

Buters, Nienke; Prochazka, Jakub; Schefbänker, Katrin; Strobl, Andreas; Teichmann, Karin

**Working Paper**

## A Comment on "Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation"

I4R Discussion Paper Series, No. 137

**Provided in Cooperation with:**

The Institute for Replication (I4R)

*Suggested Citation:* Buters, Nienke; Prochazka, Jakub; Schefbänker, Katrin; Strobl, Andreas; Teichmann, Karin (2024) : A Comment on "Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation", I4R Discussion Paper Series, No. 137, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/301426>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

No. 137

I4R DISCUSSION PAPER SERIES

# **A Comment on “Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation”**

Nienke Buters

Katrin Schefbänker

Karin Teichmann

Jakub Prochazka

Andreas Strobl

**August 2024**

## I4R DISCUSSION PAPER SERIES

I4R DP No. 137

### **A Comment on “Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation”**

**Nienke Buters<sup>1</sup>, Jakub Prochazka<sup>2</sup>, Katrin Schefbänker<sup>3</sup>,  
Andreas Strobl<sup>4</sup>, Karin Teichmann<sup>3</sup>**

*<sup>1</sup>Technical University of Munich/Germany*

*<sup>2</sup>Masaryk University, Brno/Czech Republic*

*<sup>3</sup>University of Innsbruck/Austria*

*<sup>4</sup>Lancaster University, Lancaster/Great Britain*

**AUGUST 2024**

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

#### **Editors**

**Abel Brodeur**  
*University of Ottawa*

**Anna Dreber**  
*Stockholm School of Economics*

**Jörg Ankel-Peters**  
*RWI – Leibniz Institute for Economic Research*

## **A comment on “Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation”**

Nienke Buters (Technical University of Munich, Germany)

Jakub Prochazka (Masaryk University, Brno, Czech Republic)

Katrin Schefbänker (University of Innsbruck, Austria)

Andreas Strobl (Lancaster University, UK)

Karin Teichmann (University of Innsbruck, Austria)

### **Abstract**

This report inspects the reproducibility of a study by Dizon-Ross and Jayachandran (2023), which focused on differences in parents' spending on their daughters relative to sons on a large sample of 6,673 observations in 1,084 households in Uganda. The original study found that the willingness to pay (WTP) of fathers for different goods for their daughters was lower than for their sons. We were able to computationally reproduce all original results using the original data and code. To test for recreate reproducibility, we tried to reproduce the results of the main analyses using a new code and different software. We were not able to complete the reproduction without analyzing the original code and processed dataset. It was not clear from the manuscript nor the online appendix how the authors dealt with the multilevel structure of the data and how they controlled for different goods, which served as stimulus material. Because the raw data did not have clear labels and the replication package did not include a codebook, we were also unable to identify the variables needed for each analysis. However, after analyzing the original code, we were able to reproduce the original results in MPLUS. The missing code book and missing transcription of survey questions caused complications for investigating robustness reproducibility. Although the authors collected a large number of variables and provided them in the dataset, it was not possible to identify their meaning. Therefore, we were not able to conduct further analyses regarding the main findings of the study. Consequently, we only focused on

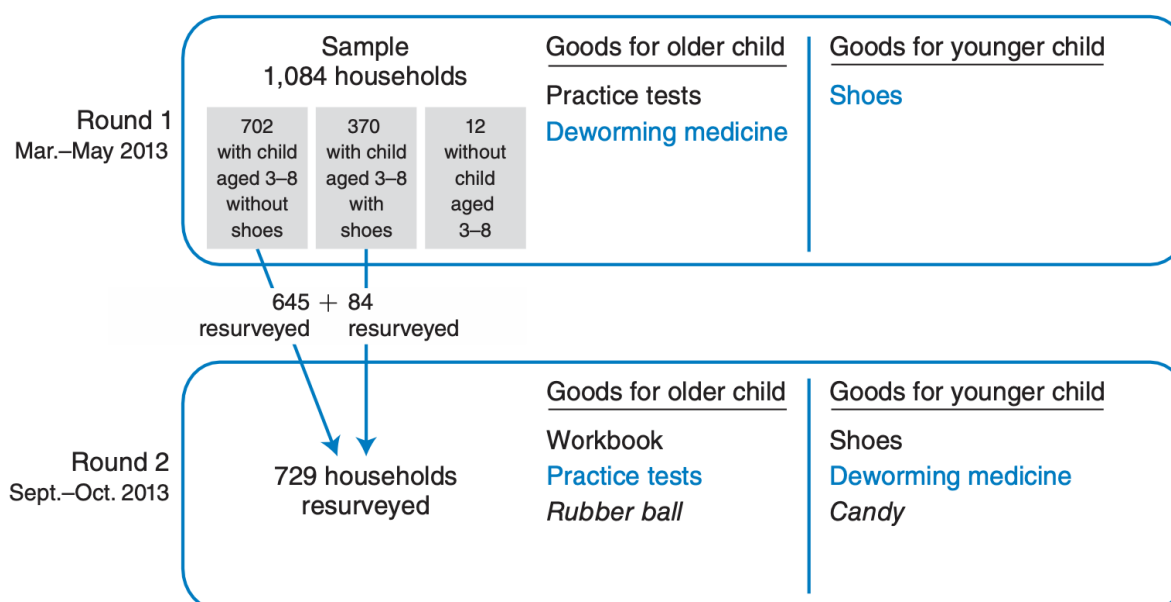
multicollinearity checks and different constellations of the control variables reported in the paper within the robustness checks. Our analyses showed that the results of the study are robust in this respect. In addition, the missing code book and transcription of survey questions did not allow for direct replicability of the study. Conceptual replicability was not investigated.

## 1. Introduction

The objective of Dizon-Ross and Jayachandran's (2023) study is to examine if and why fathers underspend on their daughters relative to mothers, and to investigate the underlying reason for the mother-father differences.

The paper measures parents' spending preferences by eliciting and comparing mothers' and fathers' willingness to pay (WTP) for various goods (human capital and enjoyment goods) for their sons and daughters. The study was conducted in Uganda over a period of two rounds. In the first round, a sample of 1,084 households with young children was surveyed and in the second round, 729 households were resurveyed (see figure 1). The authors either interview one randomly selected parent or sometimes both parents (the other parent in round 2). The study also uses fixed effects and clustering to account for household-level variation.

**Figure 1. Sampling methodology with nonincentivized goods (blue text) and enjoyment goods in *italics* (Dizon-Ross and Jayachandran, 2023).**



The authors tested between two hypotheses for gender-based spending differences:

1. The *preference-based explanation*: fathers care more about their sons than their daughters, and mothers do not.
2. The *investment-based explanation*: differences between mothers and fathers could be due to different returns to investment. Mothers may believe more than fathers that they are more dependent on support from their children in later years, and that daughters are more likely to support their parents in later years than sons are.

To test between these hypotheses, the authors examine whether there are similar gender-based differences in WTP for goods that bring joy to the children but do not add to their human capital: toys and candy (“enjoyment goods”). If the investment-based explanation holds, they expect gaps in WTP for human capital goods (i.e., shoes and practice tests) but not enjoyment goods. However, if there are no differences between the type of good, this would point to a preference-based explanation.

The results support a preference-based explanation for the mother-father differences in spending, suggesting that mothers care about daughters relatively more than fathers do. Overall, fathers spend 5% less on their daughters than on their sons. These results are robust against different specifications such as normalizing WTP for each good to its market value or the parents' reason for valuing each good. These spending differences are similar for enjoyment goods (i.e., candy) and for human capital goods (i.e., shoes), where fathers spend around 0.1 standard deviations less on their daughters than on their sons for both goods. The pattern that fathers spend less on daughters is much stronger and statistically significant only among households in which the mother is described as the more altruistic parent. Fathers' lower spending on daughters is nearly three times as large for these households compared to households where the mother is not described as the more altruistic parent. Thus, the spending differences in this sample are driven by differences in altruism toward their daughters.

The investment-based explanation in which mothers anticipate a higher financial return on investments in daughters is not supported.

In the present report prepared for the Institute for Replication (Brodeur et al. 2024), we investigate whether the analytical results of Dizon-Ross and Jayachandran (2023) computationally and recreationally reproducible. Due to missing labels and unclear descriptions of variable transformations opportunities for robustness checks were limited and a direct replication of the study is not possible.

In terms of computational reproducibility, we would like to acknowledge that the original study was successfully reproduced. Due to the well-prepared reproduction data there were no discrepancies in the results of the original paper and our analysis. In terms of recreate reproducibility, we could not create a new code without using the original code as we lacked a code book as well as variable descriptions.

## 2. Computational Reproducibility

We downloaded the replication package from here:

<https://www.openicpsr.org/openicpsr/project/179162/version/V1/view>. The authors provide a package for the computational reproducibility including a readme file with a short description to all the available files, the codes, different raw data files and ado files to install commands in Stata. The provided data is seen in Table 1. By running the master do-file, analysis data files, output tables and figures were created. With the well-prepared package for the reproduction, the same output as presented in the paper was produced. The output tables were automatically created as a TEX-file which we could not open. Nevertheless, the results were consistent and we could successfully computationally reproduce all the main results.

According to the spending on children's goods we re-run the main analysis code for the first table and for the associated robustness checks and sensitivity analysis (Original Appendix Table E.1., E.2. and E.3.). All the results are provided as in the original paper. For the original Table 1, column 2 the authors provide Figure 2 to display the results graphically. They provide additional information on the WTP gaps in percentage terms (calculated in paper statistics) and mapped the WTP to overall demand which was comprehensively described in the original Appendix D. For column 5 and 6 the authors mention that they cannot reject that the coefficients in the both columns are identical (p-



values for equality: 0.16, 0.13, and 0.92 for the Daughter, Mother  $\times$  Daughter, and Mother coefficients, respectively). These results are provided in the paper statistics and are identical to the data reported in the original study.

**Table 1 – Provided data**

|                                 | Fully | Partial | No |
|---------------------------------|-------|---------|----|
| Raw data provided               | x     |         |    |
| Cleaning code provided          | x     |         |    |
| Analysis data provided          | x     |         |    |
| Analysis code provided          | x     |         |    |
| Reproducible from raw data      |       | x       |    |
| Reproducible from analysis data | x     |         |    |

In terms of spending on human capital the results of figures 3 and 4 as well as the underlying regressions (Original Appendix Table A.3) are consistent. WTP gaps in percentage terms are again mentioned in the paper and calculated with the paper statistics code.

Next, the authors test why mothers and fathers spend differently. According to what is mentioned in Figure 4 and Appendix A3 of the original paper our results are completely in line with the original study. The percentage of different patterns of mothers and fathers mentioned in the written text are again calculated in the paper statistics and are in line with our results.

Finally, we could reproduce all the analyses successfully. Concerning the wording we recognized some small issues with the significance level. Once the authors reject the null hypotheses on a 1 percent level (page 453) and in another part they state statistical significance at the 10 percent level (page 458). We would suggest staying consistent with one significance level.

## 2.1 Pre-analysis Plan

The authors did not register a pre-analysis plan.

### 3. Recreate Reproducibility

We attempted to reproduce the main analyses presented in Table 1 by means of writing a new code and using different software. Since the original paper lacked detailed description of the analyses, we could not create the new code independently of the original code for STATA. Analysis of the original code was necessary to see how the authors dealt with the multilevel structure of the data and how they controlled for the effect of different goods. Because the original replication package lacks a code book and most variables in the dataset do not have clear labels, we had to use the original code also to identify specific variables within the large dataset.

During the analyses, we came across several issues that are not mentioned in the original manuscript.

- a. The dataset contained information about the willingness to pay for 15 different goods. Data about 3 adult goods were used to compute the general willingness to pay and these products were excluded from the main analyses as hypothesis' testing focused on child goods only. Data about 2 gender-specific toys were also excluded because they don't provide unbiased information about willingness to pay for son and daughter. According to the original code, authors also excluded data about willingness to pay for a fancy ball. The fancy ball was not mentioned in the original manuscript and it is not clear why the authors excluded this product from their analyses.
- b. Therefore, the analyses presented in Table 1 included data about willingness to pay for 9 different child products. According to the original dataset and code, candy was used as a reference category and 8 dummy variables connected to each product were included in the analyses to control for the effect of a specific product. Nevertheless, according to the STATA outputs, only 6 out of 8 dummy variables were part of the original results. Two products (shoes and tests from the first data collection wave) were automatically excluded by STATA due to the collinearity which is mentioned neither in the manuscript nor in the supplementary materials. Therefore, the analysis did not control for all types of goods.

- c. The paper also does not mention that a control for the children's age group (age 3 to 8 years) is included.
- d. The original analyses controlled for Stratum (see Table 1 in the original manuscript) which was represented by the variable "*strat*" in the dataset. However, the meaning of this variable is not clear neither from the manuscript, nor from the dataset. As the *strat* variable is clustered within the "*schoolcode*" variable, we believe that Stratum might be an indicator for a school class although the number of values (1 through 80) does not correspond to the 40 primary schools surveyed as mentioned in the paper (p. 448).

After identifying details about the original analyses, we used SPSS v29 (IBM Corp., 2023) to prepare the dataset for MPLUS (see Appendix 1) and then we wrote a new code and ran it in MPLUS v8.2 (Muthen & Muthen, 2018). The original STATA code included the "*absorb()*" command, which is specific for STATA and does not have an equivalent in MPLUS. The authors used the "*absorb()*" command to control for the effect of various strata. According to the STATA manual (StataCorp, n.d.), using "*absorb(name of nominal variable)*" command should provide the same results as using dummy variables for each category of a nominal variable. Therefore, we created dummy variables for each category of the "*strat*" variable and included them as predictors in the analyses with "*strat*" = 1 as a reference category. The full syntax for all analyses is available in Appendix 1.

As can be seen in Table 2, we successfully reproduced the analyses presented in columns 1-3 and 5-6 in Table 1 of the original manuscript. All the estimates in the main models are exactly the same as in the original study. There are only negligible differences in the p-values of supplemental tests that tested if the sum of direct and interaction effects (ie., "*Mother + Mother x Daughter*" and "*Daughter + Mother x Daughter*") differed from zero. This difference may be attributed to the difference between both statistical programs and has no practical significance. We also faced a problem with Stratum nr. 56 with only one observation within the dataset (observations connected to Practice tests in the first data collection wave). This Stratum could not be included in the analysis with nonincentivized goods (column 6) as it did not include any observation connected to the nonincentivized good. Moreover, MPLUS was not able to

reliably estimate its parameters in analyses presented in column 1 and column 5 (see Appendix 2).

We did not reproduce the analysis which is presented in the fourth column of Table 1 in the original manuscript. This analysis used the “*absorb()*” command for the household (“*hhid*”) variable. As there were 1084 households, MPLUS was not able to run the analyses with so many predictors.

**Table 2 - Fathers spend less on girls but mothers do not MPLUS reproduction**

|                   | WTP normalized by...                 |                                     |                                      |                                      |                                     |                                      |
|-------------------|--------------------------------------|-------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|--------------------------------------|
|                   | SD                                   | SD                                  | SD                                   | Market price                         | SD                                  | SD                                   |
|                   | (0)                                  | (1)                                 | (2)                                  | (3)                                  | (5)                                 | (6)                                  |
| Daughter          |                                      | -0.037<br>(.024)<br><i>0.114</i>    | -0.102**<br>(0.032)<br><i>0.001</i>  | -0.029**<br>(0.009)<br><i>0.002</i>  | -0.065<br>(0.035)<br><i>0.062</i>   | -0.159**<br>(0.051)<br><i>0.002</i>  |
| Mother X Daughter |                                      |                                     | 0.131**<br>(0.046)<br><i>0.004</i>   | 0.036**<br>(0.013)<br><i>0.006</i>   | 0.069<br>(0.052)<br><i>0.185</i>    | 0.209**<br>(0.074)<br><i>0.005</i>   |
| Mother            |                                      | -0.029<br>(0.028)<br><i>0.286</i>   | -0.095**<br>(0.036)<br><i>0.008</i>  | -0.028**<br>(0.010)<br><i>0.006</i>  | -0.091*<br>(0.039)<br><i>0.021</i>  | -0.090<br>(0.058)<br><i>0.120</i>    |
| Adult WTP         | 0.337***<br>(0.025)<br><i>0.000</i>  | 0.335***<br>(0.032)<br><i>0.000</i> | 0.335***<br>(0.025)<br><i>0.000</i>  | 0.310***<br>(0.022)<br><i>0.000</i>  | 0.506***<br>(0.027)<br><i>0.000</i> | 0.180***<br>(0.036)<br><i>0.000</i>  |
| Adult WTP X Wave  | 0.339***<br>(0.032)<br><i>0.000</i>  | 0.340***<br>(0.032)<br><i>0.000</i> | 0.342***<br>(0.032)<br><i>0.000</i>  | 0.405***<br>(0.031)<br><i>0.000</i>  | 0.249***<br>(0.035)<br><i>0.000</i> | 0.338***<br>(0.050)<br><i>0.000</i>  |
| Wave              | -2.489***<br>(0.078)<br><i>0.000</i> | -<br>(0.078)<br><i>0.000</i>        | -2.493***<br>(0.078)<br><i>0.000</i> | -0.504***<br>(0.021)<br><i>0.000</i> | -0.197**<br>(0.069)<br><i>0.004</i> | -0.382***<br>(0.105)<br><i>0.000</i> |
| Young child       | 2.013***<br>(0.044)<br><i>0.000</i>  | 2.011***<br>(0.044)<br><i>0.000</i> | 2.012***<br>(0.044)<br><i>0.000</i>  | 0.335***<br>(0.012)<br><i>0.000</i>  | -0.095**<br>(0.030)<br><i>0.002</i> | 1.927***<br>(0.052)<br><i>0.000</i>  |
| Deworm w2         | 0.053<br>(0.054)<br><i>0.203</i>     | 0.053<br>(0.041)<br><i>0.204</i>    | 0.053<br>(0.041)<br><i>0.204</i>     | 0.029*<br>(0.12)<br><i>0.014</i>     |                                     | -2.084***<br>(0.063)<br><i>0.000</i> |
| Ball              | 2.348***<br>(0.054)<br><i>0.000</i>  | 2.348***<br>(0.054)<br><i>0.000</i> | 2.352***<br>(0.054)<br><i>0.000</i>  | 0.494***<br>(0.15)<br><i>0.000</i>   | 0.243***<br>(0.035)<br><i>0.000</i> |                                      |
| Shoes w2          | 0.887***                             | 0.887***                            | 0.887***                             | 0.241***                             | 0.887***                            |                                      |

|                           |              |              |              |              |              |              |
|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                           | (0.036)      | (0.036)      | (0.036)      | (0.010)      | (0.036)      |              |
|                           | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> |              |
| Test w2                   | 2.215***     | 2.216***     | 2.129***     | 0.419***     |              |              |
|                           | (0.058)      | (0.058)      | (0.058)      | (0.016)      |              |              |
|                           | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> |              |              |
| Workbook                  | 2.104***     | 2.105***     | 2.108***     | 0.334***     |              |              |
|                           | (0.054)      | (0.054)      | (0.053)      | (0.014)      |              |              |
|                           | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> |              |              |
| Deworm w1                 | 0.091        | 0.092        | 0.092        | 0.020        |              |              |
|                           | (0.048)      | (0.048)      | (0.048)      | (0.015)      |              |              |
|                           | <i>0.057</i> | <i>0.056</i> | <i>0.055</i> | <i>0.161</i> |              |              |
| $R^2$                     | 0.468***     | 0.469***     | 0.469***     | 0.378***     | 0.457***     | 0.493***     |
|                           | (0.012)      | (0.012)      | (0.012)      | (0.013)      | (0.016)      | (0.016)      |
|                           | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> | <i>0.000</i> |
| p-val: Mother + M<br>x D  |              |              | 0.315        | 0.417        | 0.585        | 0.029        |
| p-val: Daughter+<br>M x D |              |              | 0.395        | 0.451        | 0.930        | 0.341        |
| Goods included            | All          | All          | All          | All          | Incentivized | Nonincent.   |
| N                         | 6673         | 6673         | 6673         | 6673         | 4000         | 2673         |

*Note.* *S.E.* are presented in parentheses; p-values in italics; analyses in columns 0-5 included 79 dummy variables corresponding to various strata, analysis in column 6 included 78 dummy variables; all analyses controlled for household as a clustering variable (TYPE = COMPLEX) and used MLR estimation; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

Beyond the analyses presented in the original Table 1, we also performed a regression analysis including only control variables (see column 0). We also present the statistical significance of all predictors and  $R$ -squared for each model. As can be seen from Table 2, the control variables explain a large share of the variance in willingness to pay. Adding the gender of the child, the gender of the parent, and their interaction has only a negligible incremental effect. While these effects are statistically significant, they have only small practical importance. As can be seen from Figures 2 and 3 in the original manuscript or from the standardized coefficients in our analyses (see Appendix 2), parent and child gender are only related to differences in Willingness to pay at the level of a few hundredths of one standard deviation. Such small effects could be found due to the very large sample size and therefore very high test power.

## 4. Robustness Reproduction

Although the replication package was well-prepared for the computational reproduction, there were some crucial parts missing to fully account for robustness checks and replication. Due to incomplete variable labeling and missing variable descriptions it was not possible to come up with appropriate robustness checks. We approached robustness checks of Table 1 and Table A.3 of the original paper though through checking the collinearity of predictors. In the original analysis, seemingly all dummy variables for the 15 goods (the good “candy” was the reference category) entered the regression (Table 1 explicitly states that “all goods” were included in four of the six regressions). We then re-calculated six regression analyses using the 15 single dummy variables instead of using the automated variable command for the list of goods (`_G*`). In fact, in the regressions 1 through 4 only 6 out of the 15 dummy variables were actually calculated and 9 goods omitted due to collinearity issues which the paper did not further discuss. The regressions 5 and 6 included two goods respectively. Table 1 of the original paper indicated “incentivized” (regression 5) and “nonincentivized” (regression 6) goods but it is not clear how many goods this classification actually include. To further inspect collinearity issues, we calculated the variance inflation factors (VIFs) for all predictors. To be able to do so we had to drop the absorb function and run normal regressions. Tables A through E in the Appendix 3 present the results including variance inflation factors (VIFs) and the STATA codes used. The VIFs for the data collection wave variable “Survey Round” (“fu” in the STATA code) in particular ( $VIF > 8$ ) and the age-category variable “Age 3 to 8” (“young” in the STATA code) ( $VIF > 5$ ) are rather high and point towards issues with multicollinearity. Thus, we ran the original regressions reported in the paper again and excluded the data collection wave variable “Survey Round” (“fu” in STATA) and its interaction with adult WTP in an attempt to increase the regression coefficients’ reliability. As Tables A and D in the Appendix 3 show the results differ only slightly from the original results as presented in the original paper’s Table 1 by Dizon-Ross and Jayachandran (2023). Excluding “fu” from the analysis calculates coefficients for 7 dummy variable “goods” and thereby omits 8 instead of the originally 9 omitted variables in regressions 1 through 4. In regression 5

coefficients for 3 goods-dummy variables are estimated instead of the originally 2 coefficients. To conclude, dropping “fu” does affect the results only marginally (while allowing for estimating one additional goods-dummy variable) pointing towards robust effects.

## **5. Direct Reproduction**

Given the very limited description in the paper around data collection and the survey instruments, a direct reproduction of the study is not possible. This largely goes back to a lacking description of the survey questions and their labeling. But also, the sampling is not entirely clear as our problems to identify how the Stratum variable comes about demonstrates.

## **4. Conclusion**

Our investigation of the reproducibility of a study by Dizon-Ross and Jayachandran (2023) shows that the authors have provided a very well documented STATA syntax which enables anybody to reproduce the analysis easily. In addition, the results can also be reproduced with other software packages using different analysis methods demonstrating the robustness of the findings. This being said, the lacking documentation of the data collection instruments and sampling procedures do not allow us to judge the quality of the study data and understand the exact meaning and value of different variables. This also prevented us from developing and conducting sophisticated robustness tests because we could not fully comprehend the measures used and alternative variables provided and hence it was not possible to develop alternative analysis strategies testing the robustness. While, the limited robustness tests that were possible (multicollinearity diagnoses and alternative combinations of controls) point towards robust findings, the general robustness of the study cannot be judged. Our main learning is therefore that a transparent documentation of sampling procedures and data collection instruments is key for several dimensions of reproducibility.

## References

Brodeur, A. et al. (2024). Mass Reproducibility and Replicability: A New Hope. I4R Discussion Paper 107.

Dizon-Ross, R. and Jayachandran, S. (2023). Detecting mother-father differences in spending on children: a new approach using willingness-to-pay elicitation. AER: Insights, 5(4): 445-459.

IBM Corp. (2023). IBM SPSS Statistics for Windows (Version 29.0) [Computer software]. IBM Corp.

Muthen, L. K. and Muthen, B. O. (2018). MPlus (Version 8.2) [Computer software]. Muthen & Muthen. <https://www.statmodel.com/index.shtml>

StataCorp. (n.d.). Stata Statistical Software: areg — Linear regression with a large dummy-variable set. College Station, TX: StataCorp LLC. <https://www.stata.com/manuals13/rareg.pdf>



## Appendix

### Appendix 1. Recreate reproducibility code

#### I. Prepare the dataset for MPLUS in SPSS v29

DATASET ACTIVATE DataSet1.

*\*Start with the processed dataset "analysis\_data.dat" that includes final variables for the analyses.*

*\*As Good1 (Candy) was used as a reference good, there is no variable connected to Good1. Compute Good\_1 to identify which data are connected to it: .*

```
COMPUTE @_Ggood1=1-
@_Ggood_2+@_Ggood_3+@_Ggood_4+@_Ggood_5+@_Ggood_6+@_Ggood_7+@_Ggood_8+@_Ggood_9+@_Ggood_10+@_Ggood_11+@_Ggood_12+@_Ggood_13+@_Ggood_14+@_Ggood_15+@_Ggood_16).
```

EXECUTE.

*\*Add a filter variable for child goods, that were used in the analyses for Table 1 (1 = goods used in the analyses, 0 = adult goods and excluded goods).*

```
COMPUTE
Childgood_incl=@_Ggood_1+@_Ggood_2+@_Ggood_5+@_Ggood_7+@_Ggood_8+@_Ggood_10+@_Ggood_12+@_Ggood_13+@_Ggood_15.
```

EXECUTE.

*\*Split the file to include only information about the 9 child goods used in the analyses for Table 1.*

```
SPSSINC SPLIT DATASET SPLITVAR=childgood_included
```

```
/OUTPUT DIRECTORY= "C:\Split data" DELETECONTENTS=NO
```

```
/OPTIONS NAMES=VALUES NAMEPREFIX="Childgood" PRINTLIST=YES.
```

EXECUTE.

*\*Continue with the "childgood\_1.sav" dataset that include information about 9 child goods (N = 6 697 observations).*

*\*Delete and reorder variables to keep variables relevant to the analyses only.*

```
ADD FILES file *
```

```
/KEEP hhid strat wtp_std girl mom adult_wtp_std adult_wtp_stdXfu fu young momXgirl wtp adult_wtp_pct adult_wtp_pctXfu hypo @_Ggood_2 to @_Ggood_16.
```

EXECUTE.

*\*Create dummies for the "strat" variable to enable to imitate the "absorb" command i MPLUS ("strat" must be a NOMINAL variable).*

DATASET ACTIVATE DataSet2.

```

SPSSINC CREATE DUMMIES VARIABLE=strat
ROOTNAME1=Str
/OPTIONS ORDER=A USEVALUELABELS=NO USEML=YES OMITFIRST=NO.
EXECUTE.

*Recode missing values to -999.

RECODE hhid strat wtp_std girl mom adult_wtp_std adult_wtp_stdXfu fu young momXgirl wtp
    adult_wtp_pct adult_wtp_pctXfu hypo @_Ggood_2 @_Ggood_3 @_Ggood_4 @_Ggood_5
    @_Ggood_6 @_Ggood_7 @_Ggood_8 @_Ggood_9 @_Ggood_10 @_Ggood_11 @_Ggood_12
    @_Ggood_13 @_Ggood_14 @_Ggood_15 @_Ggood_16 Str_1 Str_2 Str_3 Str_4 Str_5 Str_6 Str_7
    Str_8 Str_9 Str_10 Str_11 Str_12 Str_13 Str_14 Str_15 Str_16 Str_17 Str_18 Str_19 Str_20 Str_21
    Str_22 Str_23 Str_24 Str_25 Str_26 Str_27 Str_28 Str_29 Str_30 Str_31 Str_32 Str_33 Str_34 Str_35
    Str_36 Str_37 Str_38 Str_39 Str_40 Str_41 Str_42 Str_43 Str_44 Str_45 Str_46 Str_47 Str_48 Str_49
    Str_50 Str_51 Str_52 Str_53 Str_54 Str_55 Str_56 Str_57 Str_58 Str_59 Str_60 Str_61 Str_62 Str_63
    Str_64 Str_65 Str_66 Str_67 Str_68 Str_69 Str_70 Str_71 Str_72 Str_73 Str_74 Str_75 Str_76 Str_77
    Str_78 Str_79 Str_80 (MISSING=-999).

EXECUTE.

*Prepare a .dat file for MPLUS. In case of having commas for decimals, change them to dots manually.

SAVE TRANSLATE OUTFILE='C:\mother_father_differences_data.dat'

/TYPE=TAB

/ENCODING='UTF8'

/MAP

/REPLACE

/CELLS=VALUES.

EXECUTE.

```

## II. Code for the reproduction of Table 1 in MPLUS v8.2

### a) Column 1

```

DATA: FILE IS "mother_father_differences_data.dat";

VARIABLE: NAMES ARE

hhid strat wtp_std girl mom ad_wtpst adwtpsXf fu young

momXgirl wtp ad_wtppc adwtpXf hypo

G2dewor G3fball G4jerry G5ball G6poster G7shoes G8test G9toy G10wbook

G11cup G12dewor G13shoes G14sieve G15test G16toy

strat1-strat80

;

```

USEVARIABLES ARE

hhid wtp\_std girl mom ad\_wtpst adwtpsXf fu young  
G2dewor G5ball G7shoes G8test G10wbook G12dewor  
strat2-strat80

;

CLUSTER IS hhid;

MISSING ARE ALL (-999);

ANALYSIS: TYPE = COMPLEX;

MODEL: wtp\_std ON girl mom ad\_wtpst adwtpsXf fu young  
G2dewor G5ball G7shoes G8test G10wbook G12dewor  
strat2-strat80

;

OUTPUT: SAMPSTAT STANDARDIZED CINTERVAL;

#### **b) Column 2**

DATA: FILE IS "mother\_father\_differences\_data.dat";

VARIABLE: NAMES ARE

hhid strat wtp\_std girl mom ad\_wtpst adwtpsXf fu young  
momXgirl wtp ad\_wtppc adwtpXf hypo  
G2dewor G3fball G4jerry G5ball G6poster G7shoes G8test G9toy G10wbook  
G11cup G12dewor G13shoes G14sieve G15test G16toy  
strat1-strat80

;

USEVARIABLES ARE

hhid wtp\_std girl momXgirl mom ad\_wtpst adwtpsXf fu young  
G2dewor G5ball G7shoes G8test G10wbook G12dewor  
strat2-strat80

;

CLUSTER IS hhid;

MISSING ARE ALL (-999);

ANALYSIS: TYPE = COMPLEX;

MODEL: wtp\_std ON

```

girl (pp1)
momXgirl (pp2)
mom (pp3)
ad_wtpst adwtpsXf fu young
G2dewor G5ball G7shoes G8test G10wbook G12dewor
strat2-strat80
;
MODEL CONSTRAINT:
    NEW(p1 p2);
    p1 = pp1 + pp2;
    p2 = pp3 + pp2;
MODEL TEST:
    p1 = 0;
    p2 = 0;
OUTPUT: TECH1 SAMPSTAT STANDARDIZED CINTERVAL;

```

### c) Column 3

```

DATA: FILE IS "mother_father_differences_data.dat";
VARIABLE: NAMES ARE
hhid strat wtp_std girl mom ad_wtpst adwtpsXf fu young
momXgirl wtp ad_wtppc adwtpXf hypo
G2dewor G3fball G4jerry G5ball G6poster G7shoes G8test G9toy G10wbook
G11cup G12dewor G13shoes G14sieve G15test G16toy
strat1-strat80
;
USEVARIABLES ARE
hhid wtp girl momXgirl mom fu young ad_wtppc adwtpXf
G2dewor G5ball G7shoes G8test G10wbook G12dewor
strat2-strat80
;
CLUSTER IS hhid;
MISSING ARE ALL (-999);

```

```

ANALYSIS: TYPE = COMPLEX;
MODEL: wtp ON
girl (pp1)
momXgirl (pp2)
mom (pp3)
ad_wtppc adwtpXf fu young
G2dewor G5ball G7shoes G8test G10wbook G12dewor
strat2-strat80
;
MODEL CONSTRAINT:
    NEW(p1 p2);
    p1 = pp1 + pp2;
    p2 = pp3 + pp2;
MODEL TEST:
    p1 = 0;
    p2 = 0;
OUTPUT: SAMPSTAT STANDARDIZED CINTERVAL;

```

#### **d) Column 5**

```

DATA: FILE IS "mother_father_differences_data.dat";
VARIABLE: NAMES ARE
hhid strat wtp_std girl mom ad_wtpst adwtpsXf fu young
momXgirl wtp ad_wtppc adwtpXf hypo
G2dewor G3fball G4jerry G5ball G6poster G7shoes G8test G9toy G10wbook
G11cup G12dewor G13shoes G14sieve G15test G16toy
strat1-strat80
;
USEVARIABLES ARE
hhid wtp_std girl momXgirl mom ad_wtpst adwtpsXf fu young
!G2dewor
G5ball G7shoes
!G8test G10wbook G12dewor

```

```

strat2-strat80
;
SUBPOPULATION IS (hypo EQ 0);
CLUSTER IS hhid;
MISSING ARE ALL (-999);
ANALYSIS: TYPE = COMPLEX;
MODEL: wtp_std ON
girl (pp1)
momXgirl (pp2)
mom (pp3)
ad_wtpst adwtpsXf fu young
!G2dewor
G5ball G7shoes
!G8test G10wbook G12dewor
strat2-strat80
;
MODEL CONSTRAINT:
    NEW(p1 p2);
    p1 = pp1 + pp2;
    p2 = pp3 + pp2;
MODEL TEST:
    p1 = 0;
    p2 = 0;
OUTPUT: SAMPSTAT STANDARDIZED CINTERVAL;

```

**e) Column 6**

```

DATA: FILE IS "mother_father_differences_data.dat";
VARIABLE: NAMES ARE
hhid strat wtp_std girl mom ad_wtpst adwtpsXf fu young
momXgirl wtp ad_wtppc adwtpXf hypo
G2dewor G3fball G4jerry G5ball G6poster G7shoes G8test G9toy G10wbook
G11cup G12dewor G13shoes G14sieve G15test G16toy

```

```
strat1-strat80
;
USEVARIABLES ARE
hhid wtp_std girl momXgirl mom ad_wtpst adwtpsXf fu young
G2dewor
!G5ball G7shoes
!G8test G10wbook G12dewor
strat2-strat55 strat57-strat80
;
SUBPOPULATION IS (hypo EQ 1);
CLUSTER IS hhid;
MISSING ARE ALL (-999);
ANALYSIS: TYPE = COMPLEX;
MODEL: wtp_std ON
girl (pp1)
momXgirl (pp2)
mom (pp3)
ad_wtpst adwtpsXf fu young
G2dewor
!G5ball G7shoes
!G8test G10wbook G12dewor
strat2-strat55 strat57-strat80
;
MODEL CONSTRAINT:
    NEW(p1 p2);
    p1 = pp1 + pp2;
    p2 = pp3 + pp2;
MODEL TEST:
    p1 = 0;
    p2 = 0;
OUTPUT: SAMPSTAT STANDARDIZED CINTERVAL;
```