Navarra Center for International Development

Working Paper nº 04/2024

Welfare Eligibility Manipulation: Evidence From Georgia

Brendon McConnell
City, University of London

Jaime Millán-Quijano
University of Navarra

Welfare Eligibility Manipulation: Evidence From Georgia*

Brendon McConnell Jaime Mill´an-Quijano

This draft: March 21, 2024 Most
recent version in this link

Abstract

Optimal targeting of social aid is a fundamental issue in public policy design. A key aspect is to create welfare systems that are manipulation-proof. Using a rich tapestry of administrative and survey data, we study household-driven manipulation of a nationwide welfare program in Georgia. We start by documenting sizable bunching at a benefit discon- tinuity. Next, we build a Becker (1968)-style model of manipulation, which we use to inform our empirical strategy – a fuzzy difference-in-discontinuity design. Our estimation strategy and rich data allow us to (i) characterize those households who manipulate at a key wel- fare score threshold, (ii) document how these households manipulate their eligibility scores, and (iii) provide evidence on the downstream consequences of manipulation – in the labor market, in household expenditure patterns, and for a raft of child outcomes. We document that manipulation of scores appears to be needs-driven – it is rural and marginally poorer households who are more likely to manipulate. These households do so primarily by hiding rural assets prior to inspection. We find meaningful increases in the labor supply of women in manipulating households, and a concomitant increase in expenditure on the children in the household, likely driven by an economically-driven increase in the bargaining power of women within the household. Our labor supply findings are driven by women in households with unsuccessful manipulation attempts – we do not document a labor supply response for this in households where manipulation succeeds. This suggests a crowd-out effect of welfare income on labor market participation for women. We estimate the cost to government coffers of manipulation and find the form of welfare manipulation that we study in this work leads to a cost that is 25% of the initial expenditure on our target households. We conclude the paper by comparing our approach to the standard alternative in the literature – bunching estimators – and show, that in our setting, bunching estimators are consistent with our approach of welfare eligibility manipulation.

Keywords— Welfare Eligibility, Manipulation, Public Policy Design, Child Skill Investment
JEL Codes— H53, I38

*Author affiliations and contacts: McConnell (City, University of London, brendon.mcconnell@gmail.com); Mill´an-Quijano (NCID and CEMR, jmillanq@unav.es). We are grateful to the evaluation team at Econometr´ıa S.A., UNICEF Georgia, GeoStat and the Social Services Agency of Georgia. We are also thankful to Dirk Foremny, Alan Manning, Anastasia Terskaya, and Marcos Vera-Hern´andez for their valuable comments. The authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness Grant PID2020- 120589RA-I00.

1 Introduction

Optimal targeting of social aid is a key issue in the design public policy, irrespective of a country's level of development (Coady et al., 2004; Alatas et al., 2012). While in developed economies targeting uses rich administrative, in developing economies proxy mean tests (PMTs) are commonly used to allocate access to social programs. In both cases, the design of targeting schemes must take into account the response of potential beneficiaries to the program features, not least the incentive to game or manipulate the system (Coady et al., 2004).

We can divide the source of social benefits manipulation into two categories. First, *demand side* manipulation, where manipulators are the final beneficiaries and engage in manipulation for their direct gain. Examples include tax evasion (Friedberg, 2000; Saez, 2010; Kleven et al., 2011; Kleven and Waseem, 2013), access to health services (Miller et al., 2013), and major offices manipulating population statistics to obtain fiscal benefits (Foremny et al., 2017). Second, *supply side* or *intermediary* manipulation, whereby an intermediary or service provider manipulates access to, or elements of, the program. Examples include teachers changing student grades in high stakes tests (Machin et al., 2020), up-coding in health insurance (Geruso and Layton, 2020), or local governments allowing program access to ineligible household in order to gain votes (Camacho and Conover, 2011; Brollo et al., 2020).

In this paper we study the case of social welfare eligibility manipulation in the Targeted Social Assistance (TSA) program in the nation of Georgia. We provide compelling evidence of demand side manipulation at a key eligibility-score threshold. We build a theoretical model of manipulation, from which we bridge to our core empirical approach. We provide novel empirical evidence on the type of households that engage in welfare manipulation, the strategy followed by those households to manipulate, and, using a wealth of administrative and survey data, characterize the consequences of manipulation on a wide set of outcomes. By studying the responses of beneficiaries, notably demand side manipulation, our work contributes to the literature studying optimal welfare scheme design.

First, we provide evidence of demand side manipulation in the TSA program – a nationwide program that uses a proxy means test (PMT) with multiple cutoffs to allocate unconditional cash transfers among low income households.[1] The program allows reassessments to households that could lead to changes in the PMT score. When the household situation changes – a child is born, a household member dies, a household member becomes disabled, the household purchases a car – the household is re-assessed and a new PMT score calculated. There is a second route to PMT score reassessment – *household- initiated* reassessments – whereby households who feel their PMT score does not accurately represent their level of welfare may request a new assessment, at least one year after their initial assessment. It is these household-initiated reassessments that will be the core focus of our work on welfare eligibility manipulation in this paper.

To document manipulation we present graphical evidence comparing the initial and final observed distributions of the PMT score, paying special attention to a key threshold (a PMT score of 65,000). We supplement the visual evidence with a CJM density test (Cattaneo et al., 2020). Using the rich administrative data we have available to us, we use the reasons for reassessment to separate reassessments into two categories – household-initiated reassessments, and social security agency-initiated reassessments. This enables us to understand the *source* of any PMT score discontinuities.

Informed by the nature of manipulation attempts, we set up a model of household welfare manipu- lation, based on the insights of the Becker model of crime (Becker, 1968). The purpose of this model is twofold. First, it informs the structure of our (reduced-form) empirical specification. Secondly it makes clear the importance of household heterogeneity in unobserved willingness to manipulate, a key point when bridging with out empirical strategy.

Based on the evidence of behavioral responses of household to discontinuities in the welfare income-

[1] The PMT is based on rich survey data that incorporates information on households' demographic composition, asset holdings, income, and access to public amenities.

PMT score schedule, we bridge from the theoretical model and set up an empirical specification in the form of a fuzzy difference-in-discontinuity (FDD) design (following Grembi et al., 2016; Mill´an-Quijano, 2020). Our key endogenous treatment variable is the decision to engage in welfare manipulation, and our instrument is a binary indicator for receiving an initial PMT score about 65,000 – our key threshold of interest. In the spirit of our FDD strategy, we use variation around an alternative cutoff to account for the effect of changes in benefits on key outcomes.

We provide supportive evidence of the identifying assumptions required for this approach – that our running variable is continuous through the key cutoff of interest, and that expected potential outcomes are smooth through the cutoff. We provide evidence for the latter in two ways – first, by examining the continuity of household characteristics through the cutoff, and second by documenting the smooth evolution of key policy parameters (which may reflect both household observables and unobservables) through the cutoff. The evidence we document provides strong support for the underlying identifying assumptions. We also show evidence that the effect of additional benefits is homogeneous over the PMT distribution, the key complementary assumption of the FDD framework.

We then restrict the data to a narrow window around the key PMT score thresholds in order to consider a relatively homogeneous group of welfare recipients. The households in our working sample are poor, with a total income (labor income plus all welfare benefit income) of roughly 100 USD per month. We characterize the compliers in our IV framework – households whose manipulation status is induced by falling above/below the 65,000 PMT score threshold – using the approaches of Abadie (2003) and Dahl et al. (2014). In addition to characterizing compliers across a wide set of household and property characteristics, we can directly examine where complier households fall within the baseline income distribution. For each of the non-income characteristics, we present the partial correlation between the characteristic and baseline earned income in order to assess if the characteristic in question is positively or negatively associated with income.

Making use of the detailed administratrive data we have access to in this study, we then consider *how* households manipulate their scores. To do so, we exploit the richness of our administrative data, which includes each and every input into the combined score that yields a PMT score. We compare changes over time in PMT score input variables for households who request a repeat interview – our manipulation proxy – with household that have repeat interviews triggered by the social security agency. As such, our analysis takes the form of a difference-in-differences approach.

Finally, we document the consequences of welfare eligibility manipulation on a wide set of outcomes. This is where we make full use of the rich tapestry of data sources we have available to us – both multi-agency administrative data and survey data we collected as part of a related project. We start by considering the impact of welfare eligibility manipulation on labor market outcomes – both formal and informal – and proceed to study the consequence of welfare manipulation for expenditure patterns within the household. We end by considering the impact of household welfare manipulation on specific household members, namely children and young people.

Our first key empirical finding is to document substantial manipulation of welfare eligibility. We present graphical evidence that (i) the initial distribution of PMT scores are smooth and continuous through all key benefit cutoffs and (ii) the final observed distribution has unnatural bunching to the left

– the side that yields higher benefit income – of a key threshold (a PMT score of 65,000). We confirm the visual evidence of bunching with a CJM density test (Cattaneo et al., 2020). We then present the probability of a PMT score reassessment across the PMT score distribution. We find a large, statistically significant jump in the probability of reassessment precisely at the PMT score threshold of 65,000 and nowhere else. The probability of a reassessment is approximately 35% below the 65,000 threshold. We document a 20 percentage point jump at 65,000. Using the information on the source of the reassessment, we find that household-initiated reassessments are the exclusive driver of the total effect. The probability to be reassessed jumps precisely at 65,000 by 20 percentage points (from a base of 10 percentage points) for household-initiated cases, yet social security agency-initiated reassessments are smooth through all

PMT score-based cutoffs.

Our second finding is based on our main instrument – falling just above the PMT score threshold of 65,000. We use the approaches of Abadie (2003) and Dahl et al. (2014) to characterize the complier households within our working sample. We find the compliant sub-population – those who change their manipulation status based on their PMT score in relation to the 65,000 threshold – are more likely to live in rural areas. For instance, complier households are 24% more likely to own livestock, 22% more likely to have land for agriculture, 59% more likely to own a car or tractor, and 39% more likely to own a cattleshed or granary. These households are also relatively poorer in terms of baseline earned income
– complier households are 10% more likely to fall in each of the lower two baseline income terciles, and consequently 20% less likely to fall in the upper income terciles. This finding is particularly useful from the perspective of optimal policy design, as it shines a light on the type of households who respond to the specific design of the welfare program we study. We supplement our characterization of compliers by providing partial correlations of key household characteristics with baseline income. This means we can simultaneously examine the relative complier likelihood for a given characteristic, as well as the partial correlation of this characteristic with income. This provides a richer sense of the characteristics that relate to complier status.

We next consider the way in which households manipulate their PMT scores at a repeat interview. We document large and statistically significant falls in the probability of having rural assets, notably livestock. When combined with our work characterizing manipulating household, which highlights that complier households are more likely to be based in rural areas, the evidence here suggests that selling, or misreporting rural assets, is likely a key strategy for welfare score manipulation. This stands in contrasts to harder to move/misreport assets, such as land, properties, or the quality of the properties, for which we do not find changes. We also document evidence that households may change their reporting of the proportion of the property occupied by household members.

We then present evidence on the consequences of welfare manipulation. Our FDD results show that women in households that engage in welfare manipulation work more in the formal labor market, but their income, and the household total income, do not increase significantly. When we split households by those with successful and unsuccessful manipulation attempts, we find that it is women in households with unsuccessful manipulation attempts that drive the increase in labor market participation. We do not find such effects for women in households where the manipulation attempt is successful, pointing to a crowding out effect of welfare income on labor market engagement. We additionally study the expenditure response to welfare manipulation. A key element of our household survey involved collecting detailed household expenditure data. The FDD estimates for total expenditure are positive, but imprecisely estimated. Where we find the largest increases in expenditure is on children – this total expenditure figure comprises increases in child clothing. We find no effects on food expenditure, including eating out, nor do we find any effect of increased expenditure on alcohol and tobacco.

Given that households that manipulate their welfare eligibility status spend their additional income almost exclusively on children, we focus our attention for the remainder of the paper on the outcomes of children and young people in the household. Our setting is interesting from the perspective of childhood skill investment in that, as a consequence of manipulation, there are two countervailing forces present at the household level. Manipulating households spend more on children, yet women in the household work more. There is an active literature focusing on the child consequences of us such “time versus money” trade-offs (Caucutt et al., 2020; Agostinelli and Sorrenti, 2021; Nicoletti et al., 2023; Mullins, 2022).

We combine our administrative health data and survey data to investigate changes in early childhood investment. These investments take the form of health and time investments. We find no changes in vaccination rates of children aged 0-5. Using our survey data, we do not find evidence of drops in the number of health check-ups. Whilst we have information on child-related time use of parents, the IV estimates for these outcomes are imprecisely estimated, and we run in to issues with the strength of our instrument.[2] We additionally focus on later child investments, in the form of high school and university attendance of older children and young adults in the household. We do not find any changes in post-compulsory high school attendance, neither we do find increases in university attendance for 18-23 years old living in the household.

In concluding the paper, we present two exercises. The first is to present a conservative lower bound estimate of the cost of welfare manipulation to the state coffers. Welfare eligibility manipulation is costly, amounting to an additional 25% of the initial welfare expenditure on our target group of welfare recipient households. The second exercise is to compare our estimates of welfare score manipulation – based on a regression discontinuity approach – with estimates from two flavors of bunching estimators. Our prevalence estimates are as good as identical to those from bunching estimators.

Our work contributes to three strands of literature. First, we contribute to the literature on optimal public policy targeting, with a specific focus on the demand-side manipulation. Our main contribution to this literature, given our unique data and our empirical approach, is that we can work more directly with welfare manipulators. Using the reasons for a repeat assessment, we can isolate the key margin on which manipulation occurs. Based on our characterization of compliers, we have a considerably better sense of observable profile of households susceptible to manipulation, based on their quasi-random allocation of a proxy means score. Previous work in this area estimates counterfactual distributions around a cutoff using the size of the bunching/hole following Saez (2010); Chetty et al. (2011). Additionally, Gelber et al. (2020) introduced a method to estimate bounds when there is manipulation of the score in RDD, which is used in work including that of Deshpande et al. (2021); Howell (2022); Britto et al. (2022). Miller et al. (2013) estimates the implicit un-manipulated score to estimate the effect of access to subsidized health. In order to reconcile our approach with the current literature we compare our estimates of manipulation with the resulting estimates using the bunching estimations (Chetty et al., 2011; Zwiers, 2021) show they are consistent.

Second, we add to the academic and public debate on household responses to cash transfers (see the discussion in Banerjee et al., 2017), notably on whether such transfers have negative or positive effects on key outcomes. Targeted welfare may lead to welfare traps, whereby households are discouraged from labor market participation and making productive investments in order to keep receiving benefits.[3] Conversely, cash transfers can relieve households' liquidity constraints, thereby allowing them to search for better jobs or to invest in their children's education or productive ventures (for example Gertler et al., 2012; Carneiro et al., 2021). Even though we do not find negative effects from receiving the cash transfer it self, the program design, creates incentives to manipulate which leads to inefficient allocation of governmental funds. The lower bound estimate of the cost of manipulation that we provide at the end of the paper highlights that the cost of manipulation is substantial.

Finally, we add to the childhood skill investment literature investigating the return of different types of parental inputs at different stages of childhood (Cunha and Heckman, 2007; Caucutt et al., 2020; Agostinelli and Sorrenti, 2021; Nicoletti et al., 2023; Mullins, 2022). Making full use of both our adminis- trative health and education data, as well as our rich survey data on time use and health investments, we are able to document the impact on children and young people in the household of a setting where adults within the household have more available income, but less available time. Despite not finding statistically significant changes in children and youth outcomes, this is valuable, as much of the work that considers the competing roles of parental time versus income investments do so within the context of developed economies.

[2] In section A.2.3, we document very clearly that these instrument strength issues are a consequence of the reduced sample size with which we have to work when using our survey data – the first stage estimate is remarkably constant across all our data settings.

[3] Although theoretically possible (Banerjee et al., 2017), there is no empirical evidence of such a response.

2 Institutional Framework and Data

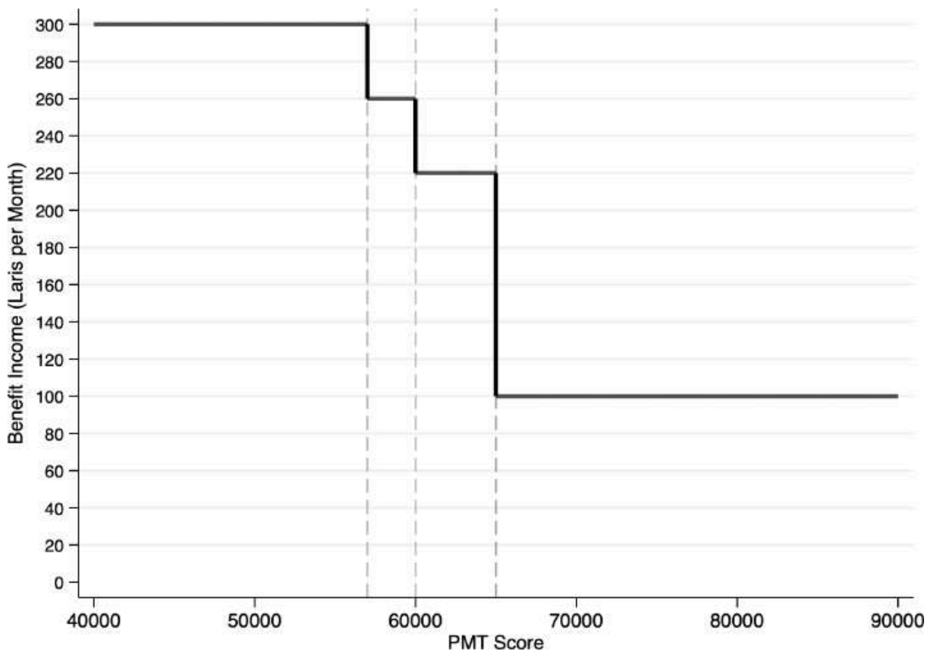
2.1 Georgia's TSA Program

In 2008, Georgia faced a deep crisis due to both the effects of the international financial crisis and the conflict with the Russian Federation in Ossetia. In response to the social consequences of this crisis, the government commenced the Targeted Social Assistance program (TSA) as part of the social safety net (World Bank, 2018). The objective of the program was to alleviate poverty by direct cash transfers to households for a country where over one third of the population lied below the poverty line. The TSA management is in the hands of the Social Service Agency of Georgia (SSA).

In 2015, the SSA introduced major changes to the TSA program. First, the Agency commenced targeting using a Proxy Means Test (PMT).[4] To do so, the Agency interviewed all households registered in the United Database for Socially Unprotected Families (UDSUF).[5] The PMT measures households welfare using data on income, consumption, expenditure, assets, and household composition.

Second, the TSA allocates benefits decreasing gradually as the PMT score increases. Third, by recommendation of UNICEF, the TSA introduced an additional benefit per child. Initially the benefit was 10 Lari per child month, but from January 2019 this increased to 50 Lari per child per month. Figure 1 provides the TSA benefit schedule for the sample median household composition. Table A1 in Appendix A provides a full summary of the TSA benefit scheme. It is important to point out that after a PMT score of 65,000, the reduction in benefits is considerably larger than the reduction in benefits in previous cutoffs.

Figure 1: The TSA benefit schedule is a Stepwise Function, With a Large Change at 65,000



Notes: Benefit Income-PMT score schedule for the sample median household structure of two adults, two children. See Table A1 in Appendix A for the full schedule.

After a household is assessed by the SSA and receives a score, they receive a monthly benefit based on their household composition and PMT score. However, households may be reassessed for various reasons. For example, changes in household composition, changes in income (observed by the SSA), or changes in

[4]The PMT formula was approved by the Resolution No. 758 (December 31, 2014) of the Government of Georgia

[5]From 2008, every household who wished to apply for receive social benefits was registered in this database.

household's location.[6] In addition, if a household feels that their PMT score does not accurately represent their welfare, they can request an additional interview after one year of being assessed. In each case, the SSA will re-interview the household and calculate a new score, which may be larger or smaller than the original score, and will adjust the benefits accordingly.

2.2 Data

We combine multiple sources of administrative data with information from a household survey that we conducted on a block-random[7] subset of households. Our core data is the universe of all PMT interviews conducted by the SSA from April 2015 – the start of the new TSA regime – to June 2019. This allows us to track every interaction a household has with the SSA and the benefits they receive since 2015. The PMT interview data contains new entrants in the welfare system, as well as existing welfare recipients, who were interviewed in order to calculate their PMT score.

For households with multiple interviews, we also observe the reason why an additional reassessment occurred. Using this information, we are able to observe if a reassessment was initiated by the household in the form of a request for a repeat interview, or was automatically triggered due to a change in the demographic or economic situation of the household.

Once a household receives a new PMT score, the previous score is annulled by the SSA. The SSA may also cancel the welfare payments to a household if (i.) the Agency finds out that the household hid changes that could alter their PMT score or cheated in any other way, or (ii.) if the household refuses a reassessment. In these cases the PMT score allocated to the household is annulled. We observe the PMT score status for each household-interview couplet.

We match households in the PMT interview database to three other administrative data sets. First, for every adult aged 18 to 64 years old we match in labor income and labor market participation infor- mation from the Revenue Service database from the Ministry of Finance. This covers only the formal sector. Using this data we can observe the extensive margin of formal labor supply, and the associated income with this job. We observe this information at four points in time – August 2018, February and August 2019, and February 2020.[8]

Second, we use administrative data from the Ministry of Education on school attendance for children aged 5 to 18 years old. We observe in which grade they enroll in September 2017 to September 2019. Primary and secondary education in Georgia is free and compulsory (grades 1 to 10). At 16, teens are expected to enroll in high-school for grades 11 and 12, where school is still free but no longer compulsory. In addition, we have information on college attendance and college graduation for individuals aged 16 to 25 years old still living within the household.

Third, we use information from the Ministry of Health regarding vaccinations for children within the household.

We supplement the wealth of administrative data with a household survey conducted in the Fall of 2019, which surveyed a random sample of 7,392 households with children in 46 municipalities. The survey includes information about income, expenditure, labor market participation (in both the formal and informal sectors), education, health and childcare.[9]

2.3 Sample Selection

We focus on households who have children when initially assessed by the SSA as a matter of internal consistency – the household survey we conducted only interviewed households with children. The analysis

[6]The SSA has access to data from different governmental sources in order to follow the TSA beneficiaries. For example, births, deaths, children dropping out school, increases in formal labor market income, disability claims.

[7]We randomly surveyed households at specific parts of the PMT score distribution – these are the "blocks".

[8]In our analysis, we only use formal labor market tranches of the data that fall after the last observed interview of the household.

[9]For more details of the questionnaire and sample selection in Econometrı́a (2020)

we present in the paper focuses on the structure of the welfare payment system in Georgia, specifically the discrete cutoffs in the welfare payment-PMT score schedule, which creates incentives for PMT score manipulation. We focus our analysis on the 65,000 cutoff because as shown in Figure 1, the changes in benefits is larger there. This, along with the fact that households can apply for reassessments result, as we will show shortly, 65,000 is the only cutoff where we find clear evidence of PMT score manipulation. For this reason, we use households with an initial PMT score in the range 60,000 - 70,000. In addition, in order to have a groups of households where benefits change but not the probability of manipulation, we add to the sample those whose initial PMT score is around the 57,000 cutoff (from 54,000 to 60,000). These households are key to disentangle the effect of score manipulation as we will explain in the following section.

Furthermore, additional interviews play a key role in our analysis. For this reason we exclude house- holds whose first interview was after December 31st 2017, to allow that all the households in our sample have the opportunity to request a second interview, within the time frame for which we have all necessary data. Given that we have detailed information on the *reason* for a repeat interview, we omit all households with more than one interview whose repeated interview was not triggered by the household, but rather triggered by the SSA. The purpose of this sample restriction is to avoid conflating manipulation with a random demographic or labor market shock. Finally, we exclude households receiving *Internal Displaced People (IDP)* benefits the first time they were interviewed as these households receive a different set of benefits from the SSA.

Our final administrative data sample, once we apply all relevant sample selection restrictions, contains 11,972 households. Our final survey data sample contains 1,682 households.[10] Table 1 summarizes the main characteristics of our sample of analysis.

In our sample, 18% of households request an additional interview. Over the course of time that we observe these households, they average 2.8 interviews. Many households are multi-generational, with an average of 5 household members – 3 adults and 2 children. Most children attend school, only 7.3% of households have at least one child not attending. The households in our working sample are poor – baseline income for these households is 288 Lari (about 100 USD). 62% of households in our working sample own some form of estate (for example, garage, additional housing), 61% have agricultural land, and 41% have some livestock.

3 Evidence of Manipulation

We start by providing initial evidence of welfare manipulation, in order to motivate both the theoretical model and the empirical specification that follows. To do so, we consider a wider range of PMT scores than used for our main analysis – specifically 40,000-90,000. We first present the distribution of initial PMT scores in Figure 2(a). A visual inspection suggests that the distribution is smooth and continuous through the 57, 60 and 65 thousand cutoffs. This is confirmed by the associated *p*-values from a CJM density test (Cattaneo et al., 2020). In Figure 2(b) we present the analogous figure for the final PMT score distribution of households, allowing for reassessments. The difference between the two distributions is stark. There is clear visual evidence of unnatural bunching of households to the left of the 65,000 threshold, and a large discontinuity precisely at 65,000. The CJM density test confirms the presence of manipulation, with a *p*-value of .000. In contrast, for the 57,000 and 60,000 cutoffs the distribution remains continuous.

[10]The fact that the survey data is an order of magnitude smaller than the administrative data leads to sample sized-based power issues in some of our later analysis. Throughout our empirical work we balance the size and accuracy of the administrative data, with the richness of the survey data.

Table 1: Summary Statistics

	Mean	Standard Deviation
Household-Initiated Repeat Interview	.181	.385
Number of interviews[1]	2.85	.99
Household Composition		
Household size	4.54	1.52
Adults	2.7	1.2
Children	1.84	.852
Child Not in School	.073	.26
Pensioner in the Household	.0352	.184
Household Head Characteristics		
Age	50.4	15.2
Female	.391	.488
Single Mother	.0333	.179
Income and Expenditure		
Total income (Lari per Month)	288	274
Utility Bills (Lari per Month)	18.9	15.1
Housing Characteristics		
Number of Rooms	3.24	1.52
Good Quality Floor	.774	.418
Assets		
Owns any Estate	.618	.486
Owns a Car or Tractor	.0399	.196
Owns Agricultural Land	.612	.487
Owns any Livestock	.408	.491
Observations	11,972	

Notes: [1] Conditional on requesting at least one additional interview. Household characteristics as measured at the time of the initial interview. Data source: PMT Interview Data.

Households may be assessed multiple times for a variety of reasons and reassessments may be initiated by both households and the SSA. In Figure 3 we present evidence that is highly consistent with household-initiated requests for PMT score reassessment being the key driver of the discontinuity we document in Figure 2(b). We start by presenting Figure 3(a), which shows the unconditional probability that a household will have multiple interviews in the period of analysis. The probability jumps by approximately 20 percentage points, or just under 60%, precisely at 65,000. We then make use of the rich administrative data we have available, and separate between reasons for a reassessment. We plot the probability of a household initiated and non-household initiated reassessments respectively in Figure 3(b) and Figure 3(c). The discontinuity at 65,000 is driven solely by household-initiated reassessment requests.

Following this evidence, in the following section we present a model in which we explain manipulation as the result of households optimally choosing whether or not to request an additional interview.

4 Modelling Welfare Eligibility Manipulation

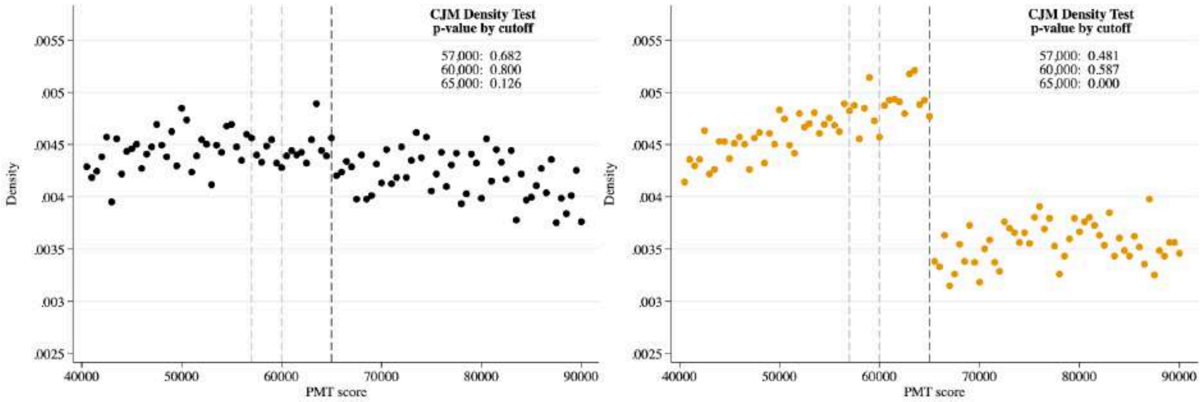
4.1 A Becker Model of Manipulation

We model welfare eligibility manipulation – here the decision to request a repeat interview – through the lens of the Becker-Ehrlich model (Becker, 1968; Ehrlich, 1973). According to this approach, an individual will choose to engage in welfare eligibility manipulation if the expected value of manipulation (V_R) exceeds that of accepting their initial benefit level (V_A):

$$E(V_R) > E(V_A). \tag{1}$$

When requesting a repeat interview, the individual may receive a higher benefit level B^+ with exoge-

Figure 2: The PMT distribution for the last interview is not continuous around 65,000.



(a) First Interview (b) Final Interview

Notes: Bin size of 500. Panel (a) shows the distribution of the PMT score for the first interview each household had. Panel (b) shows the distribution of the last PMT score each household received. The box in both figures contains CJM Density Test *p*-value from the Cattaneo et al. (2020) manipulation test using households with scores between the cutoff above and below each cutoff in the estimation, a polynomial of order 2, and data driven bandwidths, around each cutoff.

nously determined probability p, or may receive the same benefit level as their initial allocation B^0. The cost of requesting a repeat interview is C. This cost captures the administrative and time cost of request- ing a repeat interview, as well as the time cost involved in the repeat interview itself. With probability q the SSA discovers that the individual is falsifying information and imposes a sanction – a suspension of welfare payment for at least one year. B^- is the expected value of potential sanctions including any additional costs the individual may face, for example, loss of social capital due to engaging in welfare fraud (Williams and Sickles, 2002), or debt-related issues such as high interest payments, if households expect they may fall behind on bills or other payments if sanctioned. Combining these factors, we can write an expression for the expected utility of requesting a reassessment – $E(V_R)$:

$$E(V_R) = p(1 - q)U(B^+) + (1 - p)(1 - q)U(B^0) + qU(B^-) - C. \tag{2}$$

The expected value of accepting the initial PMT score is considerably simpler:

$$E(V_A) = U(B^0). \tag{3}$$

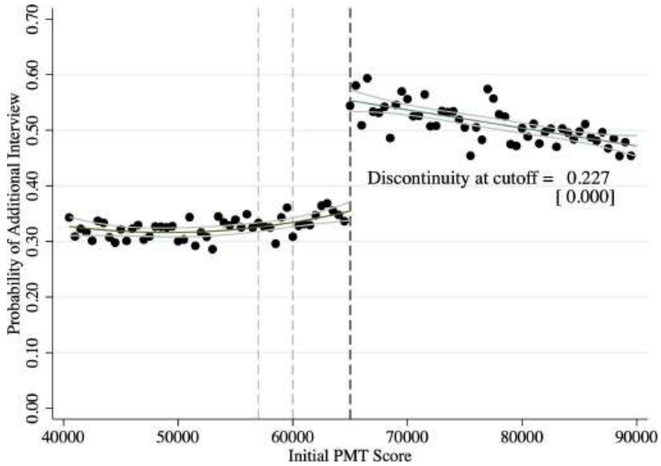
Equating $E(V_R)$ and $E(V_A)$ allows us to characterize the point at which an individual is indifferent between requesting a repeat interview and accepting their initial PMT score:

$$p(1 - q)U(B^+) + (1 - p)(1 - q)U(B^0) + qU(B^-) - C = U(B^0). \tag{4}$$

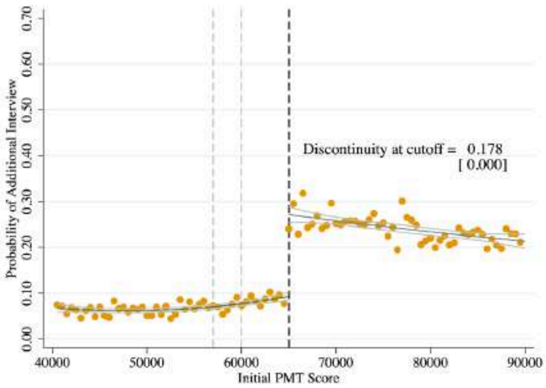
By defining the possible utility gain of requesting an additional interview as $\Delta U^+ = U(B^+) - U(B^0)$, the possible utility loss of an additional interview as $\Delta U^- = U(B^0) - U(B^-)$, and rearranging yields, we know that a given household will request a reassessment if:

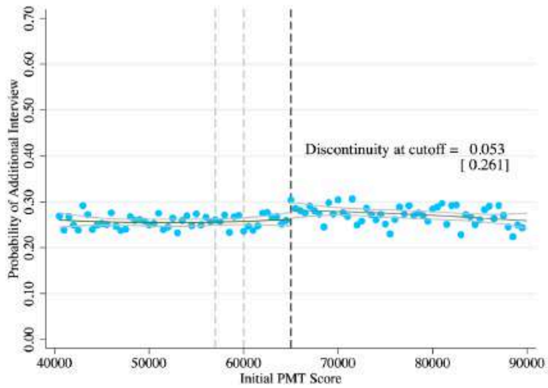
$$p(1 - q)\Delta U^+ - q\Delta U^- > C. \tag{5}$$

Figure 3: The Probability of an Additional Interview Jumps at 65,000 – an Effect Driven by Household-Initiated Reassessment Requests.



(a) All Cases





(b) Household-Initiated Reassessment (c) Other Reason for Reassessment

Notes: Each figure shows the probability of having an additional interview by the first PMT score each household obtained. Panel b plots the probability that at least one additional interview was asked by the household. Panel c plot the probability that all PMT reassessments were initiated by the SSA. We include in each figure the resulting RD estimate and p-value in brackets, following Calonico et al. (2014).

4.2 Requesting a Reassessment and Empirical Specification.

We now map our theoretical model onto a specification that we will estimate with our data. For a given household i, ΔU^+ and ΔU^- are functions of the household's initial PMT score, $z_{0,i}$, and some limited household characteristics H_i, which determine B_0. Thus, the left hand side of Equation (5) can be written as:

$$p(1 - q)\Delta U^+ - q\Delta U^- = f(H_i, z_{0,i})$$

C_i is a function of a broader set of observable variables, X_i, which encompasses H_i, and an unobservable component μ_i, which captures household-level tendency towards welfare eligibility manipulation:

$$C_i = k(X_i) - \mu_i$$

We denote $R_i = 1$ when a household requests a repeat interview. From Equation (5), a reassessment occurs when $p(1 - q)\Delta U^+ - q\Delta U^- - C > 0$. Assuming $f()$ and $k()$ are linear in X_i, and given that B_0, B^+ and B^- depend on $z_{0,i}$ and a cutoff k, we write down a latent variable model for requesting a repeat interview:

$$R_i^* = X'\beta + g(z_{0,i}, k) + \mu_i \quad (6a)$$

$$R_i = 1 \quad \text{if } R_i^* > 0 \quad (6b)$$

$$R_i = 0 \quad \text{if } R_i^* \leq 0 \quad (6c)$$

Welfare Eligibility Manipulation Attempts. From Equation 6a, we estimate the impact of welfare eligibility manipulation on a series of outcomes starting with a FRD-IV approach in the spirit of Lee and Lemieux (2010). The first stage can be expressed as:

$$R_i = \alpha_0 D_i + g_1^D(z_{0,i}) + X_i' \alpha_1 + \mu_i \quad (7)$$

where $D_i = 1[z_{0,i} > 65,000]$ and $g_1^D(z_{0,i})$ is a function of $z_{0,i}$ above and below the cutoff. Following the FRD-IV literature, the instrument in Equation (7) is D_i.[11] To measure how a welfare eligibility manipulation attempt affects a given outcome Y, for example, labor market participation, the second stage equation of the system is:

$$Y_i = \beta_0 R_i + g_2^D(z_{0,i}) + X_i' \beta_1 + \eta_i \quad (8)$$

Therefore, so long as μ_i is not correlated with D_i, α_0 captures the change in the probability of requesting an additional interview because the initial score was just above 65,000, and β_0 captures the effect of requesting a reassessment on Y.

However, as previously explained, TSA benefits also change around each cutoff. Thus, at 65,000, two variables that affect final outcomes jump, the probability of requesting an additional interview (R_i) and the initial benefit each households (B_0). Thus, we know that in equation 8, $\eta_i = \beta_2 B_{0,i} + \phi_i$. Then, the second stage is:

$$Y_i = \beta_0 R_i + g_2^D(z_{0,i}) + X_i' \beta_1 + \beta_2 B_{0,i} + \phi_i \quad (9)$$

For this reason we use variation around other cutoff, using a difference-in-discontinuities design (Grembi et al., 2016; Millán-Quijano, 2020), in order to disentangle the effect of manipulation attempts from the effect of initial benefits. The approach we take in this work, which we outline below, involves an instrumental variables approach to a difference-in-discontinuities design, which we refer to as a fuzzy difference-in-discontinuities (FDD) design.

We add to our analysis sample households with initial PMT scores just above and just below 57,000. Around that cutoff we do not find evidence of manipulation (see Figure 2), nor do we document any changes in the probability of requesting an additional interview (Figure 3). Instead, initial benefits drop by 10 Lari per person per month as shown in Figure 1. Following the difference-in-discontinuities literature, we define a dummy A_i that takes the value of one (1) for households whose first PMT is in the area around 65,000 ($A_i = 1[60,000 < z_{0,i} \leq 70,000]$). We also redefine D_i, as it now takes the value of 1 for households above their respective cutoff. Thus, $D_i = 1[z_{i,0} > 57,000 \ \& \ A_i = 0]$ or $D_i = 1[z_{i,0} > 65,000 \ \& \ A_i = 1]$. Then, we can write

$$\begin{aligned}
R_i &= \omega_1 D_i + \omega_2 A_i + \omega_3 A_i \times D_i + g_R^{D,A}(z_{0,i}) + X_i' \omega + \mu_{R,i} && (10a)\\
B_{0,i} &= \gamma_1 D_i + \gamma_2 A_i + \gamma_3 A_i \times D_i + g_B^{D,A}(z_{0,i}) + X_i' \gamma + \mu_{B,i} && (10b)\\
Y_i &= \theta_R R_i + \theta_B B_{0,i} + \theta_3 A_i + g_Y^{D,A}(z_{0,i}) + X_i' \theta + \mu_{Y,i} && (10c)
\end{aligned}$$

The latest system of equations can be also understood as an Instrumental Variable system with two

[11] In order to bridge between Equations (6a) and (7) note that we parameterize $g_1(z_{0,i}, k)$ as $\alpha_0 D_i + g_1^D(z_{0,i})$.

endogenous variables (R, B_0), and two instruments (D, $A \times D$). In this case, θ_R identifies the effect of requesting a reassessment on the outcome variable if the following assumptions are fulfilled.[12] First, the core RDD assumptions around both cutoffs. Figure 2(a) shows that $z_{0,i}$ is continuous through the cutoffs at 57,000 and 65,000, which means that D_i is randomly assigned. In addition, Figure A1 in Appendix A.2 shows that the RDD continuity assumption holds for a large set of observable variables X, using information from the households' initial PMT interviews.

We may still be concerned about the role household unobservables play in the reassessment process. We address this concern by providing direct evidence that there are no discontinuities at the thresholds for the two key dimensions governing the success of a manipulation attempt – the probability that a repeat interview will reduce a household score, and the probability of being caught and sanctioned by the SSA. These are, respectively, the parameters p and q from the theoretical model we present in Section 4.1. Using detailed data about the final status of each household, we can plot these two probabilities against our running variable. We do so in Figure A2 in Appendix A.2, which shows that both p and q are continuous through the cutoff. The p-values that we present in the graphs are based on the null that there is no discontinuity at the threshold. The respective p-values for p and q are .80 and .36, thus our statistical tests confirm what a visual inspection of the figures tells us – p and q are continuous through the cutoff.

Finally, following Grembi et al. (2016) and Mill´an-Quijano (2020), we require two further assumptions to hold. In the remainder of this section, we provide detailed evidence to suggest that these assumption do hold in our setting, thereby enabling us to proceed with our FDD design.

First, we require that the effect of B_0 is constant across the two PMT areas: $A = 0$ and $A = 1$. We use two cutoffs close to each other, thus, after controlling by A, assuming that $E(\theta_B|A = 0) = E(\theta_B|A = 1)$ is plausible, as θ_B represents the effect of one additional Lari. Figure A3 in Appendix A.2.2 shows that the impact of an additional Lari is not statistically different for a set of labor market outcome around the three cutoffs where we do not find evidence of manipulation (around 30,000, 57,000 and 60,000). One potential concern is that the change in benefits around 65,000 is three times the change in benefits at 57,000. However, the scale of the jump in benefits is taken into account by γ_1 and γ_3 in equation 10c. Figure A5 in Appendix A.2 shows that $\gamma_1 + \gamma_3 = 3\gamma_1$ regardless of the estimation sample.

Second, in order to isolate the effect of R, the change in manipulation attempts only happens around one of the cutoffs, in our case around 65,000. We already show evidence that manipulation only occurs at 65,000, as the distribution of the final PMT is continuous around 57,000 (Figure 2(b)), and manipulation attempts are also continuous around 57,000 (Figure 3(b)).

5 Results

In Section 3 we provide evidence that given (i) the structure of the benefit scheme we study and (ii) the availability of household-initiated reassessment creates the incentives to manipulate welfare eligibility. In this section we seek to answer three related questions regarding welfare eligibility manipulation. First, *what type* of households attempt to manipulate their welfare eligibility status? Second, *how* do these household manipulate their score? Third, what are the down-stream consequences of welfare eligibility manipulation?

5.1 Who are the Compliers?

In this section we consider the compliers in our 2SLS framework – households whose manipulation status is induced by falling above/below the 65,000 PMT score threshold. Whilst we cannot directly identify the compliant sub-population, we can characterize these households. To do so, we follow the approaches of Abadie (2003) and Dahl et al. (2014) in characterizing compliers. Our target statistic is the complier

[12]See details in Appendix B

relative likelihood of having a given Bernoulli-distributed characteristic, x_{1i}, which we express as $P[x_{1i} = 1 \mid R_{1i} > R_{0i}] / P[x_{1i} = 1]$. For continuous characteristics, we binarize the variable.[13] R_{1i} and R_{0i} denote the potential outcomes of R_i when $A_i \times D_i = 1$ and $A_i \times D_i = 0$ respectively. We present a series of complier relative likelihoods in Figure 4(a).

Complier households are more likely to be headed by a woman, and have an older, slightly more educated head of household than average. These households are larger in terms of total size and number of children, are more likely to have a household member with a health condition and are less likely to have a single mother present. Compliers appear to live in more rural settings, as they are more likely to own a workshop, granary or cattleshed, to have agricultural land, to own both livestock, and a car or tractor. Also, they are more likely to have a garage.

Finally, we document that complier households are poorer than the average household in the PMT score range of 60,000-70,000 – they are more likely to fall in the lower two terciles, and much less likely to be in the upper tercile of baseline income.

In Figure 4(b) we present estimates for each characteristic from a regression where the dependent variable is household earned income at baseline.[14] Combining the information in Figure 4(a) and 4(b), we can better understand the baseline economic status of the compliant households – 4(b) informs us of the partial correlation between a given characteristic and baseline income, whilst Figure 4(a) informs us of the relative likelihood a complier household will have the characteristic. With the exception that complier households are typically larger than average, all other characteristics of these households are correlated with lower economic status at baseline, particularly those related to the more rural setting in which complier households appear to be based.

5.2 How do Households Manipulate Their PMT Scores?

We next aim to understand how households manipulate their PMT scores, in order to achieve a lower score and thus receive higher benefit income. To do so, we exploit the richness of our administrative data, which includes each and every input into the combined score that yields a PMT score. To be clear, we are not privy to the factor loadings used to generate the PMT score, but we do know, and have access to, each of the individual inputs. Accordingly, when we present the results from this section of analysis, we consider a much wider array of variables than we do elsewhere in this paper.

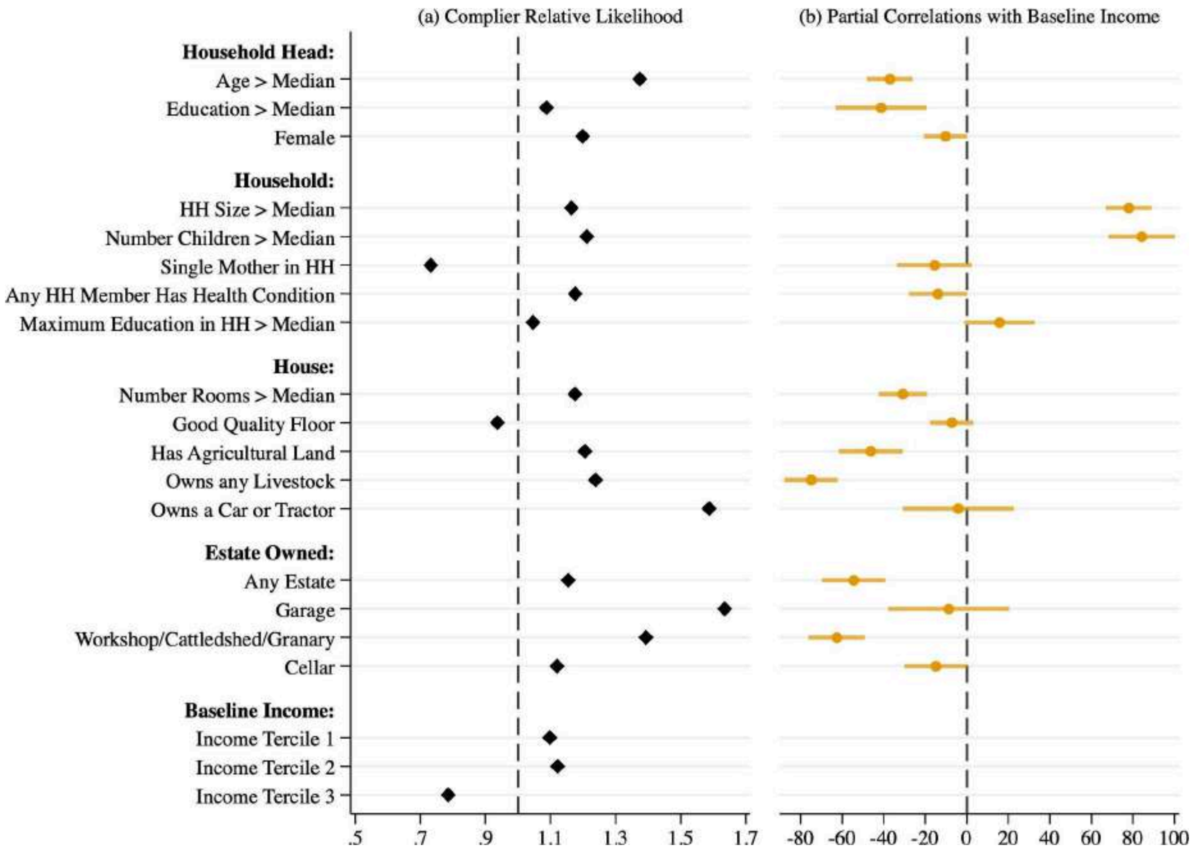
We focus in households an initial PMT score just above the 65,000 cutoff, specifically scores between 65,000 and 70,000. We constrain our working sample here to households with multiple interviews, either SSA- or household-initiated. We then calculate the change in each of the PMT score input components between the last and the first interview each household has. We use the SSA-initiated interviews to establish a baseline for changes that are natural between interviews, or that are likely to come from random shocks. The aim of this exercise is to identify which variables are more likely to change between interviews for households with household-initiated compared to SSA-initiated re-interviews. Given this approach, we can interpret the resulting differences across the two groups over time as difference-in- differences estimates.

Figure 5 shows the results of this analysis. Panel (a) the average change in each component for both household- and SSA-initiated reassessments, i.e., the group-based differences. One can see that many components change similarly for both types of reassessments, while for other components the average change after a new interview is different when the source of the request for re-interview comes from the household, not the SSA. In Panel (b) we present the difference-in-differences estimates. Clear patterns emerge, which provide suggestive evidence of how households that engage in welfare eligibility

[13]We calculate the relative likelihood using Bayes' Rule and by taking the ratio of the first-stage coefficient for the sub-group with $x_{1i} = 1$ divided by the first stage coefficient for the full sample, $P[R_{1i} > R_{0i} \mid x_{1i} = 1] / P[R_{1i} > R_{0i}]$.

[14]We condition on a common set of household-level control variables (our baseline covariates, described below), region-by-quarter and interview time fixed effects.

Figure 4: Complier Households Live in More Rural Areas and are Relatively Poorer



Notes: *Panel (A)* – We characterize compliers by presenting the ratio of the first stage coefficient on the instrument for each binary (or binarized) characteristic to the overall first stage coefficient. By Bayes' rule this ratio of first stage estimates – which we can express as $P\left[R_{1i} > R_{0i} \mid x_{1i} = 1\right] / P\left[R_{1i} > R_{0i}\right]$ – yields the complier relative likelihood of a given characteristic, $P\left[x_{1i} = 1 \mid R_{1i} > R_{0i}\right] / P\left[x_{1i} = 1\right]$. PMT range: 60,000-70,000. *Panel (B)* – We report the coefficient and 95% confidence interval for each characteristic from a regression where the dependent variable is household (own) income at baseline. We condition on a common set of household-level control variables (our baseline covariates, described below), region-by-quarter and interview time fixed effects. The estimation sample is based on a PMT range of 55,000-60,000 and 70,000-75,000, i.e., bands of 5,000 on either side of our range of interest.

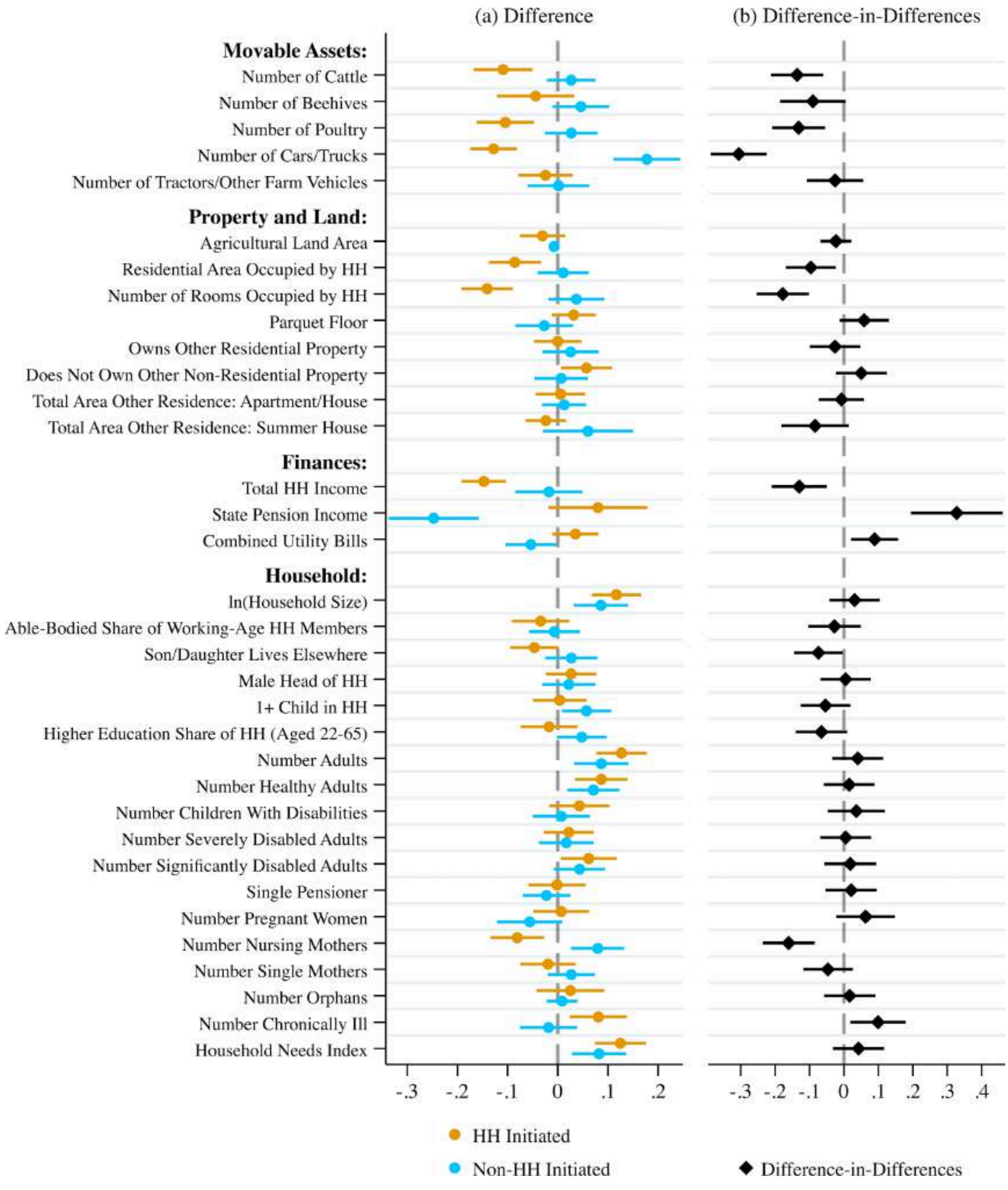
manipulation do so in practice. First, the probability of having rural assets decrease, notably livestock. When combined with the evidence we provide in Section 5.1, which highlights that complier households are more likely to be based in rural areas, the evidence here suggests that selling, or misreporting rural assets, is likely a key strategy for welfare score manipulation. This stands in contrasts to harder to move/misreport assets, such as land, properties, or the quality of the properties, for which we do not find changes. We also document evidence that households may change their reporting of the proportion of the property occupied by household members.[15]

5.3 The Labor Market and Welfare Eligibility Manipulation

In order to understand the labor market consequences of welfare eligibility manipulation, we make use of two data sources. First, administrative data on formal labor market activity and earnings from four

[15]Note that we find DD estimates statistically significantly different from zero in several other cases that are generated by SSA-initiated repeat interviews, which we would expect – if households situations change signifi- cantly, this automatically triggers a new (SSA-initiated) interview. Examples of this include the purchase of a new car, or a change of the pension income received by the household.

Figure 5: Reductions in movable agricultural assets, house reported size, and salaries, are the ways in which possible manipulators try to reduce their PMT score.



Notes: *Panel (a)* – We plot the average difference in the value of each variable from the last to the first interview for households with more than one interview and whose first interview PMT score is between 65,000 and 70,000. The line represents the 95% confidence interval. *Panel (b)* – We plot the average difference between household-initiated and SSA- initiated in the change between the last and the first interview. The line represents the 95% confidence interval.

periods of time. We supplement this administrative data with survey data that contains information on both formal and informal labor market activity. The combined use of both data sources permits us to capture a broad and comprehensive view of the labor market consequences of welfare eligibility manipulation.

We first consider formal labor market outcomes. In Table 2 we provide evidence of the labor market consequences of attempted welfare eligibility manipulation. Although the welfare eligibility manipulation occurs at the household level, the heterogeneity in formal labor market responses by gender means it is instructive to consider the results by gender. As we note above, we only consider labor market outcomes that occur after the households' final interview, thus we interpret the results in this section as the downstream consequences of welfare eligibility manipulation.

We start with men. The OLS results indicate a negative relationship between a welfare eligibility manipulation attempt and labor market outcomes. This suggests that the act of manipulation could be driven by need – those who attempt to manipulate their eligibility have lower formal labor market income and are less likely to be employed at least once – our least stringent measure of labor force attachment. This interpretation of the negative coefficient is consistent with what we document for household income in Figure 4(a). An alternative explanation for the negative OLS estimates is selection bias – those that attempt to manipulate welfare eligibility have unobservables negatively correlated with labor market outcomes. Comparing the OLS estimate of an eligibility manipulation attempt to the mean of the outcome variable for those who did not ask for a reassessment with a first PMT score below the 65,000 cutoff (Y_0 at the base of the table), we see those attempting to manipulate their scores earn 21% less in the formal labor market and are 10% less likely to be minimally attached to the formal labor force. Given that our 2SLS estimates reflect a local average treatment effect (LATE) based on the compliant sub-population, and deal with the problem of endogenous welfare eligibility manipulation attempts, we present complier-reweighted OLS estimates to bridge between our OLS and 2SLS estimates (Bhuller et al., 2020). The complier-reweighted OLS estimates are broadly in line with our main OLS estimates, which suggests that treatment effect heterogeneity is unlikely to be a primary concern when interpreting our 2SLS estimates.

Finally, we turn to the 2SLS estimates and document that the impact of a manipulation attempt are still negative but statistically insignificant. Hence, once we use the exogenous variation in the probability to manipulate, we conclude that manipulation attempts do not lead to changes in the probability to work in the formal sector, or in the income earned.

In columns 4-6 of Table 2 we present analogous results for women. Both the OLS and complier- reweighted OLS estimates suggest little correlation between a household welfare eligibility manipulation attempt and labor market outcomes for women. The 2SLS estimates paint a very different picture however. In columns 4 and 5, we document a positive effect of a manipulation attempt on the labor force participation of women. They are statistically significantly more likely to work at least once, and in all periods, following a household manipulation attempt. This increased labor market attachment leads to higher income, although this effect is not statistically significantly different from zero.[16]

In Figure A6 in Appendix A.3, we provide evidence that the labor market response of households appears strategic. If households who receive welfare benefits increase their monthly earned income by more than 175 Lari, this income change automatically triggers a further, SSA-initiated repeat interview. Figure A6 highlights that income changes stay within these limits for the vast majority (93%) of households.

5.3.1 Does success in the manipulation attempt matter?

So far, we have documented the effect of a welfare manipulation *attempt* – proxied by household-initiated repeat interview – on labor market outcomes. In Figure A2 we show that approximately half of household-initiated repeat interview lead to a PMT score reduction sufficient to increase benefit income ($p(1-q)$

[16] At the household level, the results by gender lead to an increase in formal labor market participation but a null effect on total labor market income (see Table A2 in Appendix A.3).

Table 2: Welfare Eligibility Manipulation Leads to Increased Formal Labor Market Engagement for Women

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			**Women**		
	Employed At Least Once	Employed All Periods	Mean Income	Employed At Least Once	Employed All Periods	Mean Income
OLS						
Repeat Interview	–0.027**	0.009	–24.637***	–0.011	0.004	–2.920
	(0.012)	(0.007)	(7.264)	(0.010)	(0.006)	(3.348)
CW-OLS						
Repeat Interview	–0.010	0.012	–14.898*	–0.009	0.007	–1.344
	(0.013)	(0.008)	(7.774)	(0.010)	(0.006)	(3.578)
2SLS						
Repeat Interview	–0.112	0.202	–127.498	0.312*	0.257**	64.521
	(0.311)	(0.186)	(186.933)	(0.181)	(0.110)	(75.602)
SW F -Statistic: R.I.	20.435	20.435	20.435	44.620	44.620	44.620
Y_0	0.256	0.072	105.716	0.179	0.052	43.963
Observations	11,220	11,220	11,220	14,544	14,544	14,544

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Initial benefit is measured in 100s of Laris.

from our conceptual model). This observation gives rise to the possibility of treatment effect hetero- geneity, which relates to our labor market findings above. Do women in households that attempt to manipulate their welfare eligibility work more due to the removal of (non-labor) income-based constraint that previously prevent them from working? Such a response would rely on a credit-constraint channel of some form. Or is it the converse, i.e., manipulation attempts are needs driven, and a failed manipulation attempt leads women in the household to shift towards formal labor market participation?

To make progress on answering this question, we first briefly recap the ordering of events. All households receive an initial interview. Some household request an additional interview. We find that those just above the PMT threshold of 65,000 do so at a much higher rate, consistent with welfare eligibility manipulation. After a repeat interview is requested, the SSA visits the household to reassess their need and recalculates the household PMT score. The indicator p takes the value 1 if the household moves down a PMT score category, and therefore increases their benefit income, and takes value 0 otherwise. At the repeat assessment, the interviewer completes a 9-point checklist of questions to indicate the reliability of the information supplied by the household – this is the indicator variable q, which takes value 1 if the household is adjudged to have provided unreliable information and 0 otherwise. A value of $q = 1$ leads to all SSA benefits being suspended for at least 12 months. Accordingly we classify a manipulation attempt to be a success if $p(1 - q) = 1$ and unsuccessful otherwise.

In order to answer this question, we split our sample into two, non-mutually exclusive groups: first we consider households with only a single interview plus households with successful attempts. Second we consider the same group of single-interview households plus households with unsuccessful attempts. We present two pieces of evidence to support the validity of this approach. In Figure A2(c) in Appendix A.2 shows that the probability of success is continuous through the cutoffs, with a mean of approximately 45%. In Figure A7 in Appendix A.5 we present evidence in support of the continuity assumption, showing that observable variables are continuous at the cutoffs for the two split-samples. We then present the 2SLS results for formal labor market outcomes for the two split-samples in Table 3.

Once again, we find no impact for men, irrespective of the success of the manipulation attempt. When we turn to women, we find that it is women in households with unsuccessful manipulation attempts who drive our core labor market results that we document in Table 2. Following an unsuccessful manipulation attempt, women supply more labor in the formal sector. There is no statistically significant change in labor market engagement for women in households with successful manipulation attempts. This finding highlights an additional costs that welfare manipulation imposes upon the government – successful manipulation attempts *crowd out* labor market participation, reducing tax revenue, and perpetuating a cycle of reliance on the welfare system. As a final point, even though our estimates are imprecise, it is worth noting that the income of women increases by 120 Lari per month, which is equivalent to the expected TSA benefit increase after a manipulation attempt for a household of four members.[17]

Table 3: Successful Welfare Manipulation Attempts Crowd out Female Labor Force Participation

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			**Women**		
	Employed At Least Once	Employed All Periods	Mean Income	Employed At Least Once	Employed All Periods	Mean Income
(a) Unsuccessful Manipulation Attempts 2SLS						
Repeat Interview	–0.046	0.416	–38.626	0.468*	0.375**	119.573
	(0.547)	(0.342)	(328.086)	(0.255)	(0.156)	(106.478)
SW F-Statistic: R.I.	9.281	9.281	9.281	33.717	33.717	33.717
Y_0	0.261	0.071	110.120	0.183	0.055	47.250
Observations	10,434	10,434	10,434	13,485	13,485	13,485
(b) Successful Manipulation Attempts 2SLS						
Repeat Interview	–0.200	0.343	–134.008	0.316	0.296	8.906
	(0.441)	(0.264)	(264.137)	(0.358)	(0.210)	(147.866)
SW F-Statistic: R.I.	27.158	27.158	27.158	27.701	27.701	27.701
Y_0	0.261	0.071	110.120	0.183	0.055	47.250
Observations	10,130	10,130	10,130	13,244	13,244	13,244

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%.

5.4 Household Expenditure Responses to Welfare Eligibility Manipulation

In the previous section we provided evidence of the impact of welfare eligibility manipulation on labor market outcomes, documenting a positive effect of welfare eligibility manipulation on employment for women and no significant impacts for men. We provided further evidence highlighting the mediating role that the success of the manipulation attempt plays in driving these results. Ultimately, households that engage in welfare eligibility attempts are better off financially – successful manipulation attempts bring in more welfare income, and unsuccessful attempts end up yielding more labor income due to the labor market response of women to failed manipulation attempts.

Having taken stock of the evidence in the previous section, a natural question to ask is how do these welfare-manipulating households spend the extra income? In Table 4 we present expenditure patterns based on our survey data. The penultimate row of this table displays the outcome variable for those with an initial PMT score below the 65,000 cutoff who did not ask for a reassessment (Y_0). These statistics are particularly useful to gain a sense of expenditure patterns for a control set of households.

Both the OLS and complier-weighted OLS estimates are negative for almost every single expenditure group, and typically statistically indistinguishable from zero. The 2SLS estimates, however, tell a different story. We find that households target their additional income on expenditure on children. We find no changes in expenditure on other areas, such as expenditure on food, eating out, tobacco or alcohol. We document a 88 Lari increase in total expenditure on children, the lion's share of which is on clothing, and

[17] The analysis at the household level shows a similar pattern. Formal labor market income increases by 135 Lari in households were the manipulation attempt was unsuccessful. However, this estimate is not statistically significant (See Table A5 in Appendix A.5).

a smaller share on increased education spending. Although the effect for total expenditure is statistically insignificant, we can see that the increase in spending on children is approximately 90% of the total expenditure response. A different way to benchmark the increase in child spending is to use the baseline total expenditure (Column 1, penultimate row), in which case the increase in child spending is 21% of baseline expenditure.

We also estimate the effect of manipulation on expenditure patterns splitting the sample by the result of the manipulation attempt. However, in this case, our first stage estimate are weak and our estimators are too noisy to find significant differences.[18]

Summarizing the results, we find that a manipulation attempt leads to a significant increase in labor supply for women. In addition, in Table 4 we document that children are the primary beneficiaries of the corresponding increase in household spending. As our gaze turns now to child outcomes, we note that the evidence we document so far identifies two, countervailing forces on the child skill production function within households that attempt to manipulate their welfare eligibility. The increase in income, and concomitant expenditure on children, should have a positive impact on childhood skill production, whereas the fact that parents now have less time available will likely lead to a decrease in the production of childhood skills (Cunha and Heckman, 2007; Caucutt et al., 2020; Agostinelli and Sorrenti, 2021; Nicoletti et al., 2023; Mullins, 2022).

We investigate the impact of a household manipulation attempt on a battery of child outcomes for a variety of different ages, using administrative and survey data at the household and children level. For children 0 to 5 years old we check for possible effects of manipulation effects on health investments such as vaccinations and check ups, and parental investments measured in time spent with their children. We consider the impact of manipulation attempts on educational attendance for children 15 to 18, and for enrollment into tertiary education for young adults (18 to 23 years old). The details of these analyses are in Appendix A.4. We do not find any measurable and statistically significant changes in any children related outcomes, in many cases due to small sample size.

[18]The resulting estimates of this exercise are in Table A6 in Appendix A.5.

Table 4: Welfare Eligibility Manipulation Attempts Lead to Significant Increases in Child-Related Expenditure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
						Children			
	Total	Food	Food Outside of House	Alcohol Tobacco	Adult Clothing	Total	Clothing	Education	Childcare
OLS									
Repeat Interview	–40.523	–11.749	–0.647**	–3.983	–3.041***	–2.133	0.440	–1.848*	–0.725
	(33.057)	(9.162)	(0.293)	(2.689)	(0.811)	(2.423)	(1.878)	(1.068)	(0.494)
CW-OLS									
Repeat Interview	–40.229	–12.754	–0.716**	–3.906	–3.091***	–2.785	0.267	–2.267*	–0.784
	(35.897)	(9.584)	(0.301)	(2.776)	(0.839)	(2.538)	(1.925)	(1.176)	(0.539)
2SLS									
Repeat Interview	96.333	–49.255	0.516	–51.534	0.064	87.949*	55.881*	25.879	6.188
	(330.187)	(139.620)	(4.691)	(50.765)	(12.353)	(47.617)	(31.343)	(17.387)	(9.904)
SW *F* -Statistic: R.I.	8.771	8.771	8.771	8.771	8.771	8.771	8.771	8.771	8.771
Y_0	422.331	142.361	0.667	17.681	5.916	30.909	23.806	6.187	0.915
Observations	1,670	1,670	1,670	1,670	1,670	1,670	1,670	1,670	1,670

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Each column summarizes the results for the respective outcome variable following the system of equations 7 and 8 using information from the household survey. All estimations control for the first PMT first score above and below the cutoff, first monthly household benefit awarded, , and region-by-quarter and interview time fixed effects. The CW-OLS calculation follows Bhuller et al. (2020).

6 Discussion

6.1 What is the cost of manipulation?

Having documented the extent, and the consequences, of welfare manipulation, we next pose the question: how costly is welfare eligibility manipulation to the Georgian government. The figures we provide here establish a conservative lower bound on the true cost of manipulation – for instance, we do not factor in the crowding out effect that a successful manipulation attempt has on formal labor supply, and the concomitant loss in income tax revenue.

We present our estimates of the costs of welfare manipulation attempts in Table 5. We consider two primary sources of costs to the governments. First, the costs from additional welfare payments to households with successful manipulation attempts. Second, the administrative costs of additional interviews. First, in row [1] we calculate the number of successful manipulators around the 65,000 cutoff. These are households whose initial PMT score was just above 65,000, whose final score was below 65,000, and whose request for a repeat interview did not lead to a welfare suspension. Row [2] presents the total increase in benefits of those households who succeed in their manipulation attempt. To compute the calculation of the administrative cost of reassessments, we start in row [3] with the average number of (household-initiated) requested reassessments in a month. The costs we associate to each reassessment is the sum of a 3 hour wage of an enumerator, which we assume is the median wage in Tbilisi.[19] We add 40 Lari per interview as additional administrative costs.

Table 5: The Direct Costs Associated With Manipulation Are Substantial

(a) The Cost of Manipulation	
Additional Benefit Payments	
[1] Additional Households at Higher Benefit Level	403
[2] Additional Monthly Benefit Payments (Lari)	106,150
Additional Interview Costs	
[3] Additional Visits to Households	38
[4] Additional Visit Costs	3,116
[5] Total Additional Costs (Lari) Due to Manipulation	109,266
(b) Baseline Costs as a Benchmark	
[6] Number of Households at Baseline	4,854
[7] Total Baseline Benefit Costs (Lari)	441,800
[8] **Cost of Manipulation as a Percentage of Baseline Costs**	**24.73%**

Notes: The table summarizes the monthly cost of manipulation. Rows [1] and [2] are based on the difference between households with successful manipulation attempts with initial PMT scores between 65,000 and 70,000 and those with PMT scores between 60,000 and 65,000. For row [4] we assume that an interview takes three hours, enumerators are paid the median wage in Georgia (14 Lari per hour), plus an administrative cost of 40 Lari. The total additional cost associated with manipulation [5] = [2] + [4]. For row [6] we sum all households with an initial PMT score between 65,000 and 70,000. Row [7] is the associated monthly benefit payments for these reference households at baseline. Row [8] = [5]/[7].

Combining these two costs, we document that the Georgian government loses 109,000 Lari per month (about 38,000 USD per month) to welfare eligibility manipulation. To benchmark this cost, we consider the welfare expenditure on households with initial PMT scores of 65,000-70,000. The cost of manipulations is one quarter of the initial welfare payments to this group. This benchmarking highlights the substantial costs of welfare manipulation to the Georgian government, as well as to those on low incomes living in

[19] From https://www.salaryexpert.com/salary/area/georgia/tbilisi

Georgia. This money could be spent to increase the generosity of welfare payments to existing recipients, or to extend the range of PMT scores that yield welfare benefits.

Given this observation, it is worth recapping on why we see such manipulation in the first place – the specific design of the welfare system at this time in Georgia. First, there is a step-wise benefit schedule with a particularly large drop at 65,000 (Figure 1). Second, households are freely and costlessly able to request a repeat interview. Given the costs we document of welfare manipulation, it is useful to consider how one may address this issue. Removing the ability to request an additional interview may lead to poor households not having the ability to directly address genuine errors, which does not seem palatable. Removing the extreme steps in the PMT schedule however, seems like a much more direct and simple approach to obviating the large welfare manipulation we document in this paper. This could be done by replacing the large steps multiple smaller steps, or by "smoothing through" these steps, with, for instance, a reverse cumulative distribution function. In ongoing work, we are working on precisely such an approach.

6.2 Comparing our estimations with bunching style estimations

We complete this section by comparing the estimates we derive from our FDD approach with alternative approaches – specifically different bunching methods – that are typically used in the literature. Given the data we have access to for this work, we have been able to directly approach measuring the extent of welfare manipulation, with the use of an RD design. This approach can most easily be seen in Figure 2 and Figure 3. Standard bunching estimators approach the topic from a different, less direct perspective. A useful starting point for this analysis is to recall the distributions of PMT scores we present in Figure 2. The bunching methodologies we implement will estimate the "missing" mass point, based on final PMT scores, just above the 65,000 threshold.

In Table 6, we present the results of a comparison between both first and second generation bunch- ing estimators, and our approach. For the bunching estimator approaches we present the proportion of missing households in the final PMT score distribution as a proxy of those households that successfully manipulated their score. We use two different bunching approaches. The first generation bunching esti- mator follows the approach of Chetty et al. (2011) and Foremny et al. (2017). We use the distribution of the final PMT score outside a range around 65,000, where households are more likely to be passing from above to below the cutoff (exclusion area), to estimate a counterfactual distribution without bunch- ing. The proportion of missing households are the difference between the observed and counterfactual households just above 65,000 within the exclusion area. For the second generation bunching estimator, we follow Zwiers (2021) and take advantage of our data to use the initial PMT score distribution to estimate the counterfactual distribution, having already documented an absence of manipulation in this initial PMT score distribution. Finally we present the first stage coefficient of our FDD estimation (ω_3 from Equation 10(a)), which represents the proportion of households who attempted to manipulate their welfare eligibility due to falling just above the 65,000 cutoff. We present this estimate (row [3*a*]) as an intermediate estimate – it is not the correct estimate for this exercise. We present this here for com- pleteness, as this is the estimate we have used elsewhere in the paper. Row [3*b*] presents the correct estimate for this exercise – the first stage estimate when we consider *successful* manipulation attempts, i.e., manipulation attempts that lead to the types of bunching we document in Figure 2(b).

What is particularly striking about the estimates of the degree of manipulation around the 65,000 threshold that we present in Table 6 is how similar the estimates are from the different methods. Both generations of bunching estimators and our FDD approach yield as good as identical results. This is surprising given the different manners in which these methodologies estimate the degree of manipulation. Such a finding locates our approach within a more familiar terrain for estimating manipulation in response to cutoffs and kinks in public economics. That said, it should be noted that if we were relying on a bunching estimator, we would have only been able to answer one of our research questions – how prevalent is manipulation in response to the step-function benefit schedule. The additional richness we have been able to detail in the rest of our work – characterizing manipulators, documenting how households manipulate, providing evidence on the downstream consequences of manipulation including labour market participation and household expenditure – would not have been possible.

Table 6: Existing Bunching Estimates are Consistent With Our FDD Approach

	(1)
[1] **First Generation Bunching Approach**	0.133
`a la Chetty et al. (2011); Foremny et al. (2017)	[0.050]
[2] **Second Generation Bunching Approach**	0.114
`a la Zwiers (2021)	[0.031]
[3*a*] **FDD Approach**: Welfare Manipulation Attempt	0.254 (0.026)
[3*b*] **FDD Approach**: Successful Welfare Manipulation Attempt	0.117 (0.018)

Notes: The table shows the proportion of missing households resulting for bunching estimates using different samples and estimations strategies. Rows 1 and 2 estimate the counterfactual polynomial following Foremny et al. (2017); Chetty et al. (2011). We choose the polynomial degree and the exclusion window that minimizes the difference between the excess of households below the cutoff and the missing households above the cutoff. Rows 3 and 4 use the first PMT score to estimate the polynomial following Zwiers (2021). The excess of households adds the difference between the observed distribution and the counterfactual distribution, from the highest PMT score when the counterfactual is larger than the observed distribution to the cutoff. Missing households are the difference between the counterfactual and the observed distribution from the cutoff to the lowest PMT score such that the observed distribution is larger than the counterfactual. The missing as the difference between the contrafactual count and the observed count by bin. The proportion is the missing count divided by the total number of estimated households between the cutoff and the upper-bound. Rows 1 to 4 use bins of 500 points for the calculations using PMT scores from 40,000 to 90,000. Bootstrap standard error for 1000 repetitions in brackets. Rows 5 and 6 present the estimated change in manipulation attempts ω_3 from Equation 10. Analagously, rows 7 and 8 present the estimated change in successful manipulation attempts. Robust standard error in parentheses.

7 Conclusion

In this work, we study a large, nationwide welfare program in Georgia. The program uses a typical form of targeting for a developing country – proxy means testing – and has prominent discontinuities in the schedule that links benefit income to PMT scores. Coupled with the fact that households may request repeat PMT score assessments, the program gives households incentives to manipulate their welfare eligibility. We start by showing that such manipulation is extensive at a particular threshold, a threshold with a particularly large benefit discontinuity.

We develop a Becker-style model of household manipulation, which we use to inform our empirical approach – a fuzzy difference-in-discontinuities design. We provide extensive evidence of the causal effects of welfare manipulation on labor market engagement, household expenditure, and outcomes of children and young people within the household. We find that women in manipulating households work more and find null effects for men and for the total household income. By probing this finding, we document evidence of welfare benefits crowding out labor market participation for our complier households. We document that also that households spend more in their children.

Given our setting, where we have quasi-experimental variation that leads to a simultaneous drop in parental time and a rise in parental income, we study the consequences of household manipulation behavior on a battery of child outcomes, spanning from health and time use investments for children aged 0-5, to educational investments for older children and young adults. We do not find changes in children's outcomes.

We conclude our work by providing a conservative lower bound for the cost of welfare manipulation, which we find to be substantial, amounting to roughly 25% of initial welfare expenditure on our target group of households. In a follow-up project, we are working on alternative benefit schedule designs that can maintain similar levels of benefit payments to target households, yet avoid the large discontinuities that give rise to large welfare eligibility manipulation incentives. We locate our approach to estimating the prevalence of manipulation within the wider terrain of bunching estimators more commonly used in the public economics literature. Our prevalence estimates coincide almost perfectly with those from two different bunching estimator methodologies.

References

Alberto Abadie. Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Jour- nal of Econometrics*, 113(2):231–263, April 2003.

Francesco Agostinelli and Giuseppe Sorrenti. Money vs. time: family income, maternal labor supply, and child development. *University of Zurich, Department of Economics, Working Paper*, (273), 2021.

Vivi Alatas, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. Targeting the poor: evidence from a field experiment in indonesia. *American Economic Review*, 102(4):1206–40, 2012.

Abhijit V Banerjee, Rema Hanna, Gabriel E Kreindler, and Benjamin A Olken. Debunking the stereo- type of the lazy welfare recipient: Evidence from cash transfer programs. *The World Bank Research Observer*, 32(2):155–184, 2017.

Gary S Becker. Crime and punishment: An economic approach. *The Journal of Political Economy*, 76 (2):169–217, 1968.

Manudeep Bhuller, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. Incarceration, recidivism, and employment. *Journal of Political Economy*, 128(4):1269–1324, 2020.

Diogo GC Britto, Paolo Pinotti, and Breno Sampaio. The effect of job loss and unemployment insurance on crime in brazil. *Econometrica*, 90(4):1393–1423, 2022.

Fernanda Brollo, Katja Kaufmann, and Eliana La Ferrara. The political economy of program enforcement: Evidence from brazil. *Journal of the European Economic Association*, 18(2):750–791, 2020.

Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

Adriana Camacho and Emily Conover. Manipulation of social program eligibility. *American Economic Journal: Economic Policy*, 3(2):41–65, 2011.

Pedro Carneiro, Lucy Kraftman, Giacomo Mason, Lucie Moore, Imran Rasul, and Molly Scott. The impacts of a multifaceted prenatal intervention on human capital accumulation in early life. *American Economic Review*, 111(8):2506–2549, 2021.

Matias D. Cattaneo, Michael Jansson, and Xinwei Ma. Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455, 2020.

Elizabeth M Caucutt, Lance Lochner, Joseph Mullins, and Youngmin Park. Child skill production: Accounting for parental and market-based time and goods investments. Working Paper 27838, National Bureau of Economic Research, September 2020.

Raj Chetty, John N Friedman, Tore Olsen, and Luigi Pistaferri. Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The quarterly journal of economics*, 126(2):749–804, 2011.

David Coady, Margaret Grosh, and John Hoddinott. Targeting outcomes redux. *The World Bank Research Observer*, 19(1):61–85, 2004.

Flavio Cunha and James Heckman. The technology of skill formation. *American economic review*, 97(2): 31–47, 2007.

Flavio Cunha and James J Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, 43(4):738–782, 2008.

Flavio Cunha, James J Heckman, and Susanne M Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Gordon B. Dahl, Andreas Ravndal Kostøl, and Magne Mogstad. Family welfare cultures. *Quarterly Journal of Economics*, 129(4):1711–1752, November 2014.

Manasi Deshpande, Tal Gross, and Yalun Su. Disability and distress: The effect of disability programs on financial outcomes. *American Economic Journal: Applied Economics*, 13(2):151–78, 2021.

Econometr´ıa. Impact evaluation of targeted social assistance (tsa) in georgia. final report. Technical report, Econometr´ıa Consultores. Evaluation summoned by UNICEF, September 2020.

Isaac Ehrlich. Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of political Economy*, 81(3):521–565, 1973.

Dirk Foremny, Jordi Jofre-Monseny, and Albert Sol´e-Oll´e. 'ghost citizens': Using notches to identify manipulation of population-based grants. *Journal of Public Economics*, 154:49–66, 2017. ISSN 0047- 2727. doi: https://doi.org/10.1016/j.jpubeco.2017.08.011. URL https://www.sciencedirect.com/science/article/pii/S0047272717301433.

Leora Friedberg. The labor supply effects of the social security earnings test. *The Review of Economics and Statistics*, 82(1):48–63, 2000.

Alexander M Gelber, Damon Jones, and Daniel W Sacks. Estimating adjustment frictions using nonlinear budget sets: Method and evidence from the earnings test. *American Economic Journal: Applied Economics*, 12(1):1–31, 2020.

Paul J Gertler, Sebastian W Martinez, and Marta Rubio-Codina. Investing cash transfers to raise long- term living standards. *American Economic Journal: Applied Economics*, 4(1):164–192, 2012.

Michael Geruso and Timothy Layton. Upcoding: evidence from medicare on squishy risk adjustment. *Journal of Political Economy*, 128(3):984–1026, 2020.

Veronica Grembi, Tommaso Nannicini, and Ugo Troiano. Do fiscal rules matter? *American Economic Journal: Applied Economics*, 8(3):1–30, 2016.

Anthony Howell. Impact of a guaranteed minimum income program on rural–urban migration in china. *Journal of Economic Geography*, 2022.

Henrik J Kleven and Mazhar Waseem. Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *The Quarterly Journal of Economics*, 128(2):669–723, 2013.

Henrik Jacobsen Kleven, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79(3):651–692, 2011.

David S Lee and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355, 2010.

Stephen Machin, Sandra McNally, and Jenifer Ruiz-Valenzuela. Entry through the narrow door: The costs of just failing high stakes exams. *Journal of Public Economics*, 190:104224, 2020.

Jaime Mill´an-Quijano. Fuzzy difference in discontinuities. *Applied Economics Letters*, 27(19):1552–1555, 2020.

G Miller, D Pinto, and M Vera Hernandez. Risk protection, service use, and health outcomes under colombia's health insurance program for the poor. *American Economic Journal: Applied Economics*, 5(4):61–91, October 2013.

Joseph Mullins. Designing Cash Transfers in the Presence of Children's Human Capital Formation. Working Papers 2022-019, Human Capital and Economic Opportunity Working Group, July 2022.

Cheti Nicoletti, Kjell G Salvanes, and Emma Tominey. Mothers working during preschool years and child skills: does income compensate? *Journal of Labor Economics*, 41(2):389–429, 2023.

Emmanuel Saez. Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2 (3):180–212, August 2010.

Jenny Williams and Robin C. Sickles. An analysis of the crime as work model: Evidence from the 1958 philadelphia birth cohort study. *The Journal of Human Resources*, 37(3):479–509, 2002.

World Bank. Georgia—first, second and third development policy operations. Technical Report 125186, The World Bank Independent Evaluation Group, Washington, DC, 2018.

Esm´ee Zwiers. The lifecycle fertility consequences of the great depression and wwii: Evidence from the netherlands, November 2021.

A Robustness and Ancillary Results

A.1 The TSA Benefit Schedule

Table A1: TSA benefits by PMT score (Lari per month)

PMT score	Benefit per household member	Benefit per child
0 to 30, 000	60	50
30, 001 to 57, 000	50	50
57, 001 to 60, 000	40	50
60, 001 to 65, 000	30	50
65, 001 to 100, 000	0	50
100, 000 or more	0	0

Notes: Payment scheme from January 2019. 1 USD = 2.89 Lari.

A.2 Support for the Identifying Assumptions

A.2.1 The Continuity Assumption

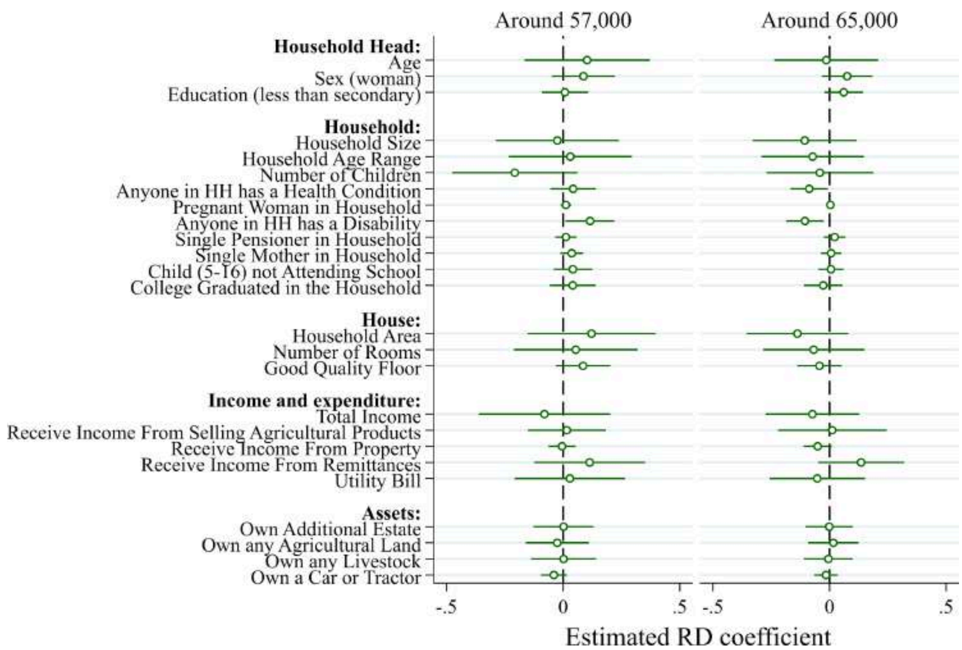
The key identifying assumption in a RD design is the continuity assumption, which states that the potential outcomes ($Y_{0,i}$ and $Y_{1,i}$) are smooth functions of the running variable $z_{0,i}$ through the cutoff, κ.

In order to provide support for this assumption, we implement a series of covariate balance tests, estimating the following specification:

$$X_i = \lambda D_i + g^D(z_{0,i}) + u_i \tag{11}$$

where $g^D(z_{0,i})$ is a polynomial of order 2 in $z_{0,i}$ on either side of the cutoff, κ. We present the estimates of λ from these regressions in Figure A1 below. In order to give support to out difference-in- discontinuities strategy, we show continuity over the 57,000 and 65,000 cutoffs. Of the 25 covariates we consider, only a single estimate is significantly different from 0 at the 5% level. We take these results as strong supportive evidence in favour of the continuity assumption.

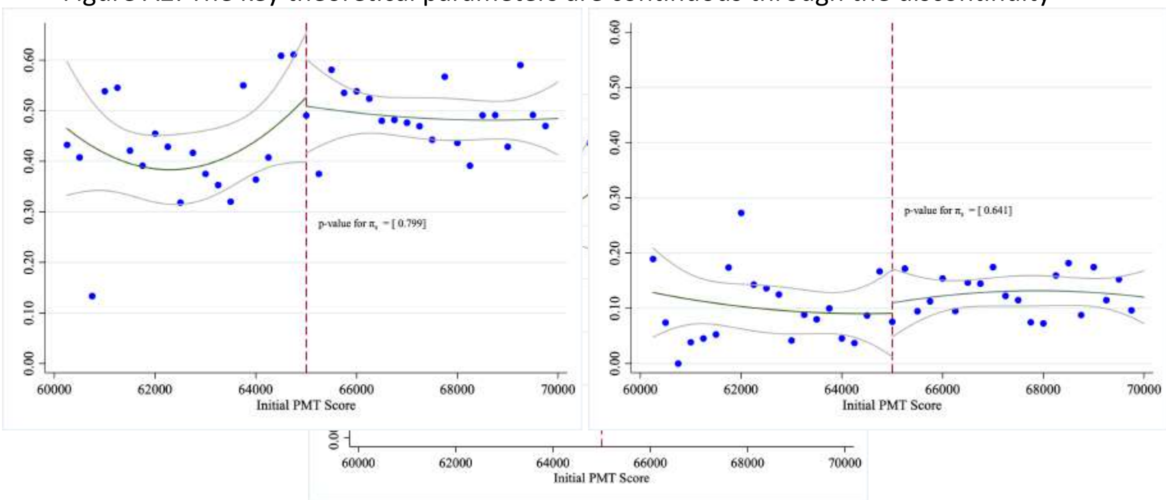
Figure A1: The Covariates at Baseline are Balanced Across the 57,000 and 65,000 PMT Thresh- old



Notes: The figures shows the resulting λ coefficients from Equation 11 for each X variable.

Finally, in Figure A2 we show that the probability of a successful reassessment and the probability of being sanctioned by the SSA, key parameters of the model exposed in section 4, are is continuous around both cutoffs.

Figure A2: The key theoretical parameters are continuous through the discontinuity

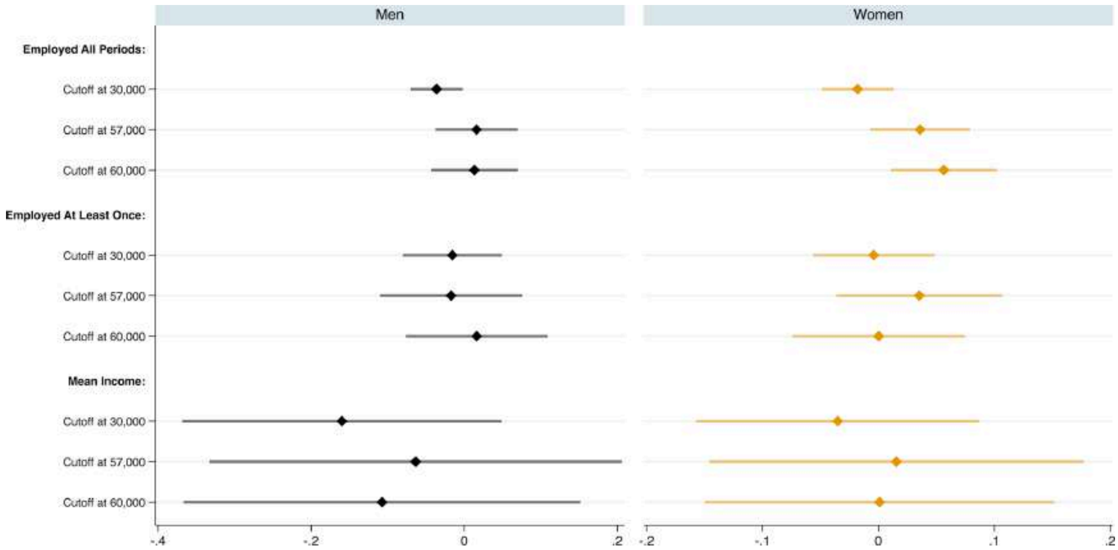


(c) $p \times (1 - q)$

Notes: Panel (a) plots the probability that a reassessment leads to a reduction in the PMT score. Panel (b) plots the probability that the household's last PMT score is recorded as invalid by the SSA. Both figures only take into account households who requested at least one additional interview. Panel (c) represents the probability of a successful manipulation attempt. In each graph we present the respective p-value for the parameter π_0 from a regression of the form $y_i = \pi_0 D_i + g^D(z_{0,i})$, where $D_i = \mathbb{1}[z_{0,i} > 65,000]$ and $g^D(z_{0,i})$ is a polynomial of order 2 in $z_{0,i}$ above and below the cutoff.

A.2.2 Homogeneity of the impact of initial benefits (B_0).

Figure A3: The Effect of Benefit Income on Labor Market Outcomes is Homogeneous Across Cutoffs

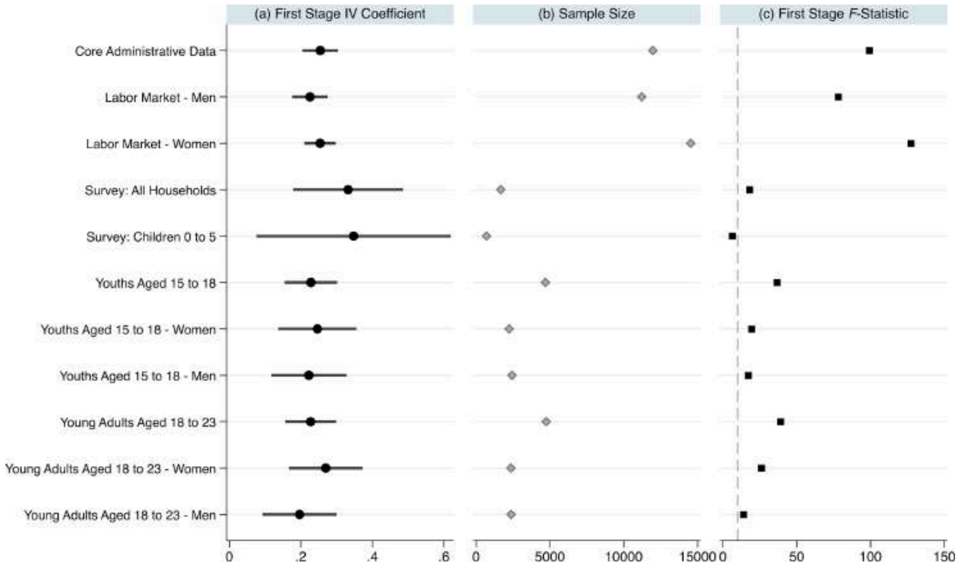


Notes: The figure shows the resulting FRD-IV coefficient of the effect of initial benefits – B_0, on each outcome variable for estimations around each cutoff.

A.2.3 First Stage Statistics

For both endogenous treatment variables, the estimated first stage coefficient on the instrument is ex- tremely stable across all data settings. Figure A4(A) shows the first stage coefficient of being above the 65,000 cutoff on the probability of requesting for a reassessment (ω_3 in Equation 10a) across various data settings we use in our analysis. The coefficients are stable, however, precision depends on sample size (Panel (B)), which can clearly be seen on Panel (C) as the F-Test for this first stage only decreases when the sample used decreases.

Figure A4: First-Stage Coefficients, Sample Size, and F -Statistics For The Effect of $A \times D$ on R

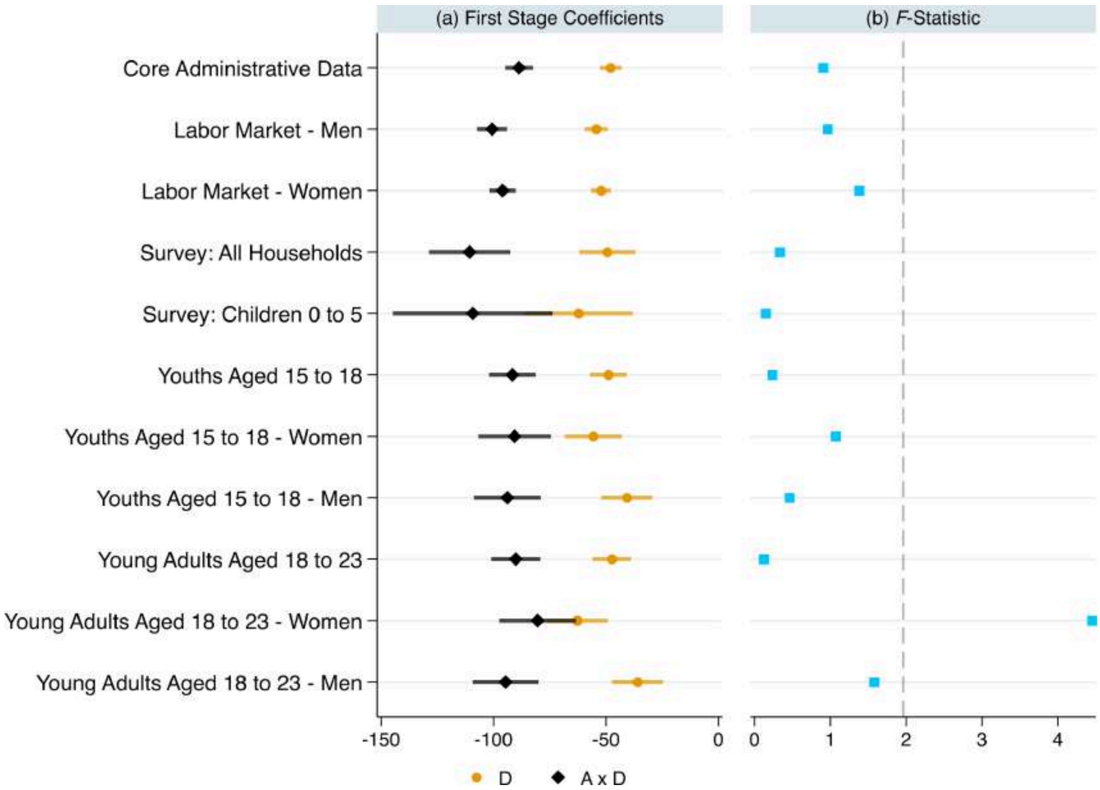


Notes: We plot the first stage coefficient of $A \times D$ for the endogenous variable R for different data sets we used over the paper. Dashed line at 10

We also estimate the effect of being above the 57,000 and 65,000 cutoffs on benefits, our second endogenous variable. As explained in Section 4.2, the FDD strategy uses variation around the 57,000 cutoff to take into account the effect of one additional Lari on initial benefits (B_0) on outcomes. It is important that the first stage considers the fact that, on the one hand, when a household score just above 57,000 it only losses 10 Lari per person with respect to a household just before 57,000. On the other hand, when a household scores just above 65,000 it losses 30 Lari pero person. Hence, we expect that the coefficients that represent the change in benefits for being above 65,000 ($\gamma_1 + \gamma_3$ in Equation 10b) is three times the change in benefits around 57,000 (γ_1). Figure A5 shows that both $\hat{\gamma}_1$ and $\hat{\gamma}_3$ are stable across different data settings. It also shows that estimated coefficients capture the differences in B_0 between both cutoffs, which is necessary for identifying the effect of manipulation attempts on household outcomes.[20]

[20]See Appendix B.

Figure A5: First stage coefficients and F-Tests for the difference of the effect of D and $A \times D$ on B_0



Notes: In Panel A we plot the first stage coefficients of D and $A \times D$ for the endogenous variable B_0 for different data sets we used over the paper. Panel B represents the F test for the null hypothesis $H_0 : \gamma_1 + \gamma_3 = 3\gamma_1$ following equation 10(a). Dashed line at 2

A.3 Additional Labor Market Results

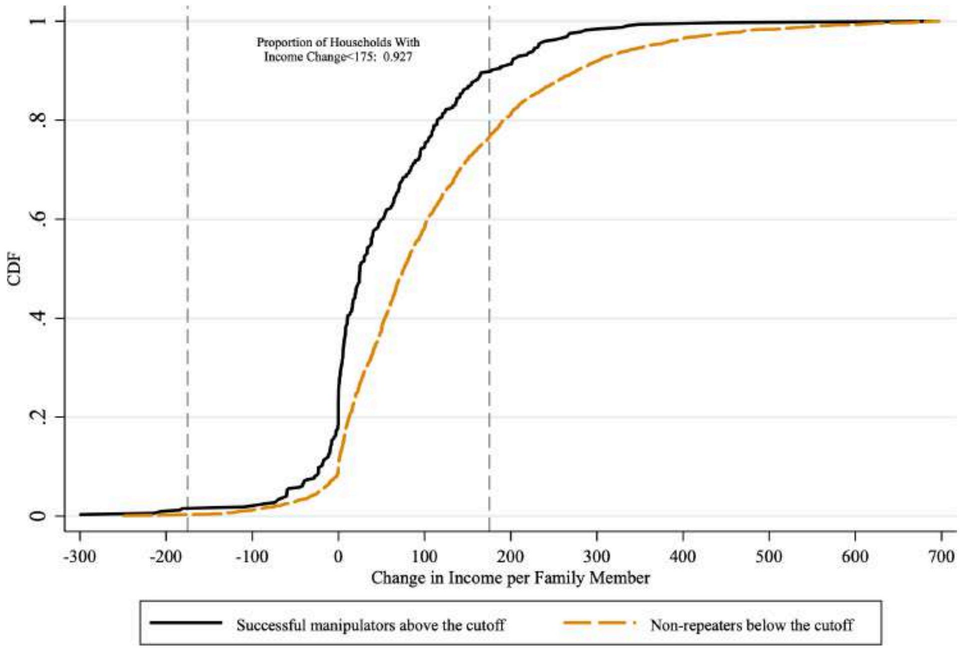
Table A2: Household Level Analysis of Labor Market Engagement Reflects What we Find at the Individual Level

	(1)	(2)	(3)	(4)	(5)
	At Least One Adult Employed at Least Once	All Adults Employed at Least Once	At Least One Adult Employed All Periods	All Adults Employed All Periods	Mean Labor Income of Household
OLS					
Repeat Interview	–0.069***	0.009	–0.009	0.007	–55.305***
	(0.012)	(0.007)	(0.009)	(0.004)	(8.805)
CW-OLS					
Repeat Interview	–0.064***	0.014*	–0.004	0.012***	–51.457***
	(0.014)	(0.008)	(0.010)	(0.005)	(9.660)
2SLS					
Repeat Interview	0.131	0.032	0.439**	0.023	26.746
	(0.269)	(0.139)	(0.202)	(0.057)	(194.333)
SW F-Stat: Repeat Interview	30.216	30.216	30.216	30.216	30.216
Y_0	0.371	0.074	0.130	0.021	156.997
Observations	11,695	11,695	11,695	11,695	11,695

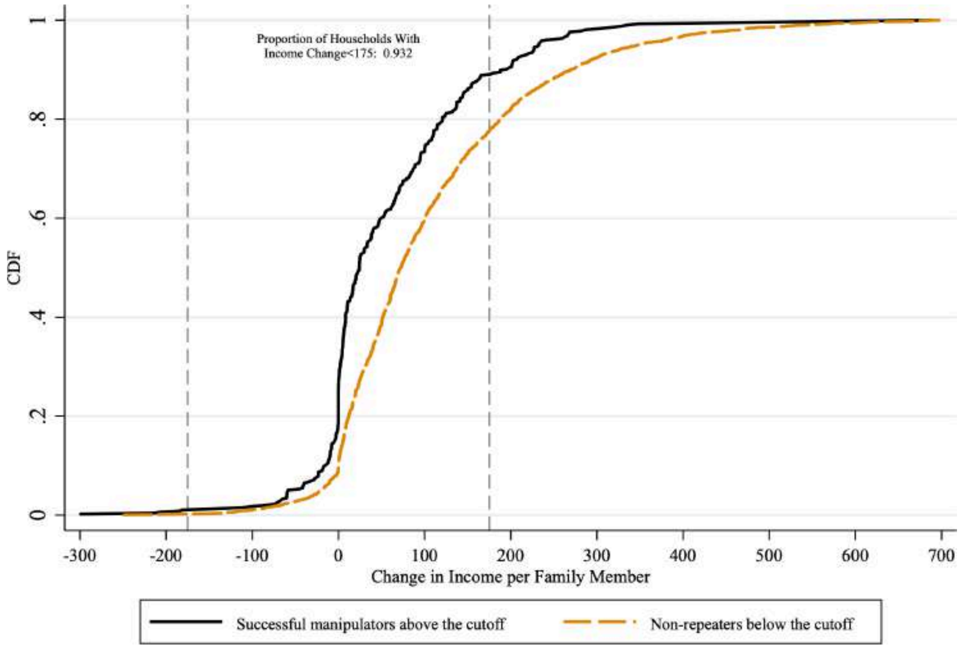
Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Initial benefit is measured in 100s of Laris.

A.3.1 Manipulation and Subsequent Income Changes

Figure A6: The Formal Labor Income Increases of Successful Manipulators Fall Almost Exclu- sively Within the 175 Lari per Month per Person Rule



(a) Unweighted



(b) Complier Reweighted

Notes: Only households-months that reported income in the Revenue Service data, with first PMT between 60,000 and 70,000. Vertical dashed lines at -175 and 175 Lari. Change in income is measured as the difference between the income earned by the household in the Revenue Service data and the Income from "Salary (including all other types of remuneration)" in the last PMT declaration filled by the household. Successful manipulators are households above the 65,000 cutoff that asked for an additional interview and the final result is a score below 65,000. Non-manipulators are households below the cutoff that do not request an additional interview.

A.4 The Impact of Welfare Eligibility Manipulation on Child Outcomes

In Section 5.3 we find that a manipulation attempt leads to a significant increase in labor supply for for women. In addition, in Table 4 we document that children are the primary beneficiaries of the corresponding increase in household spending. As our gaze turns now to child outcomes, we note that the evidence we document so far identifies two, countervailing forces on the child skill production function within households that attempt to manipulate their welfare eligibility. The increase in income, and concomitant expenditure on children, should have a positive impact on childhood skill production, whereas the fact that parents now have less time available will likely lead to a decrease in the production of childhood skills (Cunha and Heckman, 2007; Caucutt et al., 2020; Agostinelli and Sorrenti, 2021; Nicoletti et al., 2023; Mullins, 2022).

A.4.1 Early Childhood Investments

We first consider child outcomes in the first six years of life, a key period for childhood interventions if there are dynamic complementarities in investments in children across their life cycle (Cunha and Heckman, 2007, 2008; Cunha et al., 2010). We bring two data sources to bear to study this early childhood investment – administrative data on vaccinations, and survey data on health and time investments in children. The data we have available to us will predominantly reflect time costs, rather than money costs.

Once again, when working with the survey data we face a very small sample size. The consequence of this can be seen again by viewing the first-stage F statistic in Table A3, which is large for the ad- ministrative data sample, but below standard thresholds for the survey data sample. This failure of the rank condition when using the survey data occurs when we use the same specification, and considering the same PMT score range, as we do with the administrative data, so we are confident that this loss in significance reflects the small sample size of the survey data sample. Due to these power issues curtail- ing meaningful interpretation of the 2SLS estimates, we make use of the reduced form estimates in this section.

We first show in Columns 1 and 2 that a manipulation attempt has no impact on vaccinations. Given that the main parental cost of such health investments are time-based, these results are informative of household responses to changing labor supply patters as a consequence of a manipulation attempt. When we turn to the survey data, we still do not find any effect on health investments (columns 3 and 4), neither on the time parents spend with their kids.

A.4.2 Mid- and Late-Period Childhood Skill Investments

We now shift our attention to the later periods of childhood skill investments, using administrative educational data to study outcomes at two key educational margins – high school and university attendance.[21] The administrative data we use contains information on school/university attendance for the previous three years. Using this information, along with child age, we can observe if school-age children are still attending school. For 19 and 20 year olds, we can observe if they attended school in the previous years. Combining the available information, we construct an indicator for high school attendance during ages 15-18.

Columns 1-3 show the impact of a household manipulation attempt on high school attendance of children ages 15-18 within the household. Both the standard and complier re-weighted OLS estimates are both very close to zero. The 2SLS estimates are imprecise but positive. The imprecision of the estimates is not driven by small sample sizes or a weak instrument – there appears to be little effect of a household manipulation attempt (which, given what we show in previous sections, is likely best

[21] Post-compulsory education – both secondary and university education – is free in Georgia.

Table A3: There is no Strong Evidence of a Reduced Form Effect for Early Childhood Invest- ments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Administrative Data		**Survey Data**				
	Full Vaccines	Full exc. DTaP/ IPV/ Hib/ HepB	Any Health Check-ups	Number of Health Check-ups	Screen-time	Time Spent Together – Total	Time Spent Together – Reading
OLS							
Repeat Interview	–0.014	–0.029	–0.047	–0.170	–5.533	–7.714	1.023
	(0.020)	(0.022)	(0.046)	(0.450)	(6.069)	(5.906)	(2.088)
CW-OLS							
Repeat Interview	–0.019	–0.029	–0.051	–0.415	–4.305	–10.343*	0.190
	(0.023)	(0.025)	(0.047)	(0.461)	(6.146)	(5.868)	(2.060)
2SLS							
Repeat Interview	0.015	–0.190	0.567	3.445	–20.803	7.604	–4.377
	(0.327)	(0.395)	(0.557)	(5.628)	(83.031)	(74.238)	(28.642)
SW F-Stat: Repeat Interview	12.390	12.390	3.554	3.554	3.554	3.554	3.554
Y_0	0.197	0.281	0.805	5.252	64.299	74.119	15.678
Observations	3,148	3,148	701	701	701	701	701

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%.

thought of as a bundle of outcomes) on high school attendance for older children within the household. Columns 4-6 present estimates for university attendance for teens and young adults age 18-23 still living at home. The OLS show that the correlation between a manipulation attempt and university attendance is negative. The complier re-weighted OLS highlight the lack of treatment effect heterogeneity among the compliant sub-population. The 2SLS estimates tell a different story, as the coefficient for men is still negative but it is positive for women. The effect on women is large but imprecise and not statistically significant.

Summarizing, we document that manipulation attempts lead to an increase in women labor market participation (less time with children) with null effects on households' income, and an increase in children related expenditure (more money to children). We then explore what is the effect of this trade off on children's welfare, but do not find any improvements driven by a manipulation attempt.

Table A4: Household Manipulation Attempts Lead to no Changes in High School Attendance or in Unversity Attendance

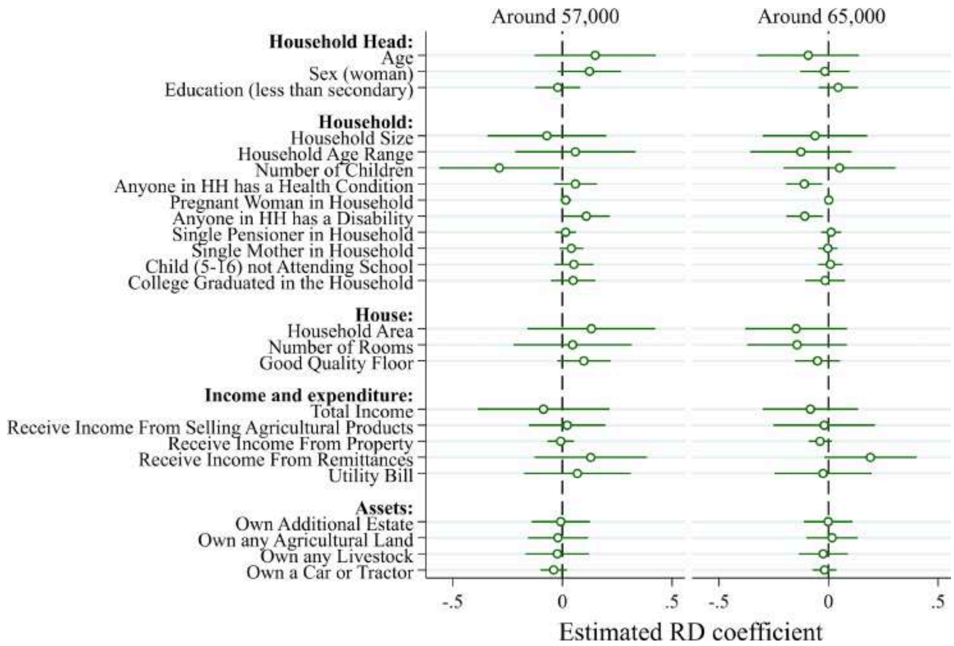
	(1)	(2)	(3)	(4)	(5)	(6)
	High School			**University**		
	All	Males	Females	All	Males	Females
OLS						
Repeat Interview	–0.007	–0.017	0.002	–0.033*	–0.039	–0.031
	(0.014)	(0.022)	(0.019)	(0.020)	(0.028)	(0.032)
CW-OLS						
Repeat Interview	–0.003	–0.013	0.018	–0.033	–0.033	–0.030
	(0.015)	(0.024)	(0.020)	(0.022)	(0.030)	(0.035)
2SLS						
Repeat Interview	0.248	0.299	0.312	–0.045	–0.724	0.318
	(0.361)	(0.581)	(0.427)	(0.342)	(0.688)	(0.393)
SW F-Stat: Repeat Interview	15.630	7.229	9.046	23.614	6.055	20.969
Y_0	0.810	0.776	0.846	0.287	0.221	0.354
Observations	6,764	3,502	3,251	4,749	2,366	2,357

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%.

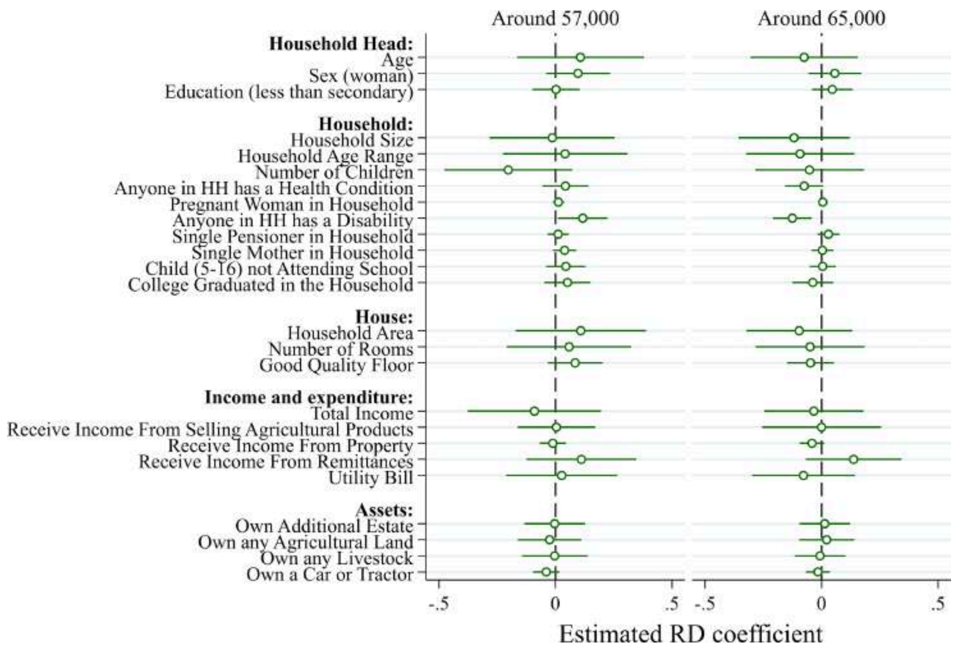
A.5 Analysis by manipulation success

A.5.1 Continuity of observables for the analysis by success of the manipulation attempt.

Figure A7: The Covariates at Baseline are Balanced Across the 57,000 and 65,000 PMT Thresh- old Even after Splitting Manipulators by Successful Status



(a) Sample with successful manipulators



(b) Sample with unsuccessful manipulators

Notes: The figures shows the resulting *λ* coefficients from Equation 11 for each *X* variable.

A.5.2 Results by success:

Table A5: The effect of manipulation attempts on labour market outcomes is stronger if the manipulation is unsuccessful

	(1)	(2)	(3)	(4)	(5)
	At Least One Adult	All Adults Employed at Least Once	At Least One Adult Employed All Periods	All Adults Employed All Periods	Mean Labor Income of Household
A. Non-repeaters and successful manipulation attempts:					
2SLS					
Repeated Interview	0.123	0.126	0.713*	0.0213	-28.99
	(0.550)	(0.282)	(0.419)	(0.104)	(390.2)
SW F-Stat: Repeated Interview	18.864	18.864	18.864	18.864	18.864
Y_0	0.371	0.074	0.130	0.021	156.997
Observations	10,517	10,517	10,517	10,517	10,517
B. Non-repeaters and unsuccessful manipulation attempts:					
2SLS					
Repeated Interview	0.193	0.0522	0.590**	0.0486	135.2
	(0.355)	(0.183)	(0.269)	(0.0750)	(260.3)
SW F-Stat: Repeated Interview	24.627	24.627	24.627	24.627	24.627
Y_0	0.371	0.074	0.130	0.021	156.997
Observations	10,811	10,811	10,811	10,811	10,811

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Each column summarizes the results for the respective outcome variable following the system of equations 7 and 8 using information from the household survey. All estimations control for the first PMT first score above and below the cutoff, first monthly household benefit awarded, , and region-by- quarter and interview time fixed effects.

Table A6: There are no differences by the result of the manipulation attempt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
						Children			
	Tot	Alcohol Tobacco			Adult Clothing		Total	Clothing	Education Childcare
A. Non-repeaters and successful manipulation attempts:									
2SLS									

B. Non-repeaters and unsuccessful manipulation attempts:

2SLS

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Each column summarizes the results for the respective outcome variable following the system of equations 7 and 8 using information from the household survey. All estimations control for the first PMT first score above and below the cutoff, first monthly household benefit awarded, , and region-by-quarter and interview time fixed effects. CW-OLS following Bhuller et al. (2020).

B Identification of θ_R

We start from Equation 10c:

$$Y_i = \theta_R R_i + \theta_B B_{0,i} + \theta_3 A_i + g_Y^{D,A}(z_{0,i}) + X_i\theta + \mu_i^{Y,i}$$

We take expectations of Y with respect to our instruments (D, $A \times D$), conditional on the score (z), and the observable (X). We define $E[Y_i|D = d, A = a, X_i, z_i] = Y^{da}$, $E[B_{0,i}|D = d, A = a, X_i, z_i] = B^{da}$, and $E[R_i|D = d, A = a, X_i, z_i] = R^{da}$. Taking into account that X is continuous around the cutoffs, and under the FDD assumption that R does not change around the 57, 000 cutoff ($R^{10} = R^{00}$), we can identify θ_B by subtracting $Y^{10} - Y^{00}$:

$$\theta_B = \frac{Y^{10} - Y^{00}}{B^{10} - B^{00}}$$

As explain before, θ_B is identified using the variation around the 57, 000 cutoff. Now, using variation around 65, 000, by subtracting $Y^{11} - Y^{01}$:

$$Y^{11} - Y^{01} = \theta_R(R^{11} - R^{01}) + \theta_B(B^{11} - B^{01})$$

Under the assumption that the effect of one additional Lari is the same around 57, 000 and 65, 000, we can plug the estimate for θ_B on the latest equation. After reorganizing, we lead to the following expression:

$$\theta_R = \frac{(Y^{11} - Y^{01}) - (Y^{10} - Y^{00})\frac{B^{11}-B^{01}}{B^{10}-B^{00}}}{R^{11} - R^{01}}$$

As in Grembi et al. (2016) and Mill´an-Quijano (2020), to identify θ_R we use the difference in the variation around 65, 000 minus the difference around 57, 000. However, given that the change in B_0 is different around 57, 000 and 65, 000, we weight the difference in outcomes by $\frac{B^{11}-B^{01}}{B^{10}-B^{00}}$, that takes into account the difference in the change in B_0.