "low" to all male observations. Profits are defined as production minus bank repayment and refer to each round's change in incentivised score. I{...} represents a dummy variable assuming a value of 1 if a participant was in a match in a given period, and 0 otherwise. Control mean refers to the DEBT-CLEAN treatment. Control variables include: English and graphic-design ability, round fixed effects, participant age, participant gender, and income of participant. Table displays coefficients from OLS regressions. Standard errors in parentheses are clustered at the level of persons. Fixed effects are on the level of persons. Stars show level of confidence: 0.1 *, 0.05 **, 0.01 ***

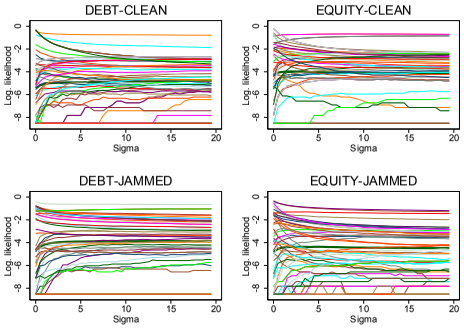
Figure 28: Maximum Simulated Likelihood – case distribution



(a) Preference for graphic-design task



(b) Indifference between tasks



(c) Preference for English task

Note: Progression of individual ranking cases' Maximum Simulated Likelihood over σ values (average of 5000 independent simulation draws shown). Sub-figure (a) imposes a ceteris paribus 30% preference towards the graphic-design task, sub-figure (c) assigns a 30% preference towards the English task. Sub-figure (b) imposes indifference between tasks. This corresponds to the different values of κ. σ refers to the standard deviation which additional hypothetical shocks to ability signals ought to have had to justify participants' ranking. σ was allowed to assume values of 0.5 increments between 0 and 20, while κ assumed values from the set of $\{0.7, 1, 1.3\}$. For display purposes only a random sub-sample of 75% of all fitted observations is shown.

B Appendix – Question Pool

B.1 Graphic-Design Questions

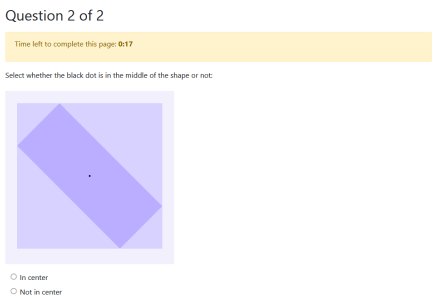
The question pool for graphic-design ability was made up of two sub-types. One sub-type of questions would present geometric shapes and participants would have to judge whether a black dot was centered within the figure or not. 20 of such figures were taken as screenshots from the website https://www.supremo.co.uk/designers-eye/ between the 07.09.2019 and 14.09.2019. The second sub-type would show four different arrangements of ten boxes each which would be painted in different color hues. Participants were then asked to judge which of the orderings is closest to the natural color spectrum of visible electromagnetic waves. A total of 20 orderings were sampled from the wesbite https://www.xrite.com/hue-test between the 07.09.2019 and 14.09.2019.

B.2 English Questions

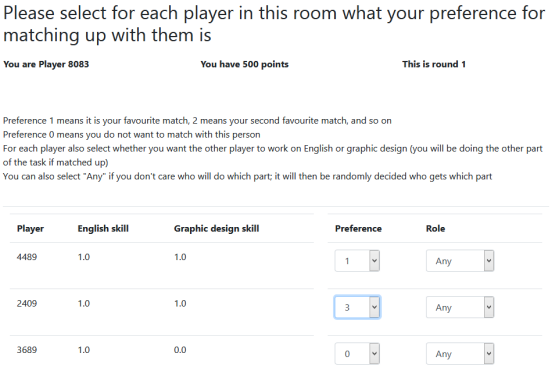
As part of the training courses predating this experiment, participants had previously been sharing an account on the online freelancing platform *upwork.com* to market services to clients across the globe. From these interactions, chat logs have been made available for the purpose of this study and two question types created. The first type consisting of 40 questions would present a real feedback message from a client stemming from the aforementioned chat logs and participants would have to judge whether the client was expressing positive, neutral or negative sentiments. The second type would again present a real chat message with three answer options – consisting of designer-side messages from the same dataset but from different conversations. Participants were then asked to select the most appropriate response. This question type consisted of a pool of 40 examples as well. Some messages were slightly altered to shorten them or correct orthographic errors.

B.3 Screenshots of Main Game Stages

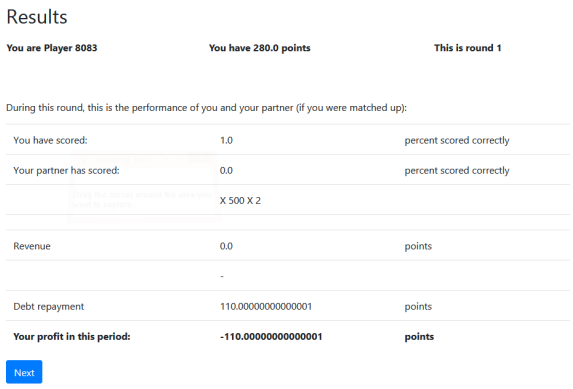
(a) Graphic-design ability question from Stage 0 or Stage 2



(b) Preference elicitation during Stage 1



(c) Presentation of pay-off calculation in Stage 3



C Appendix – Communication to Participants

This experiment and the associated messages presented here were granted ethical approval as part of the regular procedures for review of research involving human participants at the University of Oxford. Specifically, the Research Ethics Committee at the department of economics (DREC) ruled accordingly on 23.08.2019 under file number ECONCIA19-20-14.

C.1 Recruitment message

The following message was translated into Bengali and sent to potential participants by MOMODa staff.

Recruitment message

Dear freelancer,
What makes freelance entrepreneurs successful? And what tools does it take for them to work together and deliver even better services? To investigate questions like this we - a team of researchers from the University of Oxford together with the MOMODA Foundation and Ganna Unnayan Kendra – are launching a study in Gaibandha in the month of September. Since you have proven you have the potential to be a good freelancer during the mentorship program by MOMODA, we wanted to ask you to help us learn about freelancing!
In particular, we are conducting short surveys followed by a computer game where you will have to complete simple tasks alone or partnered up with another freelancer. The time commitment is about 2-3 hours and you will receive a participation fee of 5 GBP as well as a premium depending on your performance in the game.
To secure a spot, get in touch via "EMAIL ADDRESS OF MOMODA CONTACT PERSON" or "WHATSAPP OF MOMODA CONTACT PERSON" to sign up!
We are looking forward to meeting you!
Best,
Nikolai Schaffner

C.2 Elicitation of consent

The following message was translated into Bengali by local staff of MOMODa. At the beginning of each day of experimentation, this message was handed out on a piece of paper to each participant and was furthermore read out loud.

Oral assurance of consent

This survey is being conducted by University of Oxford in collaboration with MOMODa Foundation and Ganna Unnayan Kendra (GUK) to learn about the preferences of potential online freelancers like yourself. Your responses in this survey will not have any bearing on your current or future relationship with GUK and Creative IT. I am going to ask you some questions about different types of lotteries and money contracts. Our conversation here is private. Your responses to the questions will be used for educational research purposes only. The researchers will not reveal your name, or any other identifying information, to anyone. Participation in the survey is voluntary and can be withdrawn at any point. If you do not know the answer, or do not want to answer, for any question, just let me know, and I will go to the next question. If you wish to stop the survey, you can do so at any time. The survey will take approximately three hours. At the end of this session, you will receive BDT 300 plus a premium depending on your performance in the game. During the game, you will have to perform short tasks and you can partner up with other freelancers in this room. The better you perform on these tasks, the more hypothetical money you will accumulate. At the end, the hypothetical money will be converted into a cash premium. However, you must complete the whole survey and comply with the rules of the game in order to get your payment. The enumerators will record your responses in this electronic tablet. Your answers will be completely anonymous, and your data will be stored in a password-protected file and may be used in academic publications. Your private information will not be used. The University of Oxford and MOMODa Foundation will have access to the data, as well as the survey company. The supervisor for this project is Dr. Simon Quinn. Dr. Muhammad Meki from the University of Oxford is acting as a co-supervisor. The project has been designed and implemented here under the supervision of Nikolai Schaffner, a student at the University of Oxford. Do you agree to participate in the survey?

Concluding Remarks

In this thesis, I have compiled three standalone chapters – making contributions to three distinct but related areas of the applied microeconomics literature by creating and analysing novel datasets. They are unified by their common empirical roots in a developing country context, their aim to support policy makers with data-driven insight, and their ambition to incorporate machine-learning methodology to microeconomic questions where doing so adds value.

While all chapters through their contributions – suggest pathways for further research follow-up, I would like to particularly emphasise the profound potential that Gaussian Processes have unveiled in the second chapter. At the heart of many empirical microeconomic models lie methodological short-cuts which trade mathematical tractability for external validity. The potential of Gaussian Processes for economic research that the second chapter illustrated could greatly reduce the need for such short-cuts – or further validate their usage.

The insights I have demonstrated in this thesis do not only provide further research impulses – they also constitute helpful tools in shaping policy, particularly with regards to enabling economic development.