Navarra Center for International Development

Working Paper nº 04/2023

How Labels and Vouchers Shape Unconditional Cash Transfers? Experimental Evidence from Georgia

Miguel Ángel Borrella-Mas

University of Navarra

Jaime Millán.Quijano

University of Navarra

Anastasia Terskaya

University of Navarra

How do Labels and Vouchers Shape Unconditional Cash Transfers? Experimental Evidence from Georgia.*

Miguel Ángel Borrella-Mas † Jaime Millán-Quijano ‡

Anastasia Terskaya§

November 30, 2023

Abstract

We implemented a randomized control trial in Georgia to study how labels and food vouchers affect household expenditure among low-income recipients of unconditional cash transfers. Households were randomly assigned to receive only an unconditional cash transfer, a label indicating an amount intended for children's expenses in addition to the transfer, or a portion of the transfer as a food voucher usable exclusively at designated stores. We find that labeling increases the share of expenditure on children. Meanwhile, food vouchers reduce total consumption, this being likely due to the increased cost associated with shopping at voucher-accepting shops.

JEL classification: D04, I24, I38, O12

Keywords: Cash Transfers, Labeling Effect, Food Vouchers, Randomized Control Trial.

---

*We are grateful to the evaluation team at Econometría S.A., UNICEF Georgia, GeoStat and the Social Services Agency of Georgia, with whom we implemented the experiment and collected the data. We also thank participants at ESPE 2023, SEHO 2023, CEA 2023, Ridge Public Economics Workshop 2023, and seminar participants at the University of Austin and the University of Navarra. The authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness Grant PID2020-120589RA-I00. This RCT was registered as AEARCTR-0006611 (Baum,et al. 2020). IRB by Econometría Consultores S.A. 2019-08-30 No 8.

†Department of Economics and Navarra Center for International Development (NCID), University of Navarra. mborrella@unav.es.

‡Corresponding author. NCID, University of Navarra, and Centro de Estudios Manuel Ramirez (CEMR). jmillanq@gmail.com.

§Department of Economics and Institut d'Economia de Barcelona (IEB), University of Barcelona. a.terskaya@ub.edu.

1 Introduction

As nations progress and economies grow, poverty and economic inequality remain among the most pressing problems, with more than 700 million people living in extreme poverty. Furthermore, children are impacted to a greater extent. Based on estimates from the World Bank and UNICEF, 17.5 percent of children, totaling 356 million, live on less than $1.9 (PPP) a day (Silwal et al., 2020). The disadvantages experienced during childhood have profound and lasting effects, as they are associated with higher rates of morbidity and mortality (Power et al., 2013), poorer psychological well-being (Evans, 2016), as well as lower levels of academic achievement (Dahl and Lochner, 2012), among other outcomes. The far-reaching implications of poverty underscore the pressing need for the optimal design and effective implementation of public policies aimed at alleviating poverty, with a particular focus on the well-being of children.

Social assistance in the form of cash transfers or vouchers to poor households is one of the most prevalent policies for alleviating poverty.[1] The results of numerous evaluations of cash transfer policies suggest that, in most cases, cash transfers were effective in increasing expenditure and reducing poverty (Bastagli et al., 2019). However, the evidence for second-order outcomes, such as children's education and health, presents a mixed picture, with estimates varying depending on program characteristics (Baird et al., 2014; García and Saavedra, 2017; Bastagli et al., 2019; Pega et al., 2022; Haushofer et al., 2023; Hawkins et al., 2023). The lack of conclusive evidence regarding the positive impacts of cash transfers on second-order outcomes could indicate that families do not allocate their resources in a socially optimal way, potentially due to market failures. For example, parents may not allocate a significant enough portion of the transfers to their children, who generate greater returns on government investments than adults because of lack of information about the true returns on investments in children (Hendren and Sprung-Keyser, 2020). In such cases,

[1]Cash transfers were the key social protection measure enacted in response to the COVID-19 pandemic (Gentilini, 2022). Throughout this period, cash transfer programs were implemented in 203 countries and constituted about 25 percent of the overall social protection responses (Gentilini et al., 2022).

formal conditionalities (e.g., requiring school attendance), nudges or labels, or transfers specifically designated for certain purposes (e.g., food vouchers) may be more successful in achieving desired outcomes for social assistance policies.

Within this context, the primary objective of this paper is to compare the effects of different forms of unconditional transfers on household expenditure patterns. In particular, we focus on two low-cost adjustments of unconditional cash transfers: labeling and food vouchers.[2] Our analysis is built upon an experiment carried out in Georgia within a large nationwide Targeted Social Assistance (TSA) program. The program provides monthly unconditional cash transfers (UCT) to economically vulnerable households, covering about 12 percent of the country's population —or about half a million people. The TSA program comprises two components: a general benefit extended to all family members (including both children and adults) and a specific child benefit. Initially, household heads received the entire benefit amount all at once, without differentiating the portion designated as child benefit. In 2019, we conducted a randomized controlled trial (RCT) wherein households in some municipalities received a message (SMS) indicating the amount of the child benefit and its intended use for children, households in some municipalities received a portion of the child benefit in the form of a food voucher, and households in other municipalities received both vouchers and messages. Meanwhile, the procedure for households in the last set of municipalities retained the status quo, continuing to receive UCT as they had done prior to the experiment. Thus, the random allocation of the different arms of the program enables us to causally estimate *i*) the impact of labeling, *ii*) the comparison between the effects of vouchers and cash, and *iii*) the evaluation of the combined impact of labeling and vouchers. We do so by comparing expenditure patterns—–measured approximately eight months after the start of the experiment—–between municipalities assigned to the different arms of the program.

Our primary findings can be divided into two main categories. First, we

[2]By contrast, imposing a formal conditionality can substantially increase a program's cost due to the high costs of monitoring. Additionally, it might exclude specific vulnerable population groups that do not comply with program rules (Baird et al., 2011).

observe that labels lead to an increase in the proportion of total expenditure allocated to children, children's education, and children's clothing. Specifically, the labeling effect raises the share of expenditure on children by an average of 3.4 percentage points, from the baseline of 7.3 percent in municipalities with unconditional cash transfers. This increase in the portion of child-related expenditure is linked to a decrease in the share of expenditure on food. The labeling effect is not statistically distinguishable from zero for total household expenditure and the share of expenditure on adults.

Our in-depth analysis of heterogeneity based on Generic Machine Learning (Chernozhukov et al., 2018) indicates that the labeling effect on child-related expenditure is heterogeneous. The effect is positive and statistically distinguishable from zero for approximately 50 percent of households while it does not reach statistical significance for the remaining households. Notably, the labeling effect is stronger for households that received a larger child benefit and, therefore, more salient messages. This result suggests that labels may induce households to mentally segregate their resources according to the intended purpose of the funds, consistent with mental accounting (Thaler, 1999). We also find that households where the benefit recipient was a woman, and households with lower socioeconomic status are more responsive to messages.

Secondly, our comparative analysis between recipients of food vouchers and cash transfers reveals that vouchers are associated with a significant reduction in total household expenditure. However, the proportion of expenditure across various categories on different items and total household income remain comparable to households receiving only cash. Importantly, we do not find any evidence that the vouchers increase the share of expenditure on food. We posit that the most likely explanation for the adverse impact of vouchers on total expenditure lies in the elevated shopping costs associated with their restricted acceptance at specific stores, which may not align with the preferred choices of households. We further demonstrate that the shops accepting vouchers are located in less convenient places compared to the regular stores typically frequented by households. As a result, this increase in shopping costs could discourage households from using the voucher. Given that vouchers cannot be traded, and the funds within them cannot be saved or withdrawn, households may simply opt not to use them. Our heterogeneity analysis lends support to this explanation, revealing that the households most negatively affected, on average, reside closer to their regular shopping destinations.

The analysis of heterogeneity also suggests that approximately only 25 percent of households reduce their expenditure because of the voucher, whereas the remaining households do not exhibit an effect that is statistically distinguishable from zero. We provide evidence that the most negatively affected households are, on average, smaller in size, have a higher socioeconomic status, and receive smaller voucher benefits. This suggests that, for these type households, the increased shopping costs do not offset the advantages of using the voucher.

Finally, receiving the voucher and the message reduces the labeling increase on child-related expenditure and reduces the negative effect of the voucher on total household expenditure. This result aligns with the notion that some households who wanted to spend more on children were dissuaded by the higher shopping cost of the voucher. Also, some households who were faced with higher cost of shopping associated with the voucher were keen to overcome that cost in response to the nudge.

Our research makes a contribution to the extensive body of literature emphasizing the critical role of specific implementation details of cash transfer programs. For instance, Baird et al. (2011) investigate how formal conditionality shapes the impact of cash transfers on educational outcomes. Similarly, Hidrobo et al. (2014) conducts a comparative analysis, examining the effects of food assistance provided in various forms, including cash, food vouchers, and food transfers, on patterns of food consumption. A recent study by Orkin et al. (2023) based on an RCT in Kenya, demonstrates that a workshop teaching techniques to raise aspirations and plan for their achievement, when combined with UCT yields effects similar to those with only cash. We contribute to this literature by evaluating the effect of behavioral intervention (labels) and delivery method (voucher *versus* cash) within a large-scale UCT program.

Our paper also adds to the growing body of research exploring how nudges or labels can influence the economic decisions made by households.[3] While standard economic theory predicts that labels on cash transfers should not have an impact, given that money is typically regarded as fungible, several empirical and theoretical studies have suggested the opposite, i.e., that individuals do modify their consumption patterns in response to labels and do not always treat money as wholly fungible (Thaler and Sunstein, 2008; Beatty et al., 2014; Abeler and Marklein, 2017). However, there is limited empirical evidence regarding the effect of labels in cash transfers on outcomes related to children. The most closely related study to ours is the work by Benhassine et al. (2015), who compares the effects of small labeled cash transfers to conditional cash transfers in Morocco on educational outcomes. This study reveals significant improvements in school participation attributed to both types of transfers. Our research complements this study by comparing UCT to labeled cash transfers, allowing us to identify the impact of labels on child expenses in the absence of formal conditions.

Other related studies compare the effect of child benefits with the effect of income from other sources, interpreting the difference as the labeling effect on child-related expenditure.[4] For example, Kooreman (2000) concludes that child benefit increases expenditure on children more than income from other sources, while having no effect on expenditure on adults. By contrast, Blow et al. (2012) finds that the child benefit is disproportionately allocated to adult-oriented goods; meanwhile Edmonds (2002) fails to reject the hypothesis that the child benefit is allocated differently than income from alternate sources. However, the main limitation of all these studies stems from the fact that they are based on the comparison between the effects of labeled child benefit and income from other sources, such as earnings. Given that earned and unearned

[3]See Hummel and Maedche (2019); Damgaard and Nielsen (2018) for an extensive literature review.

[4]We also contribute to a broader literature on the effect of labeling on other outcomes. For example, Beatty et al. (2014) showed that the UK winter fuel payment increases household spending on fuel significantly more than what the estimated Engel curve predicts, providing evidence of a labeling effect.

income may have different effects on household expenditure, it remains unclear whether the observed effects are exclusively driven by labeling. Additionally, the estimated effects of labels might be contaminated by other characteristics of the benefit (e.g., the periodicity of payment) or intra-household bargaining, particularly considering that child benefits are more frequently received by mothers. To the best of our knowledge, no existing study has undertaken a comparison of the effects of unconditional and labeled cash transfers on child-related expenditure using random assignment of labels.

Finally, our research contributes to the body of literature focused on comparing the impact of UCT with vouchers. Our analysis is most closely related to Hidrobo et al. (2014), who conducts a randomized evaluation comparing the effects of UCT to food vouchers and food transfers. In line with our findings, they do not find a significant difference in the impact of cash or vouchers on the quantity of food consumption. Other relevant studies in this domain include Hoynes and Schanzenbach (2009) and Hastings and Shapiro (2018), who investigate the effects of the Food Stamp and Supplemental Nutrition Assistance (SNAP) programs, respectively. Interestingly, these two papers reach contradictory conclusions. While Hoynes and Schanzenbach (2009) find that households respond similarly to one dollar in cash income and one dollar in food stamps, Hastings and Shapiro (2018) find that the marginal propensity to consume SNAP-eligible food using SNAP benefits is significantly higher compared to using cash, challenging the notion that households treat money as entirely fungible.

The remainder of the paper proceeds as follows. Section 2 describes the institutional background and the experimental design, while Section 3 describes the data and the sample we use in our analysis. Section 4 describes our identification strategy and provides evidence in support of our key identifying assumption and Section 5 reports the results and discusses the main mechanisms. Finally, Section 6 concludes.

2 Institutional Background

We examine Georgia's nationwide Targeted Social Assistance (TSA) program.[5] The TSA was introduced in 2008 as a response to the economic crisis resulting from the conflict with the Russian Federation and the effects of the international financial crisis (Baum et al., 2016a; World Bank, 2018). The program involves providing monthly cash transfers to vulnerable households. The TSA covers 12 percent of Georgia's population, with a total yearly cost of approximately GEL 270 million, accounting for 9 percent of Georgia's social protection budget. Notably, it stands as the largest social protection program in terms of both coverage and costs, second only to retirement pensions (World Bank, 2018).

The program has undergone several modifications over time. Since 2015, the amount of the transfer is determined by the household composition and the resulting score of a proxy means-testing procedure, PMT score hereafter (Baum et al., 2016b). The PMT involves gathering detailed information through interviews, which includes income, assets, utility bills, and the special needs of each household. The interview process started in April 2015, initially targeting households identified as vulnerable. However, households not initially approached by the government were afforded the opportunity to request a PMT assessment. In addition, changes in a household's composition or wealth could lead to additional assessments, either by local/national authorities or at the household's request as well. In all cases, the monthly transfer amount ranges from GEL 30 to GEL 60 per household member, depending on the PMT score, provided it does not exceed 65,000. Additionally, households with a PMT score of 100,000 or less receive an extra transfer of GEL 50 per child under the age of 16.[6] A comprehensive overview of the benefit scheme is described in Table D.1 in Appendix D.

The TSA transfers constitute a significant portion of households' total income. Drawing from the data used to calculate the PMT scores, households

[5]Georgia is classified as an upper-middle-income country. As of 2020, the GDP per capita was \$4,300 (\$14,500 in PPP terms), and 21 percent of the population lived below the national poverty line.

[6]The amount of child benefit was increased from GEL 10 to GEL 50 in January 2019.

with PMT scores below 100,000 had an average income of GEL 340 per month before accounting for the TSA. To further illustrate, consider a household with a PMT score of 25,000 and 2 children under 16 and 2 adults. This household is eligible to receive GEL 60 × 4 + GEL 50 × 2 = GEL 340. By contrast, a household with a similar composition but with a PMT score of 62,000 will receive GEL 30 × 4 + GEL 50 × 2 = GEL 220.

TSA beneficiaries receive their benefits through bank transfers on the last working day of each month. Each household head is provided with a debit card, which they can use to make purchases in stores or withdraw money from the bank.[7] It is worth noting that the appointment of household heads is determined by the households themselves, and it may not necessarily correspond to the member with the highest income. As a result, the household head is the father in 64.5 percent of households, while in the remaining households, this position is held by the mother. Importantly, all households receive a message when the benefit is transferred, which also includes the total amount of the transfer.

2.1 The experiment

Since 2009, UNICEF has actively participated in enhancing the implementation of the program with the objective of making it more child-sensitive, improving monitoring, and providing technical advice to the government (Baum et al., 2016b,a). As part of this collaboration, in January 2019, the Georgian government introduced a proposal to allocate 60 percent of the child benefit in the form of food vouchers, with the intention of stimulating expenditures directed towards children's nutritional needs. Additionally, the government proposed sending a message indicating the amount of the transfer that should be allocated to children. Subsequently, we conducted a randomized controlled trial (RCT) in February 2019 —in collaboration with Econometría S.A., UNICEF, and the Social Services Agency (SSA)— to evaluate the impact of each of these pro-

[7] All beneficiaries are provided with a bank account at Liberty Bank, a nationwide institution.

posed program modifications. More precisely, municipalities were randomly assigned to the following groups:[8]

Arm 1 (cash only): In these municipalities, the total amount of the transfer (including child benefit and per individual benefit) was received in cash as a single payment.

Arm 2 (voucher): The households with children received their child benefit as GEL 20 per child in cash and GEL 30 per child on a food voucher. The remaining part of the transfer was received in cash. The food vouchers could only be used to purchase food in specific stores.

Arm 3 (cash + label): In these municipalities, the total amount of the transfer was received in cash. Households with children also received a message reminding them that GEL 50 per child should be spent on children. Specifically, they receive the following SMS every month when they receive their benefit deposit:
"Your family received social assistance benefit GEL [Total amount], which envisages GEL [Child benefit amount] for each family member under 16."

Arm 4 (voucher + label): The same as Arm 3 with the following SMS every month when they get their benefit deposit:
"Your family received social assistance benefit GEL [Total amount] from which GEL [Child benefit amount] is transferred to "child food voucher" card for each family member under 16".

In this context, Arm 1 represents the baseline (or control) group, as it mirrors the existing practices of TSA transfers before the beginning of the experiment. Arm 2 represents the case of a partial voucher, as only a portion of the transfer was restricted for purchasing food. It is worth noting that within the baseline group (Arm 1), 87 percent of households spent more on food than the value of

[8]The randomization was done at the municipality level (and not at the household level) to avoid potential conflicts between neighbors and to facilitate the voucher implementation by Liberty Bank and the grocery shops that could accept the voucher.

the food voucher they would have received had they been placed in the voucher group (see Figure D.1.a in Appendix D). According to economic theory, cash and infra-marginal food vouchers of equal value are expected to have the same impact on food consumption (Southworth, 1945). Consequently, the comparison of expenditure patterns between Arms 1 and 2 allows us to test this theoretical prediction.

It is important to point out that the food voucher was provided in the form of an additional debit card, valid exclusively at selected shops across the country. In parallel with the main TSA transfer, any unspent funds from the food voucher would be reverted to the Social Services Agency. Nevertheless, unlike the main TSA debit card, households were not afforded the option to withdraw money from the voucher card.

Arm 3 represents a labeled cash transfer, as the SMS was intended to clarify how much was earmarked as a benefit for children. However, it is important to underline that the compliance with SMS indications was neither enforced nor subjected to monitoring. Thus, this constitutes a nudging intervention, as it does not change households economic incentives (Thaler and Sunstein, 2008). Finally, Arm 4 is designed to explore the potential complementarities between labels and vouchers.

3 Data and Estimation Sample

In this section, we discuss the dataset employed in our analysis, along with a detailed description of the estimation sample used to assess our research questions.

Outcomes information was collected through a comprehensive survey targeting households with children during the period between November and December 2019. This survey was carried out approximately 8 months after the beginning of the experiment in March 2019, and it was administered by the national statistics agency —GeoStat. The selection of households was based on the PMT score from the first interview, with a random sub-sample of households near each PMT cutoff being interviewed.[9] The survey questions were directed towards the heads of households.

Additionally, the administrative data from the first PMT interview that each household carried out provides a large set of baseline characteristics. As explained above, this data compilation spanned the years between 2015 and 2019.[10]

As a result, we have information from 6,874 households with children with a valid PMT score in 46 municipalities. To this initial sample, we apply several filters. First, we exclude 1,059 households from Tbilisi since this city was not included in the random assignment.[11] Second, we eliminate 945 households with the first PMT score above 100,000 since they were not eligible to receive the benefit. Third, we discard 235 households who applied to the TSA after the experiment began. The reason behind this exclusion stems from the fact that baseline characteristics are measured after the allocation of arms, and thus the decision to apply may have been already affected by this allocation. Finally, 173 households with incomes and/or expenditures below the 1st percentile or above the 99th percentile of the income or expenditure distribution are removed. After applying all these filters, the resulting estimation sample consists of 4,462 households.

The main outcomes for our analysis are total household income, total household expenditure, and the shares of expenditure on food, adults, children, childcare, children's clothing, and children's education. To account for potential confounding factors, we incorporate a battery of control variables into our model. These include the PMT score obtained by each household during the first interview, dummy variables to identify households below and above each PMT cutoff group, household composition, household head characteristics (age, sex, educational level), household characteristics (pregnancy of the mother, single-

[9]Precisely, we selected households whose first PMT scores were around 30,000 and 100,000 using the optimal bandwidth (Calonico et al., 2014). For other cutoff groups, we used the distance to the nearest cutoff as the selection criterion (e.g., for the 57,000 cutoff, we employed a bandwidth of 3,000, considering the next cutoff at 60,000).

[10]See Econometría (2020) for more detailed information about the sample selection process.

[11]Tbilisi is the most populated city in Georgia, with a population density about 40 times larger than the country's average. The reason for this exclusion is that Tbilisi's higher population density could create an imbalance across the different experimental groups.

mother household, presence of household members with a disability, income, presence of wage earners in the household), and house characteristics (number of rooms, floor material). All these variables were measured during the first interview, when each household obtained the initial PMT score, preceding the implementation of the RCT. Additionally, we control for the household's total monthly TSA expected transfer based on their first PMT score. More detailed information about all these variables can be found in Appendix A.

Table D.2 in the Appendix provides summary statistics for the outcome measures. On average, the monthly household income and monthly household expenditure amount to GEL 629.9 and GEL 415.6, respectively. Additionally, households report spending an average of 9.5 percent of their expenditure on children, 6.0 percent on adults, and 34.7 percent on food. In addition, Table D.3 within Appendix D presents descriptive statistics for the baseline characteristics. The average total monthly household transfer amounts to GEL 207.9 (equivalent to USD 72.0), constituting approximately the 35 percent of the average monthly household income. Within our sample, families have an average of 3.2 adults and 1.9 children.

4 Empirical Strategy

In this section we outline the empirical strategy employed to estimate the effect of each of the proposed program modifications on several outcomes. The random assignment of municipalities to different arms implies that, on average, municipalities in various treatment groups have comparable background characteristics. Consequently, different treatment groups are likely to have comparable expenditure distributions in the absence of program modifications. By comparing outcomes between the randomly assigned treatment groups, we can estimate the relative effect of the voucher *versus* cash, as well as the labeling effect.

We use the following specification to estimate the effects of Arms 2-4 relative to Arm 1 (cash only).

$$Y_{mi} = \rho_0 + \rho_1 Arm_{2,m} + \rho_2 Arm_{3,m} + \rho_3 Arm_{4,m} + X^{'}_{mi}\delta + \epsilon_{mi} \tag{1}$$

where Y_{mi} denotes the outcome of household i from municipality m. $Arm_{2,m}$ is a dummy variable that equals 1 if municipality m was assigned to receive part of the child benefit in the form of a food voucher. Similarly, $Arm_{3,m}$ is a dummy variable that equals 1 if municipality m was assigned to receive an SMS informing that GEL 50 per child should be spent on children (labeled cash transfer). Additionally, $Arm_{4,m}$ is a dummy variable that equals 1 if municipality m was assigned to receive both the food voucher and the label. X_{mi} denotes the set of baseline household characteristics reported in Table D.3.

Within this model, the parameter ρ_1 captures the differential effect of the food voucher compared to cash of a similar value. Similarly, the parameter ρ_2 denotes the effect of the label, as it compares the outcomes between recipients of cash with and without the SMS, essentially measuring the effect of labeling the transfer. Finally, the parameter ρ_3 stands for the differential effect of the food voucher with the label compared to cash of a similar value.

4.1 Inference

We estimate the effects of three program arms on multiple outcomes, which means there is a possibility of finding statistically significant effects by chance alone. To address this concern and account for the multiple statistical tests conducted, we calculate the false discovery rate (FDR)-adjusted q-values following the two-stage procedure proposed by Benjamini et al. (2006) and the algorithm provided in Anderson (2008). In short, the process guarantees that the FDR, which represents the proportion of incorrect rejections, remains below a certain threshold denoted as q. The decision rule we use to determine the significance of each of the tested hypotheses guarantees that at least 90 percent of the significant test results correctly reject the null hypothesis ($q \leq 0.1$, as in Efron 2012). We report FDR-adjusted q-values when presenting the results alongside analytical standard errors clustered at the municipality level. Appendix C.1 contains a detailed description of our procedures for multiple-hypothesis adjustment.

Additionally, we enhance conventional inference, which is based on analytical standard errors, by incorporating resampling-based inference, as suggested by Young (2019). A detailed description of how we generate the resampling-based p-values is available in Appendix C.2. In Table C.1 of the Appendix, we present both the resampling-based p-values and the conventional p-values. In most cases, these two sets of p-values show similar magnitudes.

4.2 Validity of the Identification Strategy

Randomization implies that there are no statistically significant differences in baseline characteristics between municipalities assigned to different program modifications. Table D.3 in Appendix D presents summary statistics of the baseline characteristics for each program arm. In line with the random assignment, the average values of baseline characteristics are comparable across the experimental arms. Among the 21 characteristics considered, only one shows a statistically significant difference at the 5 percent level, the indicator that the PMT score falls between 30K and 39K. To further scrutinize the balance across arms, we also regress each arm indicator variable on the complete set of covariates and then we conduct F-tests to assess the joint significance of these covariates. In all cases, the F-tests are small and statistically insignificant at the 5 percent level.

Lastly, we include the complete set of control variables in our main regressions to safeguard against any potential influence stemming from the minor differences observed in the baseline characteristics on our primary findings. Additionally, we present the results without including any covariates in Tables D.4 and D.5 in Appendix D. These results exhibit a remarkable similarity to those obtained when covariates are included, providing reassurance that our results are not driven by the observed differences in a few characteristics among the program arms.

4.3 Heterogeneous effects

We complement our main results by examining the heterogeneous effects of vouchers and labels to determine which groups benefit the most (or least) from each treatment. Such insights can offer valuable guidance to policy-makers on how to strategically target vouchers and labels. Additionally, analyzing the heterogeneity of the impact may shed light on the mechanisms through which each treatment affects household behavior.

In the conventional approach, researchers typically estimate heterogeneous treatment effects by analyzing subgroups through splitting or interacting the treatment with baseline characteristics. However, this method may result in overfitting or noisy estimates if the sample was not originally designed for such division. To address this issue, we follow Chernozhukov et al. (2018) and turned to machine learning (ML), which offers a disciplined approach to identifying relevant heterogeneity in treatment effects while avoiding the risk of overfitting. ML tools prove especially effective in high-dimensional settings, such as ours, where a multitude of observable characteristics come into play.

The objective is to estimate the *conditional average treatment effect* (CATE), which is the difference in the expected potential outcomes between treated and control states conditional on covariates. Nonetheless, according to Chernozhukov et al. (2018), in the absence of strong assumptions, it is uncertain whether generic ML tools can generate consistent estimators of the CATE. Hence, Chernozhukov et al. (2018) propose an alternative approach that focuses on valid estimation and inferences on features of the CATE, rather than on the CATE itself. Specifically, this approach considers three objects of interest.

First, the best linear predictor (BLP) of the CATE (denoted by $s_0(Z)$) on the ML proxy predictor of the CATE (denoted by $S(Z)$). The BLP is the solution to $\underset{\beta_1,\beta_2}{min} E[s_0(Z) - \beta_1 - \beta_2 S(Z)]^2$, where β_1 corresponds to the ATE and β_2 denotes the heterogeneity parameter. Rejecting the null hypothesis $\beta_2 = 0$ indicates that there is heterogeneity in the CATE and that $S(Z)$ is a relevant predictor of the CATE. Second, we report the group average treatment effects (GATES), which represent the average treatment effects at quantiles of the conditional treatment effect distribution. Third, the classification analysis (CLAN), provides the average characteristics of the most and least affected units to further explore which observable characteristics are associated with the heterogeneity.

The main advantage of this method is that it can be applied in combination with any ML method.[12] To choose among various ML methods, goodness-of-fit measures for the BLP and GATES are employed. We use four types of algorithms —random forest, elastic net, support vector machine, and gradient boosting— and provide the results for the best algorithm according to the goodness-of-fit measures. We use a set of 38 predetermined characteristics to estimate the ML proxy of the CATE and provide the CLAN for 23 variables. Specifically, we include a wide range of characteristics related to household composition, characteristics of the household head, income, and other household features. We provide additional implementation details as well as sensitivity analyses in Appendix B.

5 Results

Table 1 presents the estimated effects of receiving part of the transfer as a food voucher in the first row, labeling the transfer in the second row, and a combination of both the label and voucher in the third one, compared to receiving cash. The outcome variables are total household income (column 1), total monthly expenditure (column 2), and the shares of expenditure spent on food (column 3), adults (column 4), and children (column 5). All regressions in Table 1 include the vector of control variables described in Section 3 and displayed in Table D.3.

In the analysis, several noteworthy findings emerge. First, as expected, the results in column 1 suggest that there are no statistically significant differences in total household income between the experimental arms.

Second, the results in column 2 indicate that voucher recipients had, on

[12] Alternative methods for estimation and inference on heterogeneous effects based on ML rely on specific ML algorithms (Athey and Imbens, 2016; Wager and Athey, 2018).

Table 1: The effect of the voucher and the label on income, expenditure, and the composition of expenditure

	Total income (GEL) (1)	Total monthly expenditure (GEL) (2)	Food expenditure share (percent) (3)	Adult-related expenditure share (percent) (4)	Child-related expenditure share (percent) (5)
Voucher	-27.021	-96.943	-6.054	0.820	2.436
	(25.628)	(37.109)	(3.705)	(0.895)	(1.350)
	[0.377]	[0.046]	[0.206]	[0.401]	[0.168]
Label	-26.040	-48.651	-11.272	2.517	3.368
	(28.961)	(37.126)	(3.999)	(1.329)	(0.934)
	[0.401]	[0.296]	[0.036]	[0.162]	[0.006]
Voucher & Label	-11.022	-56.419	-5.414	0.928	3.200
	(26.075)	(23.874)	(3.900)	(0.887)	(0.661)
	[0.675]	[0.068]	[0.287]	[0.377]	[0.001]
Cash arm mean	647.39	469.21	40.86	4.98	7.27
Observations	4462	4462	4462	4462	4462

Notes: The table reports the estimated effects of the program arms following Equation (1). All regressions control for all household characteristics shown in Table D.3. Standard errors clustered at the municipal level are in parentheses and the FDR q-values are in brackets.

average, GEL 96.9 lower expenditure than cash recipients, which constitutes a 20.7 percent reduction in expenditure. This reduction in total expenditure remains statistically significant even after the FDR adjustment. By contrast, the effect of the label on household expenditure is not statistically distinguishable from zero. The reduction in total expenditure is also present in Arm 4, where households spent GEL 56.4 less than households in the cash group (a reduction of 12 percent).

Turning to column 3, we find that the voucher did not lead to a significant change in the share of expenditure allocated to food, despite this being its main goal. As previously discussed, most households already spent more in food than the value of the voucher. Conversely, the label led to a reduction in the share of food expenditure by 11.3 percentage points, corresponding to a 27.6 percent reduction compared to cash-only recipients. In addition, estimates from column 4 suggest that differences in the share of expenditure spent on adults between arms are not statistically distinguishable from zero.

Finally, column 5 indicates that receiving the message specifying the amount of the transfer to be spent on children increased the share of child-related expenditure by 3.4 percentage points, equivalent to a 46.3 percent increase, in comparison to cash-only recipients. Similarly, the share of expenditure related to children is 3.2 percentage points higher for label-and-voucher recipients than for cash-only recipients. Both of these estimates are statistically significant following the FDR correction. By contrast, the difference in the share of expenditure spent on children between voucher recipients and cash-only recipients is not statistically distinguishable from zero.

Table 2: The effect of the voucher and the label on the share of child-related expenditure by type

	Share of total expenditure on:		
	Childcare	Child clothing	Education
	(1)	(2)	(3)
Voucher	0.193	1.566	0.676
	(0.120)	(1.128)	(0.334)
	[0.171]	[0.194]	[0.088]
Label	0.160	1.983	1.226
	(0.111)	(0.804)	(0.501)
	[0.194]	[0.042]	[0.042]
Voucher & Label	0.003	1.973	1.225
	(0.027)	(0.792)	(0.357)
	[0.924]	[0.042]	[0.012]
Cash arm mean	0.03	5.98	1.26
Observations	4462	4462	4462

Notes: The table reports the estimated effects of the program arms following Equation (1). All regressions control for all household characteristics shown in Table D.3. Standard errors clustered at the municipal level are in parentheses and the FDR q-values are in brackets.

Next, we examine the impact of the program arms on three specific categories of child-related expenditures: childcare, child clothing, and education.[13] Table 2 reports the estimated effects of each arm. The results from column 1 of Table 2 indicate that there are no significant differences in the share of childcare expenditure between the program modifications. Columns 2 and 3 indicate that receiving the label increases the share of expenditure spent on child clothing

[13]Expenditure on education does not include any tuition fees, as education is free and compulsory for children under 16. The expenditure in education includes other fees, transportation, materials for learning at home or school.

and education by 2 and 1.2 percentage points, respectively.

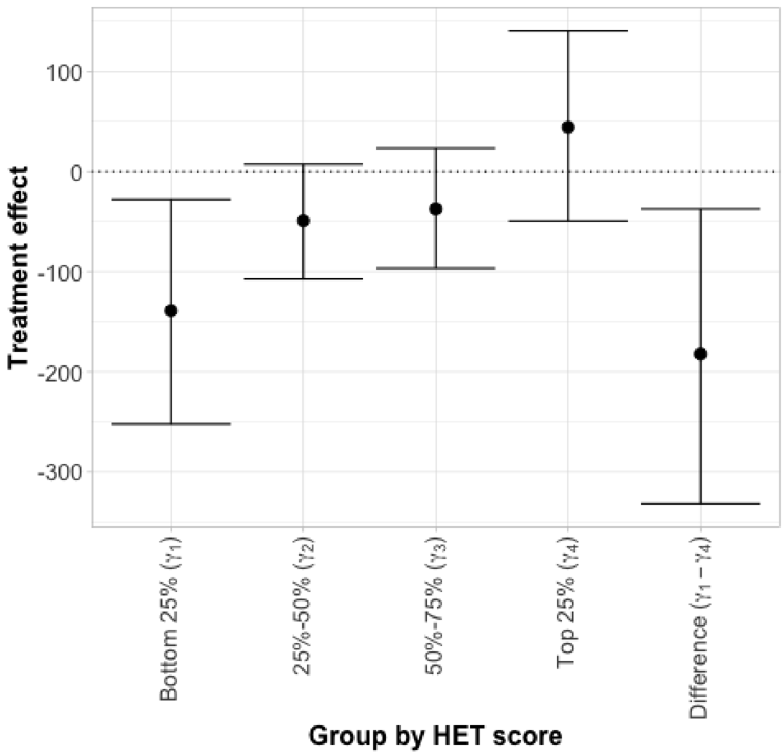
Overall, the results indicate that (1) receiving part of the transfer as a food voucher reduces total consumption; (2) the labels induce households to shift their expenditure from food to children's education or clothing.

5.1 Mechanisms

One possible explanation to differences in expenditure patterns between arms could be that compliance with the TSA changed in response to the label or the voucher. While the balancing tests indicate that there are no statistically significant differences in the PMT scores obtained from the first interview, which were used to determine the initial amount of transfers, it is still possible that there are differences in the actual amount of TSA benefit received between the program arms. Specifically, households may increase (or decrease) the TSA benefit, or some households may not receive any benefit at all, if their PMT scores change during a post-treatment re-evaluation process. Notably, households had the option to request a re-evaluation, and, in addition, some households were able to get additional interviews in response to some wealth related shocks.

To address the relevance of this potential mechanism, we estimate in Table D.6 the effect of program arms on several indicators. These variables include whether a household is an active TSA beneficiary (column 1), the actual amount of monthly TSA transfer received (column 2), the indicator for households that were re-evaluated (column 3), and the total number of interviews conducted (column 4). Reassuringly, the estimates reveal no discernible systematic differences in these variables between the program arms. Hence, one can argue that the program arms do not change the relationship between households and the TSA, and the difference in expenditure patterns we explained above are not driven by differences in the total amount of benefits obtained by each household.

Figure 1: GATES of voucher on total expenditure



Notes: The first four columns display the group average treatment effects (GATES) divided by quartiles of the conditional average treatment effect of the voucher on total expenditure. $Voucher = 1$ when $Arm = 2, 4$. The difference in the average treatment effect between the most and least affected groups is presented in the last column. The estimations are based on the best ML learner (elastic net) following Chernozhukov et al. (2018), using the parameters explained in Appendix B.

5.1.1 Why did the voucher reduce total expenditure?

To shed light on the underlying reasons for the observed reduction in total expenditure among voucher receivers, we begin by analyzing the heterogeneity of the results using the generalized ML approach described in Section 4.3. In this analysis, treatment is defined as being assigned to Arm 2 or Arm 4 —i.e., those receiving vouchers, either with or without the label.

Figure 1 displays the resulting group average treatment effects of the voucher on total expenditure, with the sample divided by the quartiles of the distribution of conditional average treatment effects. The results suggest that about 25 percent of the households in our sample are negatively affected by the voucher. For the most adversely affected group (the first quartile of the average treatment effect distribution), the reduction in expenditure amounts to GEL 139.0 per month. However, for households in the other quartiles of the average treatment effect, there is no statistically significant reduction in monthly expenditure. Notably, the negative effect of the voucher on expenditure for households in the first quartile is significantly different from the weak positive effect of the voucher observed for households in the last quartile (last column of Figure 1).[14]

Table 3: CLAN of the effect of the voucher on total expenditure for selected characteristics

	Most affected δ_1 (1)	Least affected δ_4 (2)	Difference $\delta_1 - \delta_4$ (3)	*p-value* (4)
Household composition				
Adult Members	3.059	3.658	-0.595	0.000
More than 2 children	0.190	0.283	-0.090	0.000
Children under 5 in the household	0.564	0.667	-0.104	0.000
Household's head characteristics				
Age	52.509	53.030	-0.645	0.476
Female	0.367	0.323	0.039	0.139
Low-education	0.134	0.220	-0.086	0.000
Household's characteristics				
Has agricultural land	0.815	0.849	-0.034	0.129
TSA indicators				
PMT score	60.492	53.693	7.149	0.000
Benefit in voucher	54.946	62.608	-7.527	0.000
More than 15 minutes to the closest shop	0.387	0.530	-0.142	0.000
Labeled cash transfer	0.381	0.582	-0.195	0.000
GATE	-138.999	44.068	-182.346	0.014

Notes: CLAN stands for Classification Analysis. Column 1 and 2 report the estimated sample average of each characteristic of the households belonging to the most (δ_1) and least affected group (δ_4), according to the proxy of the Conditional Average Treatment Effect (CATE) of the Voucher on total expenditure. $Voucher = 1[\text{Arm} = 2, 4]$. Column 3 reports the difference between Column 1 and Column 2, and Column 4 the p-value of such difference, where the standard errors are clustered at the municipality level. The last row shows the respective GATE as in Figure 1. Estimations based upon the best Causal Learner following Chernozhukov et al. (2018) using the parameters explained in Appendix B. The results for the complete set of characteristics are in Table B.4 in Appendix B.

Table 3 presents the results of the classification analysis, showing the average values of the characteristics of households in the bottom (most affected) and top (least affected) quartiles of the average treatment effect distribution. Our findings reveal that, on average, the most negatively affected households tend to be smaller in size, have fewer children, be headed by better educated individuals, exhibit lower economic vulnerability (indicated by a higher PMT score), receive

[14] This heterogeneity is also captured by the heterogeneity parameter β_2 in Panel A of Table B.2 in Appendix B, which is equal to 0.938 (with confidence intervals of 0.35-1.53) for the best ML learner (elastic net).

a smaller benefit in vouchers, live closer to the closest shop, and be less likely to be recipients of the label compared to their least affected counterparts.

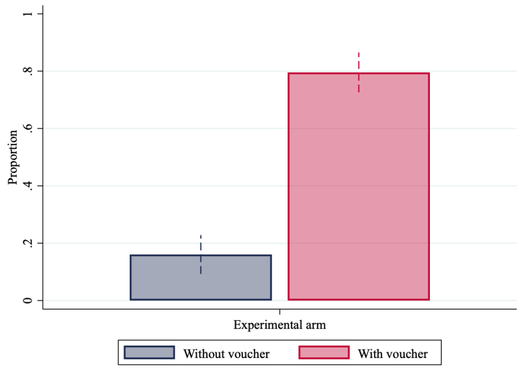
These results contribute to our discussion of the potential mechanisms explaining why households receiving part of their transfer as a food voucher experienced a reduction in total expenditure. We investigate three plausible explanations. First, the discrepancy in shopping costs between voucher and cash recipients, stemming from the limited usage of vouchers in specific shops, may have played a role. Second, we consider the possibility of increased food prices resulting from the introduction of food vouchers. Third, we test the possibility that lower expenditure comes from households reporting errors, as they might not have accurately accounted for their expenses when using the voucher.

The relative cost of shopping with the voucher. As explained in Section 2, food vouchers were only accepted in particular shops, which might not have been the preferred choice for households. For example, due to less convenient locations compared to their regular shops. Consistently with this explanation, the heterogeneity analysis in Table 3 indicates that the negative effect is primarily concentrated in households where the transfer was smaller and for households located close to a shop (voucher or non-voucher). In these households, the gains from a relatively small food voucher may not compensate for the inconvenience and potential added cost of switching to the voucher-accepting shop.

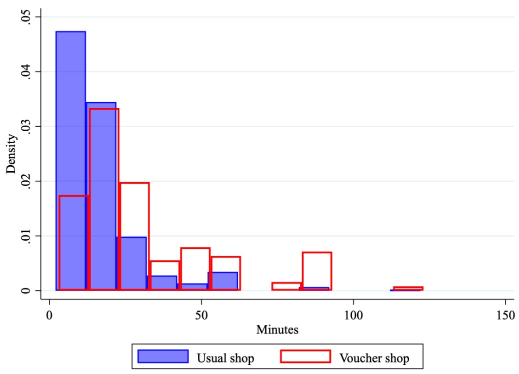
To further illustrate this point, we show in Figure 2.a that 80 percent of families in voucher-receiving municipalities opted for shops that accepted vouchers, whereas only 20 percent of families in non-voucher municipalities did the same. This difference in the use of of voucher-accepting shops supports the notion of a higher shopping cost associated with these establishments.

In line with our previous discussion, Figure 2.b shows that households' usual shops are, on average, closer to the households compared to the voucher-accepting shops. Specifically, the average time to arrive at their usual shop is 19 minutes, whereas it extends to 31 minutes for the voucher shop. Additionally, Figure 2.c shows that households were also more likely to use motor transport

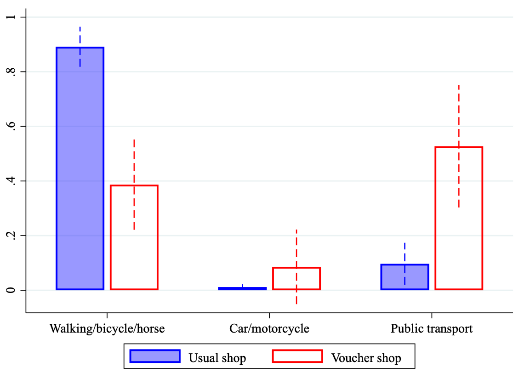
Figure 2: Use, time, and means of transport to go to voucher-accepting shops and the closest shop



(a) Proportion of households that shop at voucher-accepting shops by treatment group



(b) Time to the shop



(c) Transport means to the shops

Notes: Dashed lines represent 95 percent confidence intervals with standard errors clustered at the municipal level in parenthesis. Panel (a) compares households in Arms 1 and 3 (without voucher), with households in Arms 2 and 4 (with voucher). Estimates in panels (b) and (c) using information only from households in Arm 1 (cash only).

(either their own or public) to go shopping to the voucher shops than to their usual shops.[15] These findings collectively suggest that the reduced expenditure observed among voucher recipients may be associated with the inconvenience and additional cost associated with accessing voucher-accepting shops, which were less geographically accessible.

It is crucial to highlight that households could not save the value of the

[15] As mentioned in Section 3, it is important to note that our analysis excludes Tbilisi due to its assignment to the food voucher outside of the random allocation. Nevertheless, Figure D.2 in Appendix D shows similar patterns in Tbilisi, with voucher-accepting shops located farther from households and households in Tbilisi being more likely to use motorized transportation when shopping at voucher-accepting shops, compared to their usual shops.

voucher for the following month; i.e., any unused voucher funds would expire at the end of each month. In addition, voucher recipients were required to show their identification documents when using the voucher for payment. Consequently, they could not trade their vouchers to other people. As illustrated in Table D.7 in the Appendix, there are no significant differences between the program arms in the share of households that report having savings in the past month. However, the results indicate that the difference between total income (including the transfer and the voucher) and total expenditure is is greater in voucher-receiving municipalities, thus suggesting that some voucher recipients simply did not take advantage of their vouchers again due to the inconvenience associated with shopping at voucher-accepting stores.

Other possible mechanisms. We analyse two additional channels. First, we examine whether there are any differences in food prices between voucher municipalities and non-voucher municipalities. In isolated markets, cash transfers can lead to price increases as they boost the demand for normal goods. This effect may be more pronounced when transfers are designated for purchasing specific items, such as in the case of food vouchers (Basu, 1996; Hidrobo et al., 2014; Cunha, 2014). It is worth noting, however, that in this particular context it is unlikely that difference in food prices could account for the adverse impact of vouchers on total expenditure, given the absence of disparities in saving rates between the program arms.

In line with the approach taken by Cunha (2014), we conducted surveys to collect information on the prices households paid for a list of food items. Using this data, we constructed a price index, which is described in Appendix A.[16] Figure D.3 in Appendix D illustrates the distribution of the price index in municipalities assigned to different arms. The results indicate that the distributions of food prices are remarkably similar among all arms, suggesting that the differences in food prices between municipalities assigned to the voucher and cash do not seem to be a plausible explanation for the observed adverse impact of

[16] We use the top 20 items in terms of their importance in the household budget for calculating the CPI by the National Statistics Office of Georgia.

vouchers on total expenditure.

In addition, we explore the possibility of households in voucher municipalities misreporting their total expenses due to potentially excluding what they spent using the voucher. To address this concern, we construct a measure of the reporting error by comparing the value of the transfer reported by households with the value we calculated based on the family composition and the last PMT score. Figure D.4 in the Appendix presents the distribution of the reporting error for different program arms. The results suggest that households receiving vouchers are actually less likely to under-report the value of the transfer compared to households receiving cash. In addition, Table 3 also shows that household heads with lower education are less prevalent among households where we observe a negative effect of the voucher on expenditure. Hence, if education is correlated with the likelihood of reporting error, the CLAN analysis does not support the idea that reporting error drives the negative effect of the voucher on total expenditure.

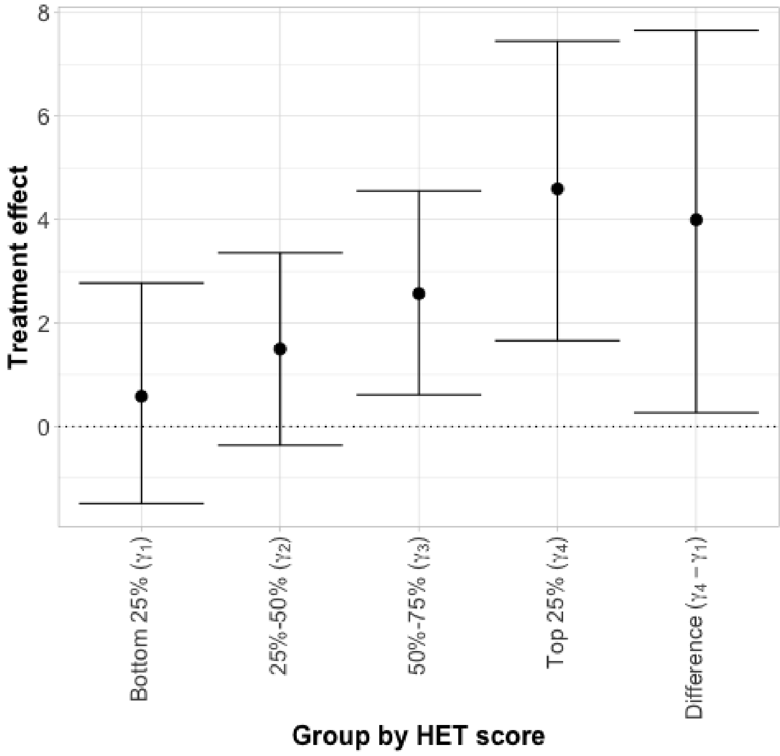
In summary, the evidence supports the hypothesis that the negative effect of the voucher on total expenditure can be attributed to the inconvenience of using vouchers at specific shops. However, we do not find any evidence suggesting that the effect is driven by differences in prices or reporting errors between voucher and non-voucher municipalities.

5.1.2 Understanding the effect of labels on child-related expenditure share

In order to understand the drivers of the labeling effect on the share of child-related expenditure, we conduct the heterogeneity analysis using the generalized ML. In this case, the label treatment is defined as being assigned to Arms 3 or 4 —either label or label and voucher.

Figure 3 presents the estimated group average treatment effects of the label on the share of child-related expenditure. The sample is divided into quartiles based on the conditional average treatment effect, with the first quartile being the least affected by the label, and the fourth quartile being the most affected. The results suggest that the effect is positive and statistically different from zero for the last two quartiles, indicating that approximately 50 percent of households increase the share of expenditure spent on children due to the label. In the most affected quartile (column 4 of Figure 3), the effect amounts to a 4.6 percentage point increase in child-related expenditure, while in the least affected group (column 1 of Figure 3), the effect is 0.6 percentage points, and it is not statistically significant. The difference in the effect between the most and least affected groups is reported in the last column of Figure 3, indicating a significant heterogeneity in the labeling effect.[17]

Figure 3: GATES of labeled cash transfer on the share of child-related expenditure



Notes: First four columns show the Group Average Treatment Effects (GATES) by quartiles of a proxy of the Conditional Average Treatment Effect (CATE) of the label on the share of child-related expenditure. $Label = 1[\text{Arm} = 3, 4]$. The difference in the ATE between the most and the least affected group is shown in the last column. Estimations based upon the best Causal Learner following Chernozhukov et al. (2018) using the parameters explained in Appendix B.

In addition, we conduct the classification analysis in Table 4 aimed at identifying those household characteristics contributing to the heterogeneity of the

[17] Consistently, the heterogeneity parameter β_2 in Panel B of Table B.2 in the Appendix is 0.58 for the best ML learner (random forest), with a confidence interval from 0.1 to 1.07.

Table 4: CLAN of the effect of the labeled cash transfer on child-related expenditure for selected characteristics.

	Least affected δ_1 (1)	Most affected δ_4 (2)	Difference $\delta_4 - \delta_1$ (3)	*p-value* (4)
Household composition				
Adult Members	3.305	3.227	-0.079	0.342
More than 2 children	0.079	0.398	0.326	0.000
Children under 5 in the household	0.720	0.514	-0.203	0.000
Household's head characteristics				
Age	52.778	51.396	-1.437	0.118
Female	0.287	0.407	0.118	0.000
Low-education	0.167	0.223	0.059	0.010
Household's characteristics				
Has agricultural land	0.665	0.910	0.249	0.000
TSA indicators				
PMT score	58.748	55.423	-3.265	0.009
Child benefit	77.016	118.280	40.681	0.000
Voucher	0.720	0.337	-0.385	0.000
GATE	0.580	4.593	3.993	0.036

Notes: CLAN stands for Classification Analysis. Column 1 and 2 report the estimated sample average of each characteristic of the households belonging to the least (δ_1) and most affected group (δ_4), according to the proxy of the Conditional Average Treatment Effect (CATE) of the label on child-related expenditure. $Label = 1[\text{Arm} = 3, 4]$. Column 3 reports the difference between Column 2 and Column 1, and Column 4 the p-value of such difference, where the standard errors are clustered at the municipality level. The last row shows the respective GATE as in Figure 3. Estimations based upon the best Causal Learner following Chernozhukov et al. (2018) using the parameters explained in Appendix B. The results for the complete set of characteristics are in Table B.3 in Appendix B.

labeling effect. Specifically, the table reports the average values of the characteristics in the least effected quartile (column 1) and the most affected quartile (column 2). This analysis provides insights into potential mechanisms underlying the labeling effect.

The results indicate that households with more than two children and where the child benefit was larger are more prevalent among those experiencing significant treatment effects. This suggests that the salience of the label enhances the treatment effect. This finding may align with the concept of mental accounting (Thaler, 1999), which describes how individuals tend to categorize their money into different accounts based on its source and purpose. Therefore, the label could induce households to mentally allocate the child benefit specifically for child-related spending, potentially increasing the total expenditure on children, especially if the child benefit is extramarginal. Unfortunately, we do not observe child-related expenditure before the program started, which prevents us from directly testing whether the child benefit is extramarginal. However, only 5 percent of households in the cash arm spent in their children more than the quantity they receive as child benefits. Even if we include children's food expenditure, which we compute by assuming that food expenditure is distributed equally among all household members, 47 percent of households still spend less on their children than the value of the child benefit. Hence, these results indicate that the child benefit is likely to be extramarginal as for a significant portion of the population there is scope to increase their expenditure in children to match the child benefit (see Figure D.1 panels b and c in Appendix D).

Furthermore, rural households and households where the household head has low education are among those experiencing a more pronounced impact of the label. This result is consistent with extensive evidence reviewed in Damgaard and Nielsen (2018), suggesting that many behavioral interventions are more effective for children from low socio-economic-status families or those with characteristics that are probably correlated with low parental socio-economic-status.

Additionally, female-headed households are more common among the most affected households, with 40.7 percent of the most affected households being female-headed, compared to 28.7 percent among the least affected households. This observation is particularly noteworthy because the household head is typically the person who opens the bank account where the benefit is transferred and, consequently, one can assume the household head has greater control over the benefit. This finding is in line with the causal evidence provided in Lundberg et al. (1997), who shows that transferring child allowance from fathers to mothers was associated with an increase in child-related expenditures in the UK. In addition, Lundberg et al. (1997) concludes that mother attach more weight to child-related expenditure, such as child clothing, compared to fathers. Furthermore, and also consistent with our results, the findings of Benhassine et al. (2015) indicate that CCT has a greater effect on the educational outcomes of children when the recipient is a woman, exploiting the randomization of the gender of the CCT recipient.

Finally, households that also received the food voucher are less common among those most affected by the label. Taking into account that the voucher had a negative effect on households' total expenditure, this may have limited the ability of households to respond to the message. Moreover, the voucher constrains parents to spend a substantial part of the child benefit on food, which may explain a smaller labeling effect on non-food child-related items for voucher recipients.

5.2 Additional Outcomes

To gain further insights into the effects of the voucher and label on food consumption, we analyze the impact of program arms on the shares of different food items over total food expenditure (Table D.8, Appendix D), as well as the consumption of specific food items for children younger than 6 —Table D.9, Appendix D. The findings do not provide evidence that program arms have a significant effect on food composition.

Finally, we also analyze whether the positive effect of the label on child-related expenditure translates into an effect on educational outcomes for children aged 6 to 16. The results presented in Table D.10 (Appendix D) suggest that there are no differences in school attendance, the probability of missing a school day, or expectations to receive university education between the program arms. This result could be explained by the fact that the survey was implemented only eight months after the program implementation, which might not have allowed sufficient time to observe significant changes in school attendance behaviors.

5.3 Sensitivity Tests

As a final robustness check, we conduct two sets of sensitivity tests to strengthen the robustness of our findings. Firstly, we excluded 628 households that were originally assigned to receive a TSA transfer according to the first interview but that were not receiving it at the time of our survey for different reasons.[18] In Table D.11 in Appendix D, we replicate Table 1 using a sample that excludes non-recipients of the TSA benefits from the estimation sample. Consistently, the estimates align with our main findings.

Secondly, we test whether our two main findings, namely the negative effect of the voucher on total expenditure and the positive effect of the label on the share of child-related expenditure, are influenced by a specific municipality. To conduct this test, we estimate equation (1) while excluding one municipality at a time from a total of 45 municipalities. Figure D.5 in Appendix D displays the distribution of the estimated effects of the voucher on total monthly expenditure (Panel A) and the label on the share of child-related expenditure (Panel B) from the 45 regressions. The effect of the voucher on total monthly expenditure ranges from -108 to -59, while the effect of the label on the share of child-related expenditure ranges from 2.8 to 4.1. This test indicates that our main findings are not driven by a specific municipality.

6 Conclusion

Cash transfer programs are widely used to reduce poverty and the implementation details may influence the effectiveness of these programs. This paper investigates the impact of two low-cost modifications of unconditional cash transfers, namely, labeling and the use of food vouchers, by exploiting the random allocation of Georgian municipalities to the different program arms of the nationwide social assistance program.

First, we demonstrate that labeling the cash transfer results in a significantly higher proportion of expenditure allocated to children, as the labels provide information about the portion of the transfer designated for child-related expenses. This effect is particularly larger in educational spending. Additionally,

[18] Households may lose their benefits if their re-evaluated PMT score is above 100K. Additionally, some households may lose their benefits if the Social Services Agency discovers that they are withholding information about changes in household composition or place of residence. Lastly, some households may voluntarily choose to opt out of the program.

we do not find any labeling effect on total household income, overall expenditure, or the share of expenditure allocated to adult-related expenses, while the share of expenditure on food decreases. Our heterogeneity analysis results suggest that labels may induce mental accounting, as the labeling effect is more pronounced when the relative value of the child benefit is higher.

Secondly, we show that recipients of food vouchers have lower total expenditure compared to those receiving cash only, even though the proportion of total expenditure allocated to food remains similar between the two groups. We find that this result is likely to be due to the fact that shops accepting vouchers are situated in less convenient locations than the regular stores typically frequented by households. This suggests that the use of vouchers induces an increase in shopping costs, potentially discouraging households from using them, ultimately leading to reduced total expenditure.

While the data allows us to study short-term outcomes measured approximately a year after the intervention, further research should also investigate the long-term effects. Overall, our findings suggest that the integration of cash transfers with information interventions may shape the short-term impacts of unconditional cash transfers, facilitating the achievement of desired outcomes without incurring additional costs.

Final note: In 2022, thanks to the input of this project, the Government of Georgia decided to eliminate the voucher as it was hindering households expenditure.

References

ABELER, J. AND F. MARKLEIN (2017): "Fungibility, labels, and consumption," *Journal of the European Economic Association*, 15, 99–127.

ANDERSON, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 103, 1481–1495.

Athey, S. and G. Imbens (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113, 7353–7360.

Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014): "Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes," *Journal of Development Effectiveness*, 6, 1–43.

Baird, S., C. McIntosh, and B. Özler (2011): "Cash or condition? Evidence from a cash transfer experiment," *The Quarterly Journal of Economics*, 126, 1709–1753.

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, and T. Schmidt (2019): "The impact of cash transfers: a review of the evidence from low-and middle-income countries," *Journal of Social Policy*, 48, 569–594.

Basu, K. (1996): "Relief programs: when it may be better to give food instead of cash," *World Development*, 24, 91–96.

Baum, T., A. Mshvidobadze, and J. Posadas (2016a): *Continuous Improvement: Strengthening Georgia's Targeted Social Assistance Program*, World Bank Publications.

——— (2016b): *Revision of the PMT Formula*, World Bank Publications.

Beatty, T. K., L. Blow, T. F. Crossley, and C. O'Dea (2014): "Cash by any other name? Evidence on labeling from the UK Winter Fuel Payment," *Journal of Public Economics*, 118, 86–96.

Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen (2015): "Turning a shove into a nudge? A" labeled cash transfer" for education," *American Economic Journal: Economic Policy*, 7, 86–125.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006): "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 93, 491–507.

Blow, L., I. Walker, and Y. Zhu (2012): "Who benefits from child benefit?" *Economic Inquiry*, 50, 153–170.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82, 2295–2326.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018): "Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India," Working Paper 24678, National Bureau of Economic Research.

Cunha, J. M. (2014): "Testing paternalism: Cash versus in-kind transfers," *American Economic Journal: Applied Economics*, 6, 195–230.

Dahl, G. B. and L. Lochner (2012): "The impact of family income on child achievement: Evidence from the earned income tax credit," *American Economic Review*, 102, 1927–1956.

Damgaard, M. T. and H. S. Nielsen (2018): "Nudging in education," *Economics of Education Review*, 64, 313–342.

Econometría (2020): "Impact Evaluation of Targeted Social Assistance (TSA) in Georgia. Final Report," Tech. rep., Econometría Consultores. Evaluation summoned by UNICEF.

Edmonds, E. (2002): "Reconsidering the labeling effect for child benefits: evidence from a transition economy," *Economics Letters*, 76, 303–309.

Efron, B. (2012): *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, vol. 1, Cambridge University Press.

Evans, G. W. (2016): "Childhood poverty and adult psychological well-being," *Proceedings of the National Academy of Sciences*, 113, 14949–14952.

García, S. and J. E. Saavedra (2017): "Educational impacts and cost-effectiveness of conditional cash transfer programs in developing countries: A meta-analysis," *Review of Educational Research*, 87, 921–965.

Gentilini, U. (2022): *Cash transfers in pandemic times: Evidence, practices, and implications from the largest scale up in history*, World Bank.

Gentilini, U., M. Almenfi, I. Orton, and P. Dale (2022): "Social protection and jobs responses to COVID-19: A Real-Time Review of Country Measures," Working paper, The World Bank.

Hastings, J. and J. M. Shapiro (2018): "How are SNAP benefits spent? Evidence from a retail panel," *American Economic Review*, 108, 3493–3540.

Haushofer, J., M. Larreboure, S. Lowes, and L. Mait (2023): "Cash Transfers and Social Preferences of Children," Working Paper 31720, National Bureau of Economic Research.

Hawkins, A. A., C. A. Hollrah, S. Miller, L. R. Wherry, G. Aldana, and M. D. Wong (2023): "The Long-Term Effects of Income for At-Risk Infants: Evidence from Supplemental Security Income," Working Paper 31746, National Bureau of Economic Research.

Hendren, N. and B. Sprung-Keyser (2020): "A unified welfare analysis of government policies," *The Quarterly Journal of Economics*, 135, 1209–1318.

Hidrobo, M., J. Hoddinott, A. Peterman, A. Margolies, and V. Moreira (2014): "Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador," *Journal of Development Economics*, 107, 144–156.

Hoynes, H. W. and D. W. Schanzenbach (2009): "Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program," *American Economic Journal: Applied Economics*, 1, 109–39.

Hummel, D. and A. Maedche (2019): "How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies," *Journal of Behavioral and Experimental Economics*, 80, 47–58.

Kooreman, P. (2000): "The labeling effect of a child benefit system," *American Economic Review*, 90, 571–583.

Lundberg, S. J., R. A. Pollak, and T. J. Wales (1997): "Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit," *Journal of Human Resources*, 463–480.

Orkin, K., R. Garlick, M. Mahmud, R. Sedlmayr, J. Haushofer, and S. Dercon (2023): "Aspiring to a Better Future: Can a Simple Psychological Intervention Reduce Poverty?" Working Paper 31735, National Bureau of Economic Research.

Pega, F., R. Pabayo, C. Benny, E.-Y. Lee, S. K. Lhachimi, and S. Y. Liu (2022): "Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low-and middle-income countries," *Cochrane Database of Systematic Reviews*.

Power, C., D. Kuh, and S. Morton (2013): "From developmental origins of adult disease to life course research on adult disease and aging: insights from birth cohort studies," *Annual Review of Public Health*, 34, 7–28.

Silwal, A. R., S. Engilbertsdottir, J. Cuesta, D. Newhouse, and D. Stewart (2020): *Global estimate of children in monetary poverty: An update*, World Bank.

Southworth, H. M. (1945): "The economics of public measures to subsidize food consumption," *Journal of Farm Economics*, 27, 38–66.

Thaler, R. H. (1999): "Mental accounting matters," *Journal of Behavioral Decision Making*, 12, 183–206.

Thaler, R. H. and C. R. Sunstein (2008): *Nudge: Improving Decisions About Health, Wealth, and Happiness*, New Haven: Yale University Press.

Wager, S. and S. Athey (2018): "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 113, 1228–1242.

World Bank (2018): "Georgia—First, Second and Third Development Policy Operations," *Project Performance Assessment Report*.

Young, A. (2019): "Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," *The Quarterly Journal of Economics*, 134, 557–598.