

NBER WORKING PAPER SERIES

MULTIPLE PRICE LISTS FOR WILLINGNESS TO PAY ELICITATION

Kelsey Jack
Kathryn McDermott
Anja Sautmann

Working Paper 30433
<http://www.nber.org/papers/w30433>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2022

We thank Kenneth Chan and Lei Yue for excellent research assistance, and the City of Cape Town for collaboration. Field work was supported by J-PAL Africa, J-PAL's Urban Services Initiative and the International Growth Center. This work was undertaken in 2018. At that time Kathryn McDermott was employed at J-PAL Africa. The research was undertaken in partnership with the City of Cape Town. Since 2021, Kathryn has joined the City of Cape Town. The research and findings as presented have no relation to her current employment arrangement, nor is it the official position of The City of Cape Town in any way. Software development was supported by a World Bank RSB grant. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, nor those of the Executive Directors of the World Bank, nor the governments they represent, nor the National Bureau of Economic Research. Declarations of interest: none. All errors are our own.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by Kelsey Jack, Kathryn McDermott, and Anja Sautmann. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Multiple Price Lists for Willingness to Pay Elicitation
Kelsey Jack, Kathryn McDermott, and Anja Sautmann
NBER Working Paper No. 30433
September 2022
JEL No. C91,C93,D46,O12,Q51

ABSTRACT

Multiple price lists are a convenient tool to elicit willingness to pay (WTP) in surveys and experiments, but choice patterns such as “multiple switching” and “never switching” indicate high error rates. Existing measurement approaches often do not provide accurate standard errors and cannot correct for bias due to framing and order effects. We propose to combine a randomization approach with a random-effects latent utility model to detect bias and account for error. Data from a choice experiment in South Africa shows that significant order effects exist which, if uncorrected, would lead to distorted conclusions about subjects’ preferences. We provide templates to create a multiple price list survey instrument in SurveyCTO and analyze the resulting data using our proposed methods.

Kelsey Jack
University of California, Santa Barbara
2400 Bren Hall
Santa Barbara, CA 93106-5131
and NBER
kelseyjack@ucsb.edu

Anja Sautmann
The World Bank
1818 H Street NW
400 Main Street
Washington, DC 20433
asautmann@worldbank.org

Kathryn McDermott
City of Cape Town
Civic Centre, 12 Hertzog Boulevard
Cape Town, NA NA
South Africa
kathryn.mcdermott@capetown.gov.za

A data appendix is available at
<http://www.nber.org/data-appendix/w30433>
A technical appendix is available at
<https://ucsb.app.box.com/v/Jack-MPL-Appendix>
A package repository is available at
<https://github.com/MPL-WTP/MPL-WTP>

1 Introduction

Multiple price lists (MPL) are a widely used and convenient tool to elicit subject preferences as part of a survey or experiment. MPL elicitation combines ease of implementation with detailed information about preferences, which makes the method attractive for field settings. Researchers have used it to elicit willingness to pay (WTP) for a diverse set of goods and services, from timed rewards to cookstoves.¹ Willingness to pay (WTP) here refers to the subject’s (relative) preferences between two options expressed in a common unit, such as money. In an MPL, preferences may be elicited by asking repeated questions that allow the subject to “buy” or “sell” indivisible goods at various prices, or to choose between divisible goods at different quantities, thus establishing the subject’s (relative) valuation of these goods. The use of MPL in WTP elicitation is not new; for example, Kahneman et al. (1990) employ it in their seminal laboratory study on the endowment effect, involving subjects who buy and sell everyday items such as pens or mugs. Since then, the literature has grown to include both methodological studies (e.g. Andersen et al., 2007) and applied measurement (e.g. Bursztyn et al., 2018).²

Researchers in applied microeconomics and development economics typically measure WTP in the context of a larger project and therefore look to the existing literature for the best measurement method. Perhaps for this reason, there has been renewed interest in validating WTP elicitation methods. Several recent papers have focused on the broader question of comparing methods, such as take-it-or-leave-it (TIOLI) vs. the Becker-deGroot-Marschak (BDM) random price mechanism (Berry et al., 2020; Cole et al., 2020). MPLs are situated somewhere between these methodologies: they provide an incentive-compatible elicitation method that delivers more information than TIOLI, but is easier to understand and more widely applicable than BDM.³

Among studies that use MPL to elicit WTP, data collection instruments and estimation methods vary considerably. Much of this variation is driven by a set of common

¹See Appendix B for an illustrative list of papers and Burchardi et al. (2021) for an overview.

²MPL are also commonly used for the elicitation of other preference parameters, such as time preferences and risk aversion (e.g. Andersen et al. (2008)), and much of the methodological research on MPL has been done in that literature. A key difference is that risk aversion and time preference elicitation are primarily used to establish ordinal preferences (through the parameters in a utility function), whereas in WTP elicitation, the cardinal valuation is often a key research interest.

³Many authors treat MPL between a single item and money as a special case of BDM (e.g., Cole and Fernando (2020); Maffioli et al. (2020)). We think of MPL more broadly as a format for eliciting preferences between any two different options, which includes that one or both of these options are money. The key feature of BDM is that the final price is selected randomly, which makes the mechanism incentive compatible with stating the true WTP, and MPL share this feature. Thus, MPL provide a modified format to implement BDM.

challenges that the existing approaches have only partially addressed. First, many subjects exhibit “multiple switching behavior” (MSB), meaning they make inconsistent choices that are difficult to interpret, by switching between “buying” vs. “not buying” multiple times as prices change monotonically (Yu et al., 2021). Based on these decisions, their WTP may lie in a very wide interval or not be recoverable at all.⁴ Second, many subjects exhibit “never-switching behavior” (NSB), meaning they choose the same option throughout the MPL, which may imply a very wide interval for the WTP including potentially $+\infty$ or $-\infty$. Sometimes, one of these choices is inconsistent because the selected option is strictly dominated.⁵ Last but not least, even consistent responses in MPLs only provide interval identification of WTP.

We begin by introducing a framework aimed at understanding choice error in WTP elicitation, and its potential to add noise to WTP estimates. We argue that MPL can be an effective tool for measuring WTP because of the repeated choice data from the same subject that permits learning about error. However, due to their format of repeated parallel questions, MPL may also be vulnerable to framing and order effects. These create bias because they introduce a correlation between the error and the (change in) value difference between the two options, which is used in an MPL to identify WTP. We argue that many widely used MPL designs do not accurately measure error and are unable to prevent or even detect order effects. Worse, some designs may introduce additional bias. Moreover, most estimation approaches do not model choice error, with implications for the estimated variation and standard errors in WTP.

We therefore propose two measures to improve how MPL-based WTP elicitation incorporates choice error and bias. The first measure is to randomize how subjects view choices in a way designed to address potential framing effects. We propose to vary both the order of binary choices within MPL, so that the value difference between options – aka the “price” – is either ascending or descending,⁶ and the order of options within each binary choice, that is, whether an option appears first or second/on the left or right.⁷ This

⁴Take the case of a “reverse switch” where the subject declines to buy at all low prices offered, but then switches and agrees at a series of higher prices. Formally, they express that their WTP is below x but also above $x + t$, with $t \gg 0$.

⁵Dominated choices are often included in the MPL as a test of rationality.

⁶As we discuss in more detail in section 3.1, we do not propose full randomization within MPL to reduce subjects’ cognitive load; see also Andersen et al. (2007) for the same argument.

⁷We think here in particular of visual presentations of the MPL, e.g. on paper or on a screen, but the order in which the two options are listed may also matter in orally administered surveys. When subjects make yes/no decisions about a good at different prices, the researcher may vary order within choice by varying whether the subject is asked to buy or sell a good, or whether the “yes” or “no” answer option is presented first.

minimizes net bias and allows us to distinguish anchoring effects that arise from simply “repeating” the same decision from explicit choices that reveal true WTP. The second measure is to model each binary choice in the MPL with a latent utility model and apply an appropriately scaled random-effects probit estimation to estimate WTP. This approach accommodates inconsistency (MSB) as well as dominated choices, and thus avoids selection biases that arise when dropping or recoding MSB or NSB choices. In addition, it provides identification of average WTP, dispersion of individual WTP, and errors in subjects’ choices, and allows us to explicitly estimate biases that arise from order effects.⁸ This ability to learn about both the distribution and mean of individual errors is a key advantage of the MPL method relative to alternative preference elicitation methods.

We show that both non-systematic errors and systematic biases play a significant role in individual choices, using a WTP elicitation example from South Africa. This finding is in line with the literature; in a survey of MPL studies, 17.1 percent of observations included inconsistencies (Filippin and Crosetto, 2016; Crosetto and Filippin, 2016). Research in populations with low literacy and numeracy may be particularly affected: for example, Dave et al. (2010) show that subjects with low math ability make significantly more inconsistent choices. Moreover, while most researchers do not randomize the order in which MPL questions are asked, others have found significant order effects (e.g. Channa et al., 2021). In our application, the order of choices within MPL is an important source of bias in measured WTP, but the order of options within choice has little effect on reported valuations.

We compare our WTP estimates with those we would have obtained using MPL design and estimation approaches found in the literature, and show important differences. First, WTP estimates obtained from non-randomized MPL cannot differentiate true WTP from existing order bias, and some MPL even create new sources of bias (such as MPL elicitation that stops after the first “switch”). Second, not accounting for choice error alters the estimated outcome variance. Together, these issues lead to markedly different conclusions about the mean and dispersion of WTP in the study population.

To facilitate use of our proposed MPL design and estimation approach, we provide a template WTP elicitation module implemented in the survey software SurveyCTO that carries out order randomization within the MPL. In addition, we provide template analysis files in Stata that carry out the random-effects latent utility estimation. The template

⁸Note that this methodology can be applied to measures of the sample average WTP, or to subgroup-level measures of WTP, but not WTP at the individual level. Based on the high rate of choice errors in typical MPL data from LMIC contexts, we caution against approaches that calculate an individual-specific WTP from MPL data, for use, for example, as an explanatory variable.

accommodates order randomization as well as a set of other features identified as good practice for MPL design from the authors’ own research and the literature. These features include practice MPL rounds, constant maximal values in each binary choice within MPL, focusing WTP intervals on points of interest for hypotheses tests, and others. Many of these can also be implemented in pen and paper or verbally elicited MPL.

Our paper contributes to the methodological literature on WTP elicitation in field applications (e.g., Berry et al. (2020); Burchardi et al. (2021)). We focus on multiple price lists, an increasingly common elicitation approach in development economics. MPL are popular because they hold the promise of both straightforward implementation for researchers and ease of comprehension for subjects. However, we show that choice error is an important feature of MPL data, and likely any WTP data. When combined with specific features of data collection or analysis, this error has the potential to introduce bias in measured WTP. While others have discussed framing effects in MPL data (e.g., Andersen et al. (2007)), order effects have been largely overlooked to date. We draw three main conclusions: first, in populations where choice error is frequent, we caution against the use of some of the most common methods used in the design and analysis of MPL. Second, researchers using MPL or interpreting MPL data should pay special attention to the potential for order bias. Third, with the right MPL design and estimation approach, we believe that MPL continue to be a valuable tool. Our proposed two-pronged method of order randomization and latent utility estimation provides researchers and practitioners with a unified approach to diagnosing and mitigating bias while correctly accounting for choice error. Our Stata and SurveyCTO templates facilitate the measurement of WTP using MPL.

The next section describes WTP elicitation via MPL. It discusses common patterns in MPL data, presents a framework to think about error in such data, and summarizes how the existing literature has addressed MSB, NSB and interval identification. Section 3 proposes to combine order randomization with a random utility model for MPL choices and uses data from an experiment in Cape Town, South Africa to show how this approach affects measured WTP. We demonstrate in particular that order biases can significantly affect elicited WTP. We also describe how researchers might apply our framework to investigate other sources of bias that arise through MPL design decisions. In section 4, we discuss features of MPLs for WTP elicitation that we identify as useful for improving data quality. These features are implemented in an accompanying template for implementing an MPL in SurveyCTO and estimating WTP from the data, described in detail in a technical appendix (Appendix S2).

2 Multiple price lists for willingness to pay elicitation

Many experimental preference elicitation procedures use some form of list experiment. We focus here on MPL (Multiple Price Lists) that elicit the subject’s relative preference for an option A over a second option B expressed in a common (monetary) value, that is, the subject’s willingness to pay (WTP) to receive option A over option B. However, MPLs can be and frequently are used for other purposes, such as measuring risk or time preferences, and much of our analysis and discussion applies.

In a WTP MPL, the subject makes repeated binary choices between two options, where the monetary value of these options is varied systematically. To fix ideas, we reproduced four examples of MPL experiments from the literature that fit this description in Figure 1. These examples show a variety of designs, and illustrate some key properties of MPL. Our analysis below draws on many other papers as well, and Appendix B provides an overview of 23 papers that use MPL to measure WTP, with a focus on incentivized MPL from low and middle income countries (LMIC). The wide range of contexts and topics covered demonstrate the versatility of the method; the large number of papers since 2019 show that MPL is increasingly used in applied research to measure WTP. We caution that this list is not exhaustive and other papers have influenced our work, including the large literature on risk aversion and time preference elicitation. However, the papers in Appendix Table B showcase a diverse set of high-quality papers directly related to WTP measurement, providing concrete examples to researchers who are designing an MPL experiment.

We use the term “option” to refer to the components that remain the same within the MPL. An option can be a specific good, a task, or a lottery; but also the format or time of delivery of the same commodity, including money, for example when eliciting time preferences. In each row of the MPL, the respondent makes a choice between two options, with varying monetary values associated with each option. The inference on the subject’s willingness to pay comes from the value difference between the two options in each row of the MPL. For example, in the MPL from Allcott and Kessler (2019) in the top left of Figure 1, one option is to receive four home energy reports (for residential electricity users), and the other option is not to receive reports. The associated cash prize varies between \$1 and \$10 for each option; from high to low when choosing the report (left) but low to high when choosing no report (right). As a result, the difference in monetary value between the two options varies between -9 and 9 US dollars.

Figure 1: Examples of MPL-style WTP elicitation in published papers.

Which would you prefer?				
1	\$10	4 more Home Energy Reports PLUS a \$10 check	\$1	A \$1 check
2	\$10	4 more Home Energy Reports PLUS a \$10 check	\$5	A \$5 check
3	\$10	4 more Home Energy Reports PLUS a \$10 check	\$9	A \$9 check
4	\$10	4 more Home Energy Reports PLUS a \$10 check	\$10	A \$10 check
5	\$9	4 more Home Energy Reports PLUS a \$9 check	\$10	A \$10 check
6	\$5	4 more Home Energy Reports PLUS a \$5 check	\$10	A \$10 check
7	\$1	4 more Home Energy Reports PLUS a \$1 check	\$10	A \$10 check

Tick if you are willing to buy at the price in each row and cross “x” if you are not willing to buy.				
		Organic	Conventional	
1	50			
2	100			
3	150			
...	...			
17	850			
18	900			
19	1000			

<i>SurveyCTO questionnaire</i>			
1	Would you buy the card for 0 Shillings (for free)?	yes	no
2	Would you buy the card for 200 Shillings?	yes	no
...
11	Would you buy the card for 2000 Shillings?	yes	no

Notes: Clockwise from top left: Allcott and Kessler (2019) sent out paper questionnaires to home energy users to measure preferences for receiving usage reports. In Squires (2021), a surveyor read out choices between receiving payments of different value either in public or in private, and varied the order of choices between subjects. In Burchardi et al. (2021), a surveyor read out yes/no questions about purchasing a voucher directly redeemable for money at increasing prices. Alphonse and Alfnes (2017) measured preferences for organic and conventional produce for tomato buyers at a market in Tanzania (prices in Tanzanian Shillings).

MPL can be used to carry out a variant of the Becker-deGroot-Marschak (BDM) mechanism, and many authors use the term BDM to refer to an MPL for WTP elicitation. In a “pure” BDM, subjects are asked to state their true value for an object, a random (purchase) price is drawn, and the purchase is realized if the stated value is higher than the price (the reverse for a “sale”). In the equivalent MPL, subjects are explicitly asked whether they are willing to buy (or sell) an object at a series of different prices, and a single row is drawn for implementation. For example, in the price experiment from Burchardi et al. (2021) in the bottom right of Figure 1, subjects are asked whether they would like to buy a voucher (“card”) at prices starting at 0 and going up to 2000 Shillings.

In most MPL, the binary choices subjects make are ordered by either ascending or descending value difference. The order of binary choices, and, where applicable, the order of the options presented in each binary choice (e.g., on the right or left side of the list), do not vary within or between subjects.⁹ On paper or on a screen, subjects may see all choices in front of them at the same time. When administered via a computerized questionnaire or oral instruction by the enumerator, choices for each row may be elicited before revealing the next row. At the end of the experiment, one binary choice is usually chosen for implementation, and the subject’s decision in that binary choice is realized.

In many research projects, MPLs are one element of a larger data collection effort. Nonetheless, since the value subjects attach to different options is often a key factor of interest, researchers put a lot of thought into the design of these MPLs. For example, the original MPL in Allcott and Kessler (2019), reproduced in Figure 1, was a paper questionnaire that used a variety of visual aids to help subjects make their decisions. Many authors report extensive piloting or test different MPL formats, and survey questionnaires often check subject comprehension, allow revisions, or provide “practice” MPLs and other tools to improve measurement. The detailed supplemental material in Berkouwer and Dean (2021) offers a nice example, and we encourage the reader to explore the papers listed in Appendix B.

⁹The MPL implemented in Squires (2021), shown on the top right in Figure 1, is one exception. In his study of WTP for private (vs. public) income, he randomized the order of binary choices: some subjects saw the choice including 500 Sh first, and some saw the choice including 150 Sh first.

2.1 Multiple switching and never-switching: Common issues with MPL data

Despite their widespread use, there are a number of issues with MPLs that many, if not most, researchers have encountered. One common issue is inconsistent choices, often termed multiple switching behavior (MSB). Take a (hypothetical) subject who chooses “no report” in row 2, but “report” in row 3 of the Allcott and Kessler (2019) MPL. This subject has the preferences $(\$5, \text{no report}) \succ (\$10, \text{report}) \succ (\$9, \text{no report})$ and, by transitivity, could be inferred to prefer \$5 to \$9. The term “multiple switching” refers to the fact that consistent choices require that the subject switches between the two options at most once, from the MPL option with decreasing (relative) value to the option with increasing value.

In addition to MSB, many subjects exhibit never-switching behavior (NSB). NSB occurs when a subject chooses the same option through the entire price list. Since the researchers typically vary the value difference between the two options across the full support of a reasonable distribution, this behavior often appears implausible. For example, take a subject in Alphonse and Alfnes (2017) (see Figure 1) who chooses to buy conventional tomatoes at all prices up to TZS 1000 – this would be implausible given that the market price at the time of the experiment ranged between TZS 300 and 400. In the MPL from Burchardi et al. (2021) (also shown in Figure 1), NSB would be an even clearer indicator of irrational choices, since valuations are elicited for a “card” that is directly redeemable for its face value of UGX 1400 in cash, and all subjects should be willing to buy the card for a price below 1400 but not for a price above; in fact, the authors use this as a rationality test when comparing different elicitation methods. In general, however, a problem for the measurement of subjective preferences is that the researcher cannot distinguish between NSB that is the expression of a very strong preference for one of the options, and NSB that is the result of an error.

A survey of 54 published risk elicitation studies employing MPL designs found that 17.1 percent of observations included inconsistencies in the form of MSB or subjects picking dominated options (Filippin and Crosetto, 2016; Crosetto and Filippin, 2016).¹⁰ In WTP elicitation where the trade is realized later, subjects often ex post decline the agreed-upon price: Maffioli et al. (2020) report a high rate of renegeing on MPL outcomes,

¹⁰A similar systematic review of the MPL for WTP literature is not available. Of the papers we surveyed, a subset of which are shown in Appendix B, MSB was rarely reported and was often suppressed by instrument design, as we discuss in greater detail in Section 2.3. NSB was more commonly reported, and ranged from 1-2 percent to 74 percent of subjects.

and summarize a number of other studies that report renegeing rates above 10 percent. This suggests that individual decisions in the MPL may not always reflect the subject's relative valuation of the two options, and these errors cannot always be detected and corrected. In the best case, this adds noise around the true willingness to pay. In the worst case, it introduces systematic bias.

2.2 Measurement error in MPL: A framework

We start with a simple framework to think about confounders and measurement error in a WTP MPL. We assume that utility is additive in monetary value. Thus, option A provides subject i with utility $U_{iA} + v_{At}$ and option B provides $U_{iB} + v_{Bt}$, where t denotes the binary choice the subject is facing, U_{ij} is the utility from option j , and v_{jt} is the monetary value associated with j in t . We assume that the researcher is interested in the mean (or other moments) of the distribution of true underlying differences in utility between options, expressed in money. We call an individual difference the subject's willingness to pay for option A: $WTP_i = U_{iA} - U_{iB}$.

A typical MPL offers T binary choices between the same two options with different associated values v_{At} and v_{Bt} , $t = 1, \dots, T$. If subjects never make mistakes and there are no outside factors that influence choices, these binary decisions provide interval identification for the WTP, because subject i will choose option A if

$$\begin{aligned} U_{iA} + v_{At} &> U_{iB} + v_{Bt} \\ \text{or } WTP_i &> v_{Bt} - v_{At}, \end{aligned} \tag{2.1}$$

option B if

$$WTP_i < v_{Bt} - v_{At},$$

and either A or B if she is indifferent. This means the WTP must lie in the closed interval between the largest value of $v_{Bt} - v_{At}$ where option A is chosen and the smallest $v_{Bt} - v_{At}$ where option B is chosen.

A first observation is that even without any errors, WTP is never point identified, and a distribution over the identified interval must be assumed. Moreover, in the NSB case, the WTP is either in the highest or lowest possible range, but the interval is unbounded: if option A is chosen for all t , we have that $WTP \in [\max_t(v_{Bt} - v_{At}), \infty)$ and if option B is always chosen, we have $WTP \in (-\infty, \min_t(v_{Bt} - v_{At})]$. Decisions about where in these intervals the WTP of never-switchers is likely to lie can have significant effects on

the distribution of elicited WTP.

Of course, some of the patterns discussed earlier indicate that there likely *is* error in individual choices; such as MSB, dominated choices (e.g., choosing not to buy a voucher that can be exchanged for cash), and implausible WTP distributions, such as a mass of never-switchers at both extremes of the MPL. Formally, the subject will weakly prefer option A if

$$\begin{aligned}
 U_{iA} + v_{At} + \epsilon_{iAmt} &\geq U_{iB} + v_{Bt} + \epsilon_{iBmt}, \\
 \text{or } WTP_i &\geq v_{Bt} - v_{At} + \epsilon_{iBmt} - \epsilon_{iAmt} \\
 &\equiv -\Delta_t + \epsilon_{iBmt} - \epsilon_{iAmt}.
 \end{aligned}$$

The terms ϵ_{ijmt} describe any disturbance to the preference for option j in choice t of MPL m . Because of this term, the subject may sometimes choose option A even though $WTP_i < v_{Bt} - v_{At}$ and option B even though $WTP_i > v_{Bt} - v_{At}$. This could lead to multiple switching as well as randomly switching “too early” or “too late”. Going forward, we will use the value difference term $\Delta_t = v_{At} - v_{Bt}$. This is the inverse of the difference in values on the right hand side above, and so subjects will tend to choose option A when Δ_t is large and option B when it is small (incl. negative).

Potential reasons for individual error, ϵ_{ijmt} , include inattention and incomprehension. For example, Yu et al. (2021) have shown in the case of risk preferences that greater attention reduces MSB in final MPL choices. Cason and Plott (2014) also provide evidence that unexpected behavior in a BDM occurs when subjects do not understand the payoff structure of the game created by the experiment, leading to behavior that looks irrational. Incomprehension may be a particular concern in populations with low literacy and numeracy. For example, Dave et al. (2010) have shown that subjects with a math ability score more than one standard deviation below the sample mean make inconsistent choices about four times more often than do other participants. On the other hand, Allcott and Kessler (2019) report only 2.4% inconsistent choices in a mail-in survey with US energy customers.

Note that Andersen et al. (2008) have argued that choices that look inconsistent simply indicate that the subject is indifferent between a range of options. Complete indifference should be a knife-edge case if the MPL is well constructed and choices are realized, but the subject may still care little about the offered payoffs compared to, for example, the time it takes to go through them carefully, and therefore choose not to incur the attentional

cost.¹¹ Whatever the exact causes for error may be – inattention, a lack of comprehension, or near-indifference – in the extreme, these confounders could lead to choices that are only weakly related to true WTP, if subjects treat the options in each binary choice essentially as “the same”. MSB with a large number of switches may in fact be a sign that some subjects randomize deliberately in choices that feel repetitive.¹²

The presence of error has consequences for the interpretation of choice data in elicitation experiments. On the one hand, even when detectable errors occur, the subject’s choices still hold information, because they are correlated with true preferences, but this information is more difficult to extract. On the other hand, a large number of choice errors may remain undetected when the subject’s choices are internally consistent, and any given MPL may not identify a subject’s WTP. Most alternative elicitation methods, such as TIOLI or BDM, do even worse in this respect: they cannot identify inconsistencies or likelihood of error at all because the subject makes only one choice. We think of the ability to learn about the distribution of the ϵ_{jimt} from MSB as one key advantage of the MPL method.¹³

MPL design and order biases in WTP. If the difference $\epsilon_{iBmt} - \epsilon_{iAmt}$ is symmetrically distributed around zero and uncorrelated with the systematic component of the subject’s preferences, the errors will introduce noise into measured WTP. Additional issues arise if the error term is correlated with the value difference Δ_t or has a nonzero mean. In this case, the data may lead to a systematic over- or underestimation of WTP. As a stylized example, suppose that $\epsilon_{iBmt} - \epsilon_{iAmt}$ is small (negative) whenever $v_{Bt} - v_{At}$ is large, i.e. Δ_t is small. Then the subject may frequently choose option A even though the utility from B is higher, leading to an overestimate of WTP.

Biases often occur due to framing or anchoring effects, because they introduce systematic (rather than random) deviations from true preferences: framing means that subjects choose one option more often only due to the way the choices are presented, either because the framing affects choice errors, or because it directly influences preferences (e.g.,

¹¹Indifference, or near indifference, could occur in hypothetical MPL, and in situations where the utility difference between options is very small, such as for subjects who strongly discount future payoffs and choose between two payments far in the future, or very risk averse subjects who make choices between two lotteries.

¹²Agranov and Ortoleva (2017) have shown that subjects exhibit stochastic choice in repeated identical decisions that directly follow each other.

¹³Others have previously pointed out both that ignoring measurement error in experimental preference elicitation can lead to biased estimates (e.g. Gillen et al. (2019)), and that detected measurement error may contain valuable information (Andersen et al., 2008).

Tversky and Kahneman (1985)). We argue that the MPL format has the potential to introduce framing effects because the two options are often presented in the same order both within MPL (e.g. with the value difference Δ_t always ascending) and within each binary choice (e.g. with the same option always shown on the left).

The potential for framing bias in binary evaluation tasks is well known; a famous example is the Galaxy Zoo project which asked “volunteer scientists” to assess the rotational direction of images of galaxies. Inflated counts of counter-clockwise rotation in early data led the researchers among other things to vary the arrangement of classification buttons on the screen.¹⁴ Anchoring effects could also arise because of the order within MPL. For example, in an MPL with decreasing Δ_t , the subject may select option A for the first few choices t , based on the utility difference from the options. As she is presented with repeated similar choices, her attention may decrease, or her early choices may even (temporarily) affect her subjective preferences. In both cases she will choose A more often as Δ_t decreases than she would have if presented with the same choice in isolation. In this case, the error term is negative across t (and possibly increasing in size), and as a result, the researcher may overestimate WTP for option A.¹⁵

Some forms of framing and anchoring in MPL are discussed in the existing literature, e.g. Andersen et al. (2007) (see also section 3.4), but to our knowledge, there is no widespread acknowledgement of potential order biases. Of the 23 studies in Appendix Table B, only three randomize the order of choices within the MPL, and none randomize the order of options within choice.

2.3 Approaches in the literature to MSB, NSB and interval data

Before discussing our proposal for dealing with choice error in MPL, we review the solutions that have been applied in the literature and point out some of their drawbacks. The existing research has introduced a variety of MPL design features aimed at reducing multiple switching and dominated choices *ex ante*, at the elicitation stage. In addition, a range of approaches aim to deal with observed MSB and NSB and the related problem of

¹⁴Between 2007 and 2019, over 400,000 volunteers on Galaxy Zoo carried out more than 11 million classification tasks (Raddick et al., 2019). In early data, it was found that a significantly higher than 50:50 share (around 52%) of galaxies were classified as rotating counter-clockwise. In a test, volunteers were therefore asked to rate mirror images of the same galaxies, and they again classified over 51% of galaxies as rotating counter-clockwise (Land et al., 2008), suggesting that the universe has no directional preference, but humans either have such a preference or click buttons in certain screen locations more frequently.

¹⁵We assume here that the interest is in a reproducible measure of WTP in contexts outside the MPL.

Table 1: Solutions in the literature for addressing MSB, NSB and interval data.

Stage	Challenge	Solution in the literature	Main issues with the solution
Elicitation	MSB	Elicit a single switchpoint: sMPL, iMPL	Error misattributed to WTP variation, changed outcome variance
	MSB	Stop MPL questions after first switchpoint	Measurement bias; error misattributed to WTP variation, changed outcome variance
Analysis	Inconsistent choices	Drop observations	Selection bias; loss of information, reduced outcome variance
	NSB	Impose minimal/maximal WTP	Measurement bias; potentially changed outcome variance
	Interval data, NSB	Assign single point in the interval, impose min./max. WTP (NSB)	Measurement bias; reduced outcome variance
	Interval data, MSB	Impose interval bounds (MSB) and use interval regression	Loss of information, potentially changed outcome variance

Notes: Overview of common approaches to addressing MPL data challenges arising from inconsistent choices (MSB and NSB with dominated choices) and interval identification (NSB and interval data). “Stage” refers to whether the solution is implemented during data collection (elicitation) or after (analysis). “Solution” and “Main issues” are discussed in more detail in the text.

interval identification when analyzing MPL data *ex post*, at the analysis stage. However, we argue that these solutions do not full address the issues with both idiosyncratic and systematic choice error in MPL. They often misattribute the individual error to variation in WTP, fail to account for order biases, and potentially even introduce new sources of bias, for example through selection. Table 1 summarizes these issues. In this section, we discuss how these solutions can lead to biased WTP measurement or incorrect standard errors. In Section 3.6, we illustrate this using data from South Africa.

At the elicitation stage, MPL design approaches often focus on avoiding or suppressing MSB by enforcing a single switch. The problem is that these solutions may not change the incidence of error, only its visibility. As a result, errors are never observed; variation that may simply be due to error is instead attributed to heterogeneity in WTP. For example, a “switching MPL” or sMPL shows subjects the full MPL and ask them at which point they would switch from option A to option B (e.g., Tanaka et al. (2010), Andersen et al.

(2007)). Andersen et al. (2008) proposed the so-called “iterative MPL” or *iMPL* which narrows down WTP in steps. In Berkouwer and Dean (2021), for instance, subjects are given a price in each step and asked if they would like to buy an improved cook stove. If they say yes, the price is increased, and otherwise it is decreased. In this manner, the procedure arrives iteratively at a small interval for the WTP. An issue with this approach may be that a wrong decision early on artificially increases measurement error.

Another common, and particularly problematic, approach is to elicit pairwise choices only up to the point of the first switch from one option to the other (e.g., Maffioli et al., 2020). Any choice errors in this format will bias WTP downward in price lists with ascending value difference, and upward when the value difference is descending. What is more, if subjects know that the MPL will end as soon as they switch options, it creates an additional incentive to switch early (e.g., if the subject’s opportunity cost of time is greater than the utility difference between the options). Note that the bias arising from recording only the first switch works in the opposite direction of any order bias from framing or anchoring described above. This may obscure order effects in data where only the first switch is observed.

Approaches to minimizing MSB that address the underlying sources of error, such as inattention or incomprehension, are rare. In one exception, Yu et al. (2021) successfully reduce MSB by “nudging” subjects to review and potentially revise their answers. Guiteras and Jack (2018) employ a similar approach in eliciting WTA for a piece rate casual labor contract. By repeating each choice four times, and providing an interpretation of the expected take-home pay after each, MSB was effectively eliminated. Note that, by reducing choice error, these approaches may also reduce framing and anchoring biases. However, they also increase implementation costs and may create other biases of their own, such as social desirability bias.

In cases where the data collection procedure does allow for inconsistent choices, methods for dealing with them at the data cleaning or analysis stage tend to introduce new sources of measurement error and bias. Some authors remove inconsistent choices (observations with MSB and sometimes NSB, typically when the experiment includes dominated choices, e.g., Dave et al. (2010)). This approach shrinks the sample size, reduces the outcome variation, and may introduce selection bias.¹⁶

A challenge presented by all MPL elicitation is the estimation of WTP from the discrete data that arises from a series of binary choices. The experimenter must make

¹⁶Some authors also count the number of times an option was chosen, without regard for the identity of the subjects who made these choices (see e.g. Holt and Laury, 2002).

assumptions about “true” WTP from the identified interval associated with a switch from one option to the other, by selecting either a single location or imposing a distribution of values within this interval. In many cases, either the midpoint or one end point of the interval is used, which results in artificially low variation and may introduce measurement error. In addition, observations that exhibit NSB must either be dropped from the dataset or modified to impose an arbitrary endpoint on the open interval.

The method that most closely respects the structure of the data is the interval regression approach proposed by Andersen et al. (2007), which employs a generalization of the tobit model. However, MSB observations remain a problem; the researcher has to make assumptions about the interval in which the subject’s WTP lies for MSB. Typically, the first and last observed switch are used, ignoring information from in-between switches.

In summary, existing approaches to MPL deal with choice error only incompletely. Moreover, the problem of order bias has been largely overlooked. This may be because in many contexts, uniform bias is not overly distorting: for example, in many applications of MPL, the main interest may be ordinal preferences or a simple sample split (e.g., into a more or less risk averse group). However, in WTP elicitation, the (cardinal) monetary value has meaning: a downward or upward shift of the distribution of elicited WTP due to bias may make the difference between a majority of the population expressing a positive vs. negative valuation for a service or policy.

In the next section, we propose a two-pronged approach to MPL design and estimation that uses randomization to detect order biases and a random utility model that includes fixed effects and order indicators to explicitly account for choice error and bias. In section 3.6 we use our approach to show with data from South Africa that (i) there is significant order bias in MPL data that affects estimated WTP, and (ii) many existing estimation approaches may exacerbate measurement issues.

3 Accounting for choice error and bias in MPL design and WTP estimation

3.1 Accounting for bias: MPL design

As discussed above, while the MPL format has many advantages, such as ease of understanding and rich data, it is plausible that it also induces some framing or anchoring effects. A challenge in examining order bias is that typical MPL implementations do not

have enough variation. Specifically, most MPL vary neither the order of binary choices presented, nor the order of options within the binary choice. If option A, say, appears before option B in every binary choice t , then any framing bias towards the first over the second option will be attributed to a higher WTP for option A. Similarly, if subjects are biased towards the option that is associated with the higher value in the first choice (due to anchoring), and all MPL are presented with a descending value difference Δ_t , then this leads to an upwardly biased estimate of the WTP for option A.

The MPL design we propose randomizes both the order of the choice sequence (ascending vs. descending value difference Δ_t) and which option is presented first (e.g. on the left or right, or as the first vs. second option in a verbal question). We call this randomizing *order within MPL*, and randomizing *order within binary choice*.

Randomizing order within MPL allows us to diagnose bias that arises from anchoring effects based on which option has higher value in the first binary choice (see estimation approach below). If order effects are symmetric and the randomization is balanced, it also minimizes the expected net bias in the data. Randomizing order within each binary choice has two roles. First, as above, if subjects have a preference for the option presented first vs. second, random variation in which option appears first/second will allow the experimenter to estimate this systematic error component and minimize net bias. Second, this randomization approach reduces concerns that inattention, or incomprehension, possibly combined with anchoring, could create “never-switching” that is interpreted as a strong preference but in reality only reflects how the MPL is presented.

Specifically, suppose a subject’s choices are primarily guided by the layout of the binary choice, say, they always choose the option on the left. In an MPL where all binary choices use the same order, this leads to NSB. In an MPL that randomizes within binary choice, it leads to MSB and the choice pattern is thus correctly attributed to error. In other words, randomization helps distinguish subjects who exhibit NSB simply because they do the exact same thing in each binary choice from subjects who actively choose the same option A or B each time. In addition, it will minimize average bias arising from subjects who always or often choose one side due to reasons unrelated to true WTP.

How does randomization affect the rate of error? As Andersen et al. (2007) have argued, full randomization of binary choices within MPL (i.e. the order of the rows) is unattractive because it makes the MPL task significantly more difficult from a cognitive perspective. We think of randomly presenting choices in ascending vs. descending order as a good compromise. However, we do randomize the order of options within each binary choice. The reader may worry that this may similarly increase attentional cost and lead

to higher error (e.g. the subject might mark the wrong option even after evaluating the options correctly due to the change in side). The randomization within binary choice may also make comprehension harder if the exact parallel organization of each decision helps subjects understand the structure of the experiment. We believe that these risks are small compared to the potential benefits. In fact, we conjecture that within-choice randomization may increase attention, since the choices look less repetitive and are less susceptible to “automation”. That said, future work could test the optimal presentation of binary choices to minimize a potential trade-off between detecting hidden, existing error and introducing additional error. Our MPL template survey instrument (described below) allows the user to implement any pattern of randomizing the order of options within binary choice, including no randomization and full randomization.¹⁷

3.2 Random effects latent utility estimation

Even if the MPL design would allow it, existing estimation approaches do not account properly for either systematic or idiosyncratic error. Building on our framework from section 2.2, we propose to estimate WTP with a latent utility binary choice model that captures both individual-level variation and choice-specific error and can also estimate the size of the systematic bias using data from a randomized MPL. This estimation approach takes seriously the structure of the data as a series of choices between two options, and it deals in natural ways with MSB, NSB, and interval-identified observations. The basic idea appeared early on in the literature on estimating risk aversion parameters from MPL (Holt and Laury, 2002; Harrison and Rutström, 2008; Harrison, 2008), and it is partially reflected in the use of interval regression for MPL data as well as in random-utility approaches found in the wider literature on WTP elicitation (e.g. Berry et al., 2020).¹⁸ However, binary choice models do not seem to be commonly used to estimate preferences from survey-based MPL experiments.

For simplicity, let us assume we have MPL data that compares only two options A

¹⁷In line with the view that full randomization is not desirable, the template does not support a fully random order of choices within MPL. This would require additional coding.

¹⁸Recently, (Apesteguia and Ballester, 2018) pointed out that the estimation of time and risk preferences from MPL data with a random-utility model may be affected by the fact that the (cardinal) utility difference between two options may not be monotonic in the risk or time preference parameter. As Conte and Hey (2018) note, this non-monotonicity is a property of the (systematic component of) the utility function; issues may arise here from pooling subjects and implicitly making comparisons across subjects. In our case, the monetary value serves a “normalizing” function. We essentially equate WTP with the (relative) subjective value of each option to the subject, and so the utility difference is, by definition, monotonic in WTP.

and B. The dependent variable is an indicator y_{it} that equals 0 if option B is chosen and 1 if option A is chosen. Following Section 2.2, we assume that person i chooses option A, and thus $y_{it} = 1$, if

$$WTP_i + v_{At} - v_{Bt} > \epsilon_{iBmt} - \epsilon_{iAmt}.$$

Here, the suffix m simply denotes the different ways in which the MPL may be presented. We substitute $\epsilon_{iAmt} - \epsilon_{iBmt} = -e_{it} + b_{mt} + s_{mt}$, where b_{mt} and s_{mt} capture order biases in the error term that depends on m and t (see below), and the remaining term $-e_{it}$ is assumed to have mean zero (the negative sign is assumed for convenience). We also decompose WTP_i into average willingness to pay WTP and an individual-specific component η_i . Finally, as above we let $\Delta_t = v_{At} - v_{Bt}$. Therefore, option A is chosen if

$$WTP + \eta_i + \Delta_t + b_{mt} + s_{mt} > e_{it}.$$

We adopt the convention that the binary choice t is synonymous with the value difference between options Δ_t and that Δ_t is decreasing in t . The MPL may be shown to subjects in ascending or descending order of Δ_t . If the MPL is shown starting at $t = 1$, Δ_t is descending and option A is most attractive relative to B early in the MPL. If the presented choices start at $t = T$, the value difference is ascending, and B starts out most attractive and then becomes less so in the course of the MPL. The order of choices is captured by the index m . In addition, especially in visual MPL presentations, different MPL versions m may reverse the order in which option A and B are presented to the subject within a given binary choice t (i.e. the MPL may vary side of screen or side of page).

We capture any effects that the variation in order has on choices with two bias parameters, b_{mt} for the order of choices within MPL, and s_{mt} for the order of options within binary choice. Suppose for some subject i , the value Δ_t starts high and declines, and the subject chooses A in the first presented choice. As discussed in Section 3.1, the subject may continue to choose A longer than if A started with a lower value relative to B. This situation would be captured by a “boost” to option A relative to B, and therefore a positive b_{mt} , and conversely for cases where the value difference is ascending. Similarly, s_{mt} captures whether one option is favored due to ordering within the binary choice. For example, if subjects prefer options shown on the left of the MPL, we would expect s_{mt} to be positive when A is on the left and negative when it appears on the right.

Assuming a normal distribution for the error term, we can estimate this model with a random effects probit procedure. To do so we write the probability of choosing option A

as

$$P(e_{it} < \alpha + \Delta_t + \beta_b x_{mt}^b + \beta_s x_{mt}^s + u_i),$$

where x_{mt}^b and x_{mt}^s are appropriately defined vectors of dummies used to estimate order biases (see below), the constant α provides an estimate of the average willingness to pay, and the term u_i accounts for individual variation in WTP. Even after controlling for preference variation and order biases, the e_{it} may be correlated within MPL, and we recommend clustering standard errors at the subject level.¹⁹

3.3 Identification of the WTP probit model

Average willingness to pay α . In a probit, utility is taken to be ordinal and all terms are scaled so that the variance of the error term equals 1. However, here we would like to express all terms, and in particular willingness to pay α , in terms of money. We achieve this by restricting the coefficient on Δ_t to be 1 and letting the error variance σ_e be identified off of the data. Equivalently, we can estimate a (standard normal) probit model without the coefficient restriction and re-scale all coefficient estimates. The coefficient estimate for Δ_t represents the inverse of σ_e in this case.

Individual willingness to pay $\alpha + u_i$ and choice error e_{it} . In most latent utility models, the error term is assumed to capture unobserved variation in true underlying utility. Here, we differentiate between the individual-level willingness to pay $\alpha + u_i$, and error e_{it} due to inattention, lack of comprehension, or near indifference (see section 2.2). The two distributions are separately identified because we observe a panel of T choices for each subject, as long as preferences are stable during the administration of the MPL, so that inconsistencies in expressed preferences within the same MPL are the result of choice errors. Both terms are assumed here to have a normal distribution, but other distributions are in principle possible; the most obvious being a logistic distribution for e_{it} which leads to a random effects logit.

The distributional assumptions on u_i and e_{it} discipline the distribution of WTP within each WTP interval. In the probit, the assumed distribution of latent indifference points across MPL intervals is normal with mean α . The location of the sample average will thus partly determine where the mass of indifference points within a given MPL interval is placed. If this seems restrictive, note that any method deriving point estimates for WTP from MPL data must make distributional assumptions. For example, when assigning the

¹⁹For datasets with multiple MPL per subject, we recommend clustering at the subject-by-MPL level.

midpoint of the interval as the switchpoint, the justification is often that any value in the interval is a priori equally likely and that the midpoint represents the average of these values. However, assigning the midpoint ignores the uncertainty from the distribution of WTP within the interval. Moreover, at the population level, this argument implies a “step function” for the probability distribution of WTP that is sensitive to the choice of MPL cutoff points and for which no latent utility model can provide a consistent explanation.

A drawback of the random-effects probit approach is that the model cannot distinguish individual variation in preferences from choice error when the subject does not exhibit MSB. For example, suppose the error terms e_{it} are strongly correlated within subject, perhaps due to anchoring. We cannot distinguish if a subject always chooses B due to anchoring or due to an idiosyncratic, strong preference for option B. This highlights the problem that even choices that are internally consistent may contain undetected error. In the most extreme case, no subject may exhibit MSB, which means that the estimated variance of the error σ_e will tend to zero, and the random-effects probit model may not be identified. Alternative estimation approaches in this situation are a probit without random effects, or interval regression (tobit). Both retain the normal distribution assumption but do not distinguish between systematic preferences and choice error. We argue that the absence of MSB does not guarantee the absence of error; subjects may grasp the structure of the MPL and therefore make choices consistent with a single switchpoint, yet still make errors when indicating this switchpoint. As discussed above, randomizing order within binary choice may help to reveal underlying choice error by inducing MSB in such cases.

Order bias b_{mt} and s_{mt} . So far we have not placed any restrictions on the order bias terms. Note, however, that a bias towards option A, e.g. when A appears high value first, measured by some dummy $x_{mt}^{b+} \in \{0, 1\}$, and a bias for option B and against option A, measured by another dummy x_{mt}^{b-} , creates a collinear set of variables when only one MPL is available. Intuitively, if we see a higher WTP in a descending MPL than in an ascending one, we do not know if this is due to the order bias in favor of option A, or the order bias in favor of B, or both. Absent other information, a natural assumption is that the order bias symmetrically favors whichever option appears high value first. If we additionally assume that this order effect on preferences is constant across binary choices t , we might define a combined indicator x_m^b which equals 1 when the MPL starts at $t = 1$ (descending Δ_t) and -1 when it starts at $t = T$ (ascending Δ_t). The coefficient on x_m^b measures the perceived monetary “boost” across binary choices for the MPL option that appears high value first. It is also possible to estimate the bias specific to each choice t ,

using T variables x_{mt}^b that take values -1 and 1 in ascending vs. descending MPL in choice t and 0 otherwise. In a parallel manner, we may define x_{mt}^s to equal 1 when A is presented on the left and -1 when it is on the right (either across choices or choice specific).

In order to relax the symmetry assumption and pin down the relative size of the bias for each option without additional MPL data (see below), researchers may choose to conduct outside validation. For example, they could ask (some) subjects to review and potentially revise the MPL responses, as in Yu et al. (2021), or measure the rates at which subjects renege on their original decision, as in Maffioli et al. (2020).

3.4 Extensions

Multiple MPL. We assumed above that we are measuring only one WTP parameter which expresses the relative willingness to pay between options A and B. In practice, researchers may want to estimate the WTP of the same subject for different options from different MPL. This can be achieved in one estimation by creating dummy variables that identify WTP for different option pairs. When estimating relative WTP between a set of different options, it is also possible to impose constraints on the WTP coefficients, such as transitivity. For example, the researcher may estimate $P(e_{it} < \alpha^1 z_m^1 + \alpha^2 z_m^2 + \alpha^3 z_m^3 + \Delta_t + \beta_b x_{mt}^b + \beta_s x_{mt}^s + u_i)$, where α^1 denotes the relative WTP for option 1 over option 2, α^2 denotes the relative WTP for option 2 over option 3, α^3 denotes the WTP for 1 over 3, and the dummies z_m^j indicate the corresponding MPL. We would expect that $\alpha^1 + \alpha^2 = \alpha^3$.

Covariates and WTP heterogeneity. The researcher may also want to allow for heterogeneity in WTP for groups of subjects or examine treatment effects of an experimental intervention. This can be straightforwardly accommodated by including group dummies or covariates and treatment indicator variables.

Other framing effects. We have focused our arguments on order biases. However, our proposed approach offers a framework for investigating other sources of variation in measured WTP. For example, as in Andersen et al. (2007), the researcher might want to test the effect of varying the range of values covered – the range of Δ_t – or the total number of binary decisions T within a given range. Differences in elicited WTP can be measured using dummies, in the same way we estimate order effects above. Again, the researcher would have to impose assumptions about the relative size of bias in any two different ways

of framing the MPL, or find a method to externally validate the estimates, for example by comparing WTP to market price, as in Andersen et al. (2007) and Burchardi et al. (2021). Alternatively, the researchers can impose that the bias is symmetric and the true WTP is the average of the estimates from differently framed MPL, or design an experiment that can measure the framing-specific order bias.

Option-specific order bias. As discussed above, it is possible that framing biases are not symmetric; in the case of order bias this would be true if option A receives a stronger “boost” with descending Δ_t than option B does with ascending Δ_t . With just one MPL, the two bias terms are not separately identified. However, if subjects complete a set of MPL in which multiple options are all compared with each other, an option-specific order bias can be estimated. This would require at least 3 options and 3 MPL that test all option combinations in the same experiment (see also example in “Multiple MPL” above). We briefly discuss an example of this in section 3.6.

3.5 Implementation

The baseline specification as well as all extensions can be implemented in the Stata procedure included in our estimation package ((Jack et al., 2022), see section 4 and Appendix S4). The user provides a data set with MPL, individual, and choice IDs, the binary outcome of each choice, and a variable containing Δ_t . In addition, covariates, group dummies, or order and framing dummies can be specified. The `mplwtp.ado` file uses the pre-programmed `xtprobit` routine for panel data and rescales the coefficients and standard errors so that the coefficient on Δ_t equals 1. The user can carry out various diagnostics and choose between scaled point estimates and standard errors generated by the delta method or a cluster bootstrap.

3.6 Application: WTP for prepaid electricity credit in South Africa

We use a field experiment in South Africa to demonstrate our approach to measurement and estimation and show evidence that order effects cause bias in MPL measurement. As described in greater detail in Jack et al. (2022), the project used MPL to measure WTP for prepaid electricity (delivered as a voucher that could be directly loaded onto the meter), with the original purpose of understanding the role of transaction costs and liquidity constraints.

Here we use the experimental data to illustrate four things: (1) implementation of the randomized order within MPL and order within binary choice, (2) MSB and NSB in these data, (3) estimation with a scaled random-effects probit, and (4) alternative approaches in the literature to addressing inconsistent choices and interval identification.

Implementation. Three different incentivized MPLs were administered as part of a survey and each subject was randomly assigned to receive two out of the three MPLs (N=767).²⁰ Here, we focus on one of the MPLs, administered to 506 subjects, which elicited a relative preference for receiving a transfer in the form of an electricity voucher or cash. The other two MPL elicited preferences over one voucher vs. receiving two vouchers on two different days, and two vouchers vs. cash. We include results for estimating WTP in all three MPL in Appendix A (see also below). One binary choice within one of the MPLs was drawn for implementation and the payoffs were immediately realized. We discuss some of our implementation choices in more detail in section 4.

The MPL randomized order both within MPL and within choice. We developed a SurveyCTO template and accompanying randomization files that can be used to carry out an MPL with these features and are described in section 4 and in Appendix S2. The survey module uses a visual representation of the options and shows binary choices one at a time. Subjects select their choice by tapping the screen of the tablet (with the help of an enumerator, if needed).

The structure of the MPL – here with descending value difference Δ_t and showing options in each choice in the same order – is shown in Figure 2. The MPL is designed so that the highest possible value that can be obtained in each choice was 100 Rand. By varying the value of the other option, the value difference increases strictly monotonically between binary choices.

MSB and NSB in our data. Figure 3 shows the number of “switches” observed in the MPL. Around 30% of subjects always choose electricity or always choose cash. Another 46% of subjects switch exactly once. The remainder – 24% – exhibit MSB.²¹ It is apparent that there is a higher share of odd than even switches, suggesting that subjects sometimes switch by mistake and then correct themselves. For example, 8% percent of subjects (33%

²⁰Subjects also went through one practice MPL before being presented with real choices.

²¹We code 12 subjects who switch from the option whose value is ascending to the option whose value is descending as switching twice (thus counting the first choice of the option with increasing value as a “switch”). This avoids lumping inconsistent and consistent choices.

Figure 2: MPL used to measure WTP for prepaid electricity in South Africa

Electricity vs. cash		
Binary choice	Electricity voucher value	Cash value
t	Option A	Option B
Choice 1	R100	R60
Choice 2	R100	R80
Choice 3	R100	R92
Choice 4	R100	R97
Choice 5	R100	R99
Choice 6	R99	R100
Choice 7	R97	R100
Choice 8	R92	R100
Choice 9	R80	R100
Choice 10	R60	R100

Notes: choices 1-10 were randomly shown in order (decreasing value difference) or in reverse order (increasing value difference). In addition, option A (electricity) and B (cash) were randomly shown either on the left or the right of the screen. Each binary choice was shown separately, accompanied by images showing the amounts either on a cash bill or on a stylized voucher. The respondent made choices by tapping directly on the image on the screen and then confirming the selection. See also figure S1 in appendix S2.

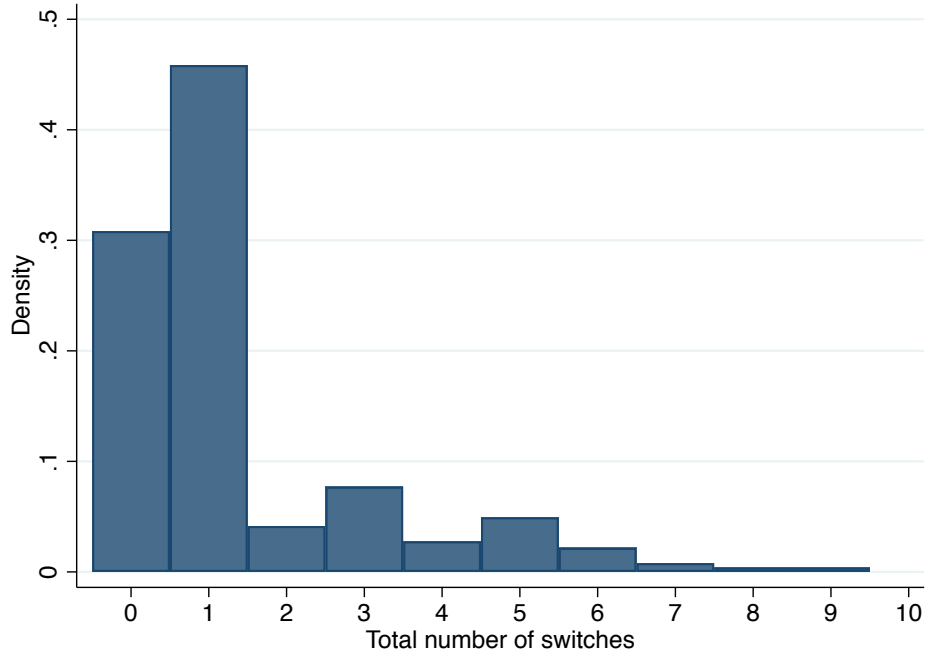
of those exhibiting MSB) have three switches, consistent with one error and subsequent correction.

The rate of MSB is relatively high, although not completely out of line with other MPL data from low-income populations. As discussed, one reason could be that within-choice randomization reveals errors that may otherwise remain undetected (e.g., in the form of NSB). That said, it is possible that the within-choice randomization increased error due to comprehension issues or difficulty using the screens.

The share of never-switchers in the data is also high. Since options were randomized within binary choice, it is unlikely that this is due to inattention that made the subject simply repeat the same choice many times. Only one individual always chose the same side of the screen, that is, exhibited a form of framing-based never-switching. The never-switchers give us a first opportunity to look into the magnitude of order biases. Never switchers are an interesting case because their choices are consistent with a very strong preference for one of the two options. Figure 4 shows the proportion of never-switchers who prefer the electricity voucher.

The left panel shows NSB subjects as a whole. In our data, 54% of all never-switchers

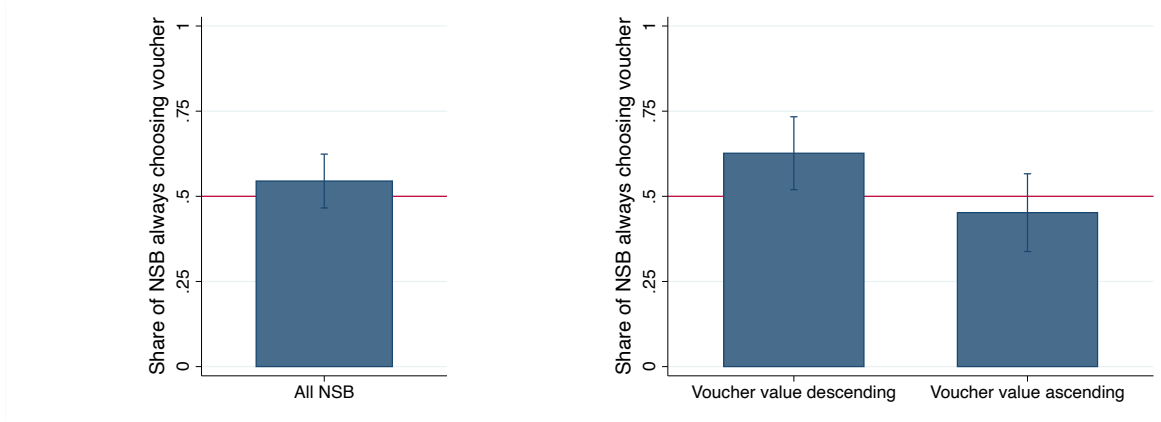
Figure 3: NSB and MSB in South Africa data



Notes: Total number of times the subject switched between options in the 10-item MPL. 0 switches indicate NSB. 0 or 1 switch are consistent with utility maximizing behavior. 2 or more switches constitute MSB and indicate inconsistent choices.

express a strong preference for the voucher, while 46% choose cash in all choices. Thus, on average, this group has a slight preference for electricity. However, the right panel reveals that the share of NSB who prefer the voucher varies depending on the order in which the choices were presented. The figure shows the share of never switchers who always choose the voucher conditional on whether the value of the voucher was descending or ascending in value. When the value is descending, the share of NSB choosing electricity is 63%. However, when it is ascending, the share is only 45%. At the sample level, this suggests that order biases can be quite powerful and shift preferences towards the high-value option in the first MPL choice. At the individual level, however, we cannot distinguish errors or order biases from true preferences. In Jack et al. (2022), we consider reasons why respondents may have strong preferences for both vouchers or cash: frictions due to liquidity issues and transaction costs, combined with unexpected shocks, may lead households to experience unplanned shortages and exhibit a (potentially temporary) high marginal rate of substitution of one for the other.

Figure 4: Share of subjects always choosing voucher, among all NSB



Notes: The figure includes only subjects who exhibit never-switching behavior (NSB), that is, they chose either always the electricity voucher, or always cash. The left panel shows the total share of NSB subjects choosing the voucher always. The right panel shows the share choosing the voucher always, conditional on the order in which the MPL was presented (voucher value descending or ascending). The red line marks an even share of 0.5.

Latent utility estimation. Next, we show results from implementing the random effects probit estimation as in Section 3.2. As described, we scale the estimated coefficients to express all terms in money (here South African Rand).²²

The first row of Table 2 shows estimated WTP for receiving an electricity transfer over cash in four different model specifications. Columns (1) and (2) restrict the sample to either decreasing or increasing value difference, to mimic many standard MPL designs. In column 1, the value difference is decreasing, so option A (electricity) is higher relative value at the start; in column (2) it is increasing. Consistent with the anchoring effects we discussed above, the WTP estimate for a voucher over cash is positive and significantly different from zero in the first column, but negative and imprecisely estimated in the second column. Using one or the other MPL design, we would have drawn very different conclusions about the value that households attach to receiving a voucher over cash. This is evidence for the importance of randomizing order: it shows that a single ascending or descending MPL would have resulted in biased WTP estimates, even when using the latent-utility estimation approach.

In column (3), the data are pooled so that approximately half of the sample are

²²In Table 2, we show results with standard errors calculated using the delta method. In Appendix table A.1, we show results after bootstrapping the estimation. See Appendix S4 for further discussion of these approaches.

Table 2: Scaled random effects probit estimates.

	Biased sample		Full sample	
	(1)	(2)	(3)	(4)
WTP for elect. voucher	8.879*** (2.58)	-2.360 (3.13)	4.015** (2.01)	3.590* (2.00)
Order within MPL				5.811*** (2.00)
Order within choice				0.435 (0.65)
SD of individual error	20.950	26.339	23.388	23.377
SD of RE	39.551	44.458	42.111	41.676
N(choices)	2,710	2,350	5,060	5,060
N(subjects)	271	235	506	506

Notes: Columns 1 and 2 use observations from subjects who saw the value of the electricity voucher (option A) descending and ascending, respectively. Columns 3 and 4 pool all observations. Column 4 adds order controls. Standard errors are clustered at the subject (MPL) level.

presented with either order within MPL. Note that an MPL design with descending Δ_t as in column (1) would have led us to conclude that subjects are willing to pay a nearly 10% tax on a transfer in order to receive that transfer in the form of an electricity voucher. Column (3) demonstrates that this yields a WTP for the voucher that is 100% higher than the sample average.

The table also reports the standard deviation of the choice-specific error term and the random effect. Both are large, indicating considerable variation in preferences as well as significant error rates – a reflection of the high share of NSB and MSB in our data.

In Appendix A, Table A.2, we show estimates for all three of the MPLs that we implemented in the original data collection (pairwise comparisons of one voucher immediately, two vouchers sent two days apart, and cash). We note that the individual error variance is somewhat lower, and the WTP for one voucher is higher and highly significant, in the MPL that compares one electricity voucher and two electricity vouchers. Both options are in the same “domain” and it appears that subjects had stronger preferences; it is possible that outside factors or inattention therefore confounded elicited preferences less.

Appendix Table A.2, column (4) also demonstrates that we can pool the data and

estimate relative willingness to pay for each pairing of options offered at once. This approach makes it possible, for example, to test joint hypotheses about the coefficients, such as transitivity.²³

Order biases. Column (4) of table 2 adds basic controls for order biases, by defining a variable that equals 1 (-1) when the MPL is descending (ascending), and a second variable that is 1 (-1) when option A is on the left (right) of the screen. The coefficients on these variables measure the average bias for the option shown with high value first (“order within MPL”) and on the left (“order within choice”), respectively.

The estimates show that order within MPL is an important determinant of choice. Specifically, subjects on average express a preference for the option that appears of higher value first that is equivalent to a payment of nearly 6 Rand. By comparison, order within choice does not seem to affect expressed preferences significantly.²⁴ Similarly, the estimates in Appendix Table A.2 all show significant effects of order within MPL, but not order within binary choice.

The approach in column (4) of Table 2 assumes both that the bias induced by the order within MPL is constant across t , and that it symmetrically favors option A and option B. An inspection of the proportion of subjects choosing option A for each binary choice by MPL order shows that this is likely a simplification and the bias is not constant for all binary choices. In Appendix A, Table A.2, column (5) we therefore include bias terms for each binary choice t separately. Interestingly, the relative bias first increases, then decreases in t . This suggests that order bias might be more pronounced when the options are closer in value and therefore harder to distinguish. However, the WTP estimates themselves are not much affected by including bias terms for each binary choice.

To investigate the symmetry assumption, column (6) in Table A.2 replaces the symmetric “order within MPL” variable with three variables that separately measure order bias in favor of each of the three options. Note that this replaces the symmetry restriction with the assumption that the bias in favor of a given option is constant across MPL comparing different options. The results suggest that order bias for cash is stronger than for the two electricity options. While the qualitative conclusions about the WTP do not

²³It would also be possible to directly impose transitivity or other restrictions when estimating these three WTP values in one model. This option is not currently implemented in our Stata routine and we do not show it here.

²⁴We do not have enough variation in the total number of times options switched sides, so we cannot test interaction effects, that is for example, whether more or less variation in the order within binary choice affects the magnitude of the within-MPL order bias.

change much, the point estimates shift somewhat. Note, however, that a different interpretation of the data is that the (symmetric) order bias is simply lower in the MPL that compares one and two electricity vouchers vs. the MPLs that compare electricity and cash (see columns (1)-(3) of Table A.2). Note also that this approach to estimating bias is only possible with data from at least 3 MPL, and so in many cases may not be a strategy available to the researcher.

(Re-)Introducing bias through standard procedures In Section 2.3 and Table 1, we described some common approaches to addressing MSB, NSB and the interval nature of MPL data. Recall that all approaches discussed above distill the information from a subject’s MPL into an interval or even just a single point; some additionally drop inconsistent information or otherwise restrict the data. Here, we replicate some of these approaches to demonstrate how this influences estimated WTP.

We make our points mostly using interval regression estimates. Interval regression is a generalization of a tobit specification and uses one upper and lower bound for the WTP of each subject to estimate average WTP for an electricity voucher. Therefore the researcher has to assign an interval for WTP to each subject based on the observed MPL choices. This is often done by using the first and last observed “switch”. Subjects with MSB are thus assigned a wider interval to accommodate all of their choices, and switches between the interval endpoints are effectively not used.²⁵ Subjects with NSB can be assigned one-sided (open ended) intervals. They may alternatively be assigned a maximum or minimum of WTP based on the researcher’s assessment which values of WTP are plausible. For example, in the South Africa data, a WTP below -100 would indicate that a subject would rather pay cash than receive the R100 electricity voucher.

Table 3 columns (1) to (4) implement these two variants of interval regression (unrestricted max./min. WTP or restriction to ± 100). For all regressions in the table, odd columns use observations with descending value difference, and even columns with ascending value difference, mimicking again typical MPL data. They can therefore be compared with columns (1) and (2) in Table 2, respectively. As in the scaled probit, the WTP estimates are sensitive to choice order, but they are also attenuated toward zero; slightly more so when a maximum and minimum WTP are imposed in columns (3) and (4).²⁶

Some MPL implementations also suppress MSB observations, either ex ante or ex

²⁵This follows the approach in Andersen et al. (2007).

²⁶Note that the number of subjects is lower in (1) and (2) because of subjects whose upper and lower bounds are both undefined. This happens if a subject chooses option B (the lower-value option) in $t = 1$ but option A in $t = 10$ (again the lower-value option).

post. Columns (5)-(8) continue to use interval regression to maintain comparability, and implement two versions of these sample restrictions. One common approach is to stop the data collection once a subject switches for the first time. We approximate this by using the WTP interval associated with only the first switch and re-estimating the interval regression. As shown in columns (5) and (6), the point estimates change substantially and standard errors increase. Interestingly, comparing column (5) with column (1) shows that the measurement bias from using the first switch essentially erases the order bias in the opposite direction (see discussion in section 2.3). We note that Channa et al. (2021) and Fuller and Ricker-Gilbert (2021) both test WTP for different kinds of grain quality verification. Both papers randomize the order within MPL (a small minority among MPL implementations), but also only use the first switch. Fuller and Ricker-Gilbert (2021) do not find an order effect, whereas Channa et al. (2021) document a strong order effect among farmers but not traders. It is possible that a full MPL elicitation (i.e., not stopping after the first switch) would have shown order bias in all samples. Given the popularity of the “first switch” MPL, it is possible that other researchers have piloted order randomization before us but dismissed order bias concerns in MPL based on the data.

Another, fairly drastic approach to addressing inconsistency is to throw out observations with MSB. We implement this in columns (7) and (8) of Table 3. In our data and within the interval regression approach, excluding MSB leads to more extreme estimated WTP (depending on order within MPL). We are agnostic whether this is due to selection bias or a sharper estimate, but note that the reduced sample size shows that we would be discarding a large number of data points. This seems in general an unsatisfactory approach.

Many times, MPL data analysis does not use interval regression, but instead reduces the WTP interval to a single point and assigns the minimum, midpoint or maximum of the interval as the individual measure of WTP. In our data, this has a drastic effect on implied WTP. For example, when imposing a WTP minimum of -100 and maximum of 100 (for NSB observations) and using the minimum, maximum or midpoint of the WTP intervals defined by the first and last switch, we get an average WTP of -18.0, 2.7 and 23.4 Rand, respectively.

Overall, neither the estimates from Table 3 nor the single-point estimates reliably reproduce our scaled-probit results, and the different methods lead to a dizzying array of estimated WTP levels. This confirms our conjecture from section 2.3 that the way the literature has dealt with choice error is inconsistent and unreliable in the context of

Table 3: Alternative approaches: Sample or data restrictions

<i>Sample/Data</i>	Inf. interval		100 endpoint		First switch only		No MSB	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WTP	6.366*** (2.45)	-1.551 (2.84)	6.086*** (2.31)	-1.459 (2.62)	1.824 (2.73)	3.695 (3.11)	8.732*** (2.87)	-3.943 (3.96)
N(subjects)	266	228	271	235	271	228	223	166

Notes: Interval regression estimates, using the first and last switch to define the interval in the case of MSB. Columns 1 and 2 allow an infinite interval for NSB. Columns 3 and 4 impose an endpoint of ± 100 on the open interval for NSB observations. Columns 5 and 6 use the first observed switch only. Columns 7 and 8 drop subjects who exhibit MSB. In odd-numbered columns, the value of the voucher is descending; in even-numbered columns, it is ascending.

WTP estimation. In particular, the single-point approaches yield WTP estimates that are so sensitive to the choice of point within the interval to be more or less useless. But even the interval regression estimates deviate from the probit estimates. Moreover, in all cases order effects reverse the sign and alter the significance level of the WTP estimate, emphasizing the need for order randomization.

We note one exception regarding using scaled probit over interval regression, and that is when subjects do not exhibit MSB. In this case, which may be more common in high-income settings or in the laboratory, interval regression or simple probit is the correct approach due to the identification problems in random-effects probit when there are no observed choice errors (see above).

4 MPL implementation

In the course of our own work and while surveying the literature, we identified a number of other design and implementation features besides order randomization that can help to decrease bias and individual error. We discuss some of these design features below. The SurveyCTO template that we provide as part of this paper, described in Appendix S2, implements these features along with a very flexible specification for order randomization. Many of these features may also be beneficial in other implementations of MPL, such as pen-and-paper questionnaires or laboratory experiments.

4.1 Best practices

Keeping maximum payout constant. As seen in Table 2, the MPL we used in South Africa keeps the maximum payout $\max\{v_a, v_b\}$ in each choice constant. This feature is also implemented in Allcott and Kessler (2019), see Figure 1. Keeping overall value constant makes it more likely that attention levels remain similar across the MPL. Moreover, the MPL incentive appears fair to subjects: regardless which binary choice is selected for implementation, the maximal monetary value the subject could have obtained remains the same. In manual implementations of an incentivized MPL where the realized choice is determined by a draw or dice roll, it also avoids irregularities in the randomized selection of the implemented choice.²⁷

Value difference decreases. It can be helpful to elicit smaller WTP (value difference) intervals around points that matter for hypothesis testing. In the South Africa Electricity example, this is the point where the value difference and therefore WTP is zero. If the interval where subjects switch from one option to the other is small, the estimation of willingness to pay will be more precise. We recommend calibrating the value steps in the MPL in piloting. Again, this feature is also implemented in Allcott and Kessler (2019).

Binary choices presented separately. If MPL are displayed as a list on a single page, some subjects may evaluate the MPL as a whole instead as a series of binary choices. This may increase anchoring effects. From our piloting work we concluded that showing each new binary choice on a new page or screen, combined with within-choice randomization, helps increase attention.

Practice MPL. Before the MPL of interest is implemented, it is in our experience helpful to let subjects complete a practice MPL. Our practice MPL pays out a randomly selected binary choice for items of relatively low value and unrelated to the research question; in South Africa these were types of candy. In addition to improving comprehension of the format of an MPL, demonstrating the realization of one binary choice increases attention and emphasizes the independence of individual binary choices. We recommend short practice MPLs (2 - 3 binary choices). Their format should be as similar as possible to the actual MPL.

²⁷One of the authors (Sautmann) found in a pilot implementation of a time-preference MPL that the distribution of randomly drawn implemented choices was markedly skewed towards high-value choices, likely because surveyors let subjects take a second random draw when the first draw yielded a low payout.

As part of the practice MPL, subjects may review the choices they made in each practice binary choice. They may be asked to consider whether they would have been pleased with the MPL payout if a given choice had actually been randomly selected. This can decrease the chance that subjects view the MPL as a test with a “correct” answer. Many papers carry out practice elicitation, including, for example, Fuller and Ricker-Gilbert (2021).

Formatting. An engaging design and clear formatting and layout can increase attention, improve understanding, and decrease fatigue and mistakes. When subjects complete the MPL themselves, formatting choices such as a large font size and images can increase comprehension, including in contexts with low literacy. Attention may also be increased by prompting subjects to verify their response to a binary choice before moving on to the next one. Other visual aids, such as showing an animated coin flip at the end of the MPL, can help convey features of the MPL, here that the realized choice is selected randomly.

4.2 Implementation Package

As part of this paper, we provide an implementation package for MPL data collection, processing, and analysis, described in detail in a technical appendix.

The package can create a SurveyCTO template that implements the design features above, along with randomization within MPL and within binary choice. SurveyCTO is a common survey tool used for conducting in-field surveys via an app on a phone or tablet. The SurveyCTO file is fully automatically generated from a Stata script together with a set of user-defined inputs, as described in Appendix S2. The template supports flexible randomization specifications for various elements of the MPL and can accommodate a pre-existing survey sample. With small changes, the SurveyCTO template can be used for a wide range of preference elicitation experiments, including risk aversion or time preferences. It can also be adapted for other Open Data Kit (ODK) data collection platforms.

We also provide Stata programs that prepare the data generated by the SurveyCTO questionnaire for analysis (Appendix S3), and an ado file that specifies a command to estimate WTP with our proposed scaled random-effects probit (Appendix S4). The command can output standard errors via the delta method or via cluster bootstrap and supports various diagnostics to ensure accurate estimates.

5 Conclusion

Measuring willingness to pay for goods and services is an important part of many research studies. Among other things, information on WTP can provide insights about welfare (e.g., Allcott and Kessler, 2019), allow cost-benefit comparisons (e.g., Cole and Fernando, 2020), and help explain treatment effects (e.g., Guiteras and Jack, 2018). Multiple price list elicitation is attractive for many reasons. It offers a compromise between the precise information revealed by BDM and the simplicity of TIOLI, and data from repeated choices can in principle be used to learn about error from inconsistent and dominated choices. As we show in this paper, in combination with standard MPL design features this error can introduce bias in WTP estimates.

We propose a straightforward, two-part approach to revealing both idiosyncratic and systematic choice error and incorporating it into WTP estimation. First, random variation in the order of choices within MPL and within binary choice reveals bias and minimizes its net impact on the estimated WTP. Second, latent utility estimation using random effects probit accommodates multiple switching, never switching and interval data, and can be used to estimate bias terms and both individual choice error and subject-level variance in WTP. To support other researchers interested in adopting these innovations, we offer a SurveyCTO package and a Stata ado file, along with the necessary instructions.

We focus on a single case study to demonstrate both the challenges with alternative approaches and the strengths of our approach. This case is intended to be illustrative and we are well aware of the limited general conclusions we can draw. For example, WTP estimation in populations that exhibit less error and bias will benefit less from both the randomized choice implementation and the estimation approach. Note, however, that only by randomizing the order of choices in the MPL and options within each binary choice researchers can test for any framing or anchoring effects in new samples.

A second limitation of our approach is that the correction of order bias can be applied to the study sample, or to large enough subgroups within the study sample, but likely not to individual-level data. The implication is that MPL data – carefully implemented and analyzed – is appropriate for measuring sample average WTP, but not individual-specific WTP. It should also be noted that inattention, incomprehension, and framing effects can lead to choice error in other elicitation methods too. Given the prevalence of these errors in WTP estimation, the method we propose in this paper offers a way forward: an estimation approach that correctly accounts for individual error and a measurement approach that reveals bias.

In the MPL implementation in this paper, we only varied whether the MPL has an ascending or descending value difference, while we allowed order within each binary choice to be fully randomized. Future work might test whether the degree of random variation in MPL implementation can be further optimized to minimize the incidence of biases and errors. Specifically, there could be a trade-off between too little randomization introducing net bias and preventing the detection of latent choice errors, and too much randomization increasing the cognitive burden and thereby provoking choice error that did not previously occur.

References

- Agranov, M. and P. Ortoleva (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy* 125(1), 40–68.
- Allcott, H. and J. B. Kessler (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics* 11(1), 236–76.
- Alphonse, R. and F. Alfnes (2017). Eliciting consumer WTP for food characteristics in a developing context: Application of four valuation methods in an african market. *Journal of Agricultural Economics* 68(1), 123–142.
- Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2007). Valuation using multiple price list formats. *Applied Economics* 39(6), 675–682.
- Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008). Eliciting risk and time preferences. *Econometrica* 76(3), 583–618.
- Apestequia, J. and M. A. Ballester (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy* 126(1), 74–106.
- Asioli, D., A. Mignani, and F. Alfnes (2021). Quick and easy? Respondent evaluations of the Becker-DeGroot-Marschak and multiple price list valuation mechanisms. *Agribusiness* 37(2), 215–234.
- Bartling, B., F. Engl, and R. A. Weber (2015). Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap. *Journal of the Economic Science Association* 1(1), 72–85.
- Ben Yishay, A., A. Fraker, R. Guiteras, G. Palloni, N. B. Shah, S. Shirrell, and P. Wang (2017). Microcredit and willingness to pay for environmental quality: Evidence from a randomized-controlled trial of finance for sanitation in rural Cambodia. *Journal of Environmental Economics and Management* 86, 121–140.
- Bénabou, R., A. Falk, L. Henkel, and J. Tirole (2020). Eliciting moral preferences: Theory and experiment. *Working paper*.
- Berkouwer, S. B. and J. T. Dean (2021). Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households. *Working paper*.
- Berry, J., R. Dizon-Ross, and M. Jagnani (2020). Not playing favorites: An experiment on parental fairness preferences. *NBER Working Paper 26732*.
- Berry, J., G. Fischer, and R. Guiteras (2020). Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana. *Journal of Political Economy* 128(4), 1436–1473.
- Björkegren, D. (2020). Competition in network industries: Evidence from the rwandan

- mobile phone network. *Working paper*.
- Burchardi, K., J. De Quidt, S. Gulesci, B. Lerva, and S. Tripodi (2021). Testing willingness to pay elicitation mechanisms in the field: Evidence from Uganda. *Journal of Development Economics* 152, 102701.
- Burchardi, K., J. De Quidt, S. Gulesci, and M. Sulaiman (2021). Credit constraints and demand for remedial education: Evidence from Tanzania. *Working paper*.
- Bursztyn, L., B. Ferman, S. Fiorin, M. Kanz, and G. Rao (2018). Status goods: Experimental evidence from platinum credit cards. *The Quarterly Journal of Economics* 133(3), 1561–1595.
- Cason, T. N. and C. R. Plott (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy* 122(6), 1235–1270.
- Cassar, A., F. Wordofa, and Y. J. Zhang (2016). Competing for the benefit of offspring eliminates the gender gap in competitiveness. *Proceedings of the National Academy of Sciences* 113(19), 5201–5205.
- Channa, H., J. Ricker-Gilbert, H. De Groote, and J. Bauchet (2021). Willingness to pay for a new farm technology given risk preferences: Evidence from an experimental auction in Kenya. *Agricultural Economics* 52(5), 733–748.
- Cole, S., A. N. Fernando, D. Stein, and J. Tobacman (2020). Field comparisons of incentive-compatible preference elicitation techniques. *Journal of Economic Behavior & Organization* 172, 33–56.
- Cole, S. A. and A. N. Fernando (2020). ‘Mobile’izing agricultural advice technology adoption, diffusion and sustainability. *The Economic Journal* 131(633), 192–219.
- Conte, A. and J. D. Hey (2018). Rehabilitating the random utility model. A comment on Apesteguia and Ballester (2018). *Working paper*.
- Crosetto, P. and A. Filippin (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics* 19(3), 613–641.
- Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty* 41(3), 219–243.
- Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies* 83(2), 587–628.
- Filippin, A. and P. Crosetto (2016). A reconsideration of gender differences in risk attitudes. *Management Science* 62(11), 3138–3160.
- Fuller, A. J. and J. Ricker-Gilbert (2021). Estimating demand for third-party quality testing in rural grain markets: Evidence from an experimental auction for measuring

- moisture content in Kenya. *Journal of African Economies* 30(5), 389–417.
- Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. *Journal of Political Economy* 127(4), 1826–1863.
- Guiteras, R. P. and B. K. Jack (2018). Productivity in piece-rate labor markets: Evidence from rural Malawi. *Journal of Development Economics* 131, 42–61.
- Harrison, G. W. (2008). Maximum likelihood estimation of utility functions using Stata. *Working paper*.
- Harrison, G. W. and E. E. Rutström (2008). Risk aversion in the laboratory. In J. C. Cox and G. W. Harrison (Eds.), *Risk Aversion in Experiments*, Volume 12 of *Research in Experimental Economics*, pp. 41–196. Emerald Group Publishing Limited.
- Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *American Economic Review* 92(5), 1644–1655.
- Jack, B. K., K. McDermott, J. Romine, and A. Sautmann (2022). Paying for power: Utility bills and the spending patterns of the poor. *Working paper*.
- Jack, K., K. McDermott, A. Sautmann, K. Chan, and L. Yue (2022). Implementation package: Multiple price lists for willingness to pay elicitation. URL: <https://github.com/MPL-WTP/MPL-WTP>.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy* 98(6), 1325–1348.
- Land, K., A. Slosar, C. Lintott, D. Andreescu, S. Bamford, P. Murray, R. Nichol, M. J. Raddick, K. Schawinski, A. Szalay, D. Thomas, and J. Vandenberg (2008, 08). Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 388(4), 1686–1692.
- Lang, G. and B. Lanz (2021). Energy efficiency, information, and the acceptability of rent increases: A survey experiment with tenants. *Energy Economics* 95, 105007.
- Maffioli, A., D. McKenzie, and D. Ubfal (2020). Estimating the demand for business training. *Working paper*.
- Raddick, M. J., E. E. Prather, and C. S. Wallace (2019). Galaxy zoo: Science content knowledge of citizen scientists. *Public Understanding of Science* 28(6), 636–651.
- Squires, M. (2021). Kinship taxation as an impediment to growth: Experimental evidence from Kenyan microenterprises. *Working paper*.
- Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review* 100(1), 557–71.

Tversky, A. and D. Kahneman (1985). The framing of decisions and the psychology of choice. In *Behavioral decision making*, pp. 25–41. Springer.

Yu, C. W., Y. J. Zhang, and S. X. Zuo (2021). Multiple switching and data quality in the Multiple Price List. *The Review of Economics and Statistics* 103(1), 136–150.