

Welfare Eligibility Manipulation: Evidence From Georgia*

Brendon McConnell

Jaime Millán-Quijano

This draft: November 28, 2022

Abstract

Optimal targeting of social aid is a key issue in the design public policy. A key aspect of policy design is to create welfare systems that are manipulation-proof. We study demand side manipulation of a large, nationwide welfare program in Georgia. We use a rich tapestry of administrative and survey data to study a group of welfare recipient households, whose total baseline income is roughly 100USD per month. We start by documenting sizable jumps in household manipulation behavior at a key benefit discontinuity. We build a Becker (1968) style model of manipulation, which we use to inform our empirical strategy – a fuzzy regression discontinuity-instrumental variables design. Using administrative data, we find that those in households that engage in manipulation subsequently also engage more in the formal labor market, particularly men. We explore various channels for this finding. Using a marginal treatment effect analysis, we rule out that households are selecting on the unobserved (labor market) gains to manipulation. Manipulating households consequently have higher levels of income, from welfare and the labor market. Using survey data that we collect, we find that welfare-manipulating households spend the vast majority of their additional income on children, increasing expenditure on clothes and education. Given that adults in manipulating households are working more and have more income it is not clear *a priori* whether childhood outcomes will improve or not, given the countervailing inputs of parental time and money. With an early childhood skill production model in mind, we study a battery of outcomes for children and young people in the household. We find no significant changes for health outcomes for 0-5 year olds, nor do we find changes in educational engagement for teenagers. We do find an increase in university attendance for young people still living at home, particularly young women.

Keywords— Welfare Eligibility, Manipulation, Public Policy Design, Child Skill Investment

JEL Codes— H53, I38

*Author affiliations and contacts: McConnell (University of Southampton, brendon.mcconnell@gmail.com); Millán-Quijano (NCID and CEMR, jmillanq@unav.es)

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 PMT score threshold. We provide 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, 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 which uses a PMT¹ with multiple cutoffs to allocate unconditional cash transfers among low income households. 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. 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 of the initial distribution of PMT scores and the final observed distribution of the scores, pay 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), which is a statistical test of bunching. 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.

Second, we set up a model of household welfare manipulation, 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 point we return to later on in the paper when we discuss our marginal treatment effect (MTE) results.

Third, based on the evidence of behavioral responses of household to discontinuities in the welfare income-PMT score schedule, we bridge from the theoretical model and set up an empirical specification

¹The PMT is based on rich survey data, and incorporates information on households’ demographic composition, asset holdings, income, and access to public amenities. Once a household receives a score, benefit payment are allocated based on a scheme that has multiple cutoffs.

in the form of a fuzzy regression discontinuity-instrumental variables (FRD-IV) strategy (following Lee and Lemieux, 2010). 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. 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 of the FRD-IV approach.

Fourth, we restrict the data to a narrow window around the key PMT score threshold 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.

Fifth, 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 gives access to 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 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, 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.

Third, our IV results show that households that engage in welfare manipulation work and earn more in the formal labor market. This effect is driven primarily by men. Given that the OLS estimates for manipulation are significantly different from zero, but opposite-signed, we investigate if the signs of our OLS and IV results differ because the OLS estimates are plagued by selection bias, or rather a reflection of treatment effect heterogeneity coupled with the fact that IV estimates are based only on the compliant sub-population. To make progress on separating between these two sources of the sign differences of our estimates, we follow the work of Bhuller et al. (2020) and present complier-weighted OLS results. If the OLS and complier-weighted OLS are similar, we can conclude that selection bias is the source of the sign difference. This is, indeed, what we find.

We consider a set mechanisms that may underpin our findings here. We consider changing job-search behavior as one mechanism, and the relieving of certain barriers to labor supply (including transport- and childcare-related barriers). Whilst we find some supportive evidence that the additional manipulation-based benefit income may have changed job search behavior of men, we find no evidence for other barrier-related mechanisms. We additionally estimate marginal treatment effect estimates. We follow Brinch et al. (2017), who show how to identify the MTE in the presence of a binary instrument using the separate MTE estimation approach developed by Heckman and Vytlacil (2007). In practice, we estimate linear MTEs using the separate approach. Our MTE analysis finds no support for the hypothesis that households select into manipulation based on the unobserved labor market gains – the MTE curves are as good as flat for all labor market outcomes we consider, for both genders and for all functional form specifications we consider.

Fourth, we use our survey data on labor market engagement, covering formal, informal and unpaid work, to provide a broader perspective on the labor market consequences of manipulation. Such a wider view is important – at least 35% of individuals are estimated to supply at least part of their labor in the informal sector (DTDA, 2021). We find an impact of manipulation for both the extensive and intensive labor supply margins for waged work for men, yet no responses for self-employment or unpaid home work.² We find no significant labor supply responses for women.

Fifth, we study the expenditure response to welfare manipulation. A key element of our household survey involved collecting detailed household expenditure data. The IV 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 and educational expenditures. We also find a small increase in spending on adult 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 have more income, yet men in the household work more and women do not change their labor supply patterns. 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., Forthcoming; Mullins, 2022).

Sixth, 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

²Almost no men in Georgia appear to not engage in unpaid home work. Whilst over half of women do unpaid home work, only .1% of men do so.

changes in vaccination rates of children aged 0-5. Using our survey data, we find some reduced form 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.³

Seventh, we 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, but we do find increases in university attendance for 18-23 years old living in the household. This effect is large – a 10 percentage point increase off of a base of 28 percent – and driven primarily by university attendance of young women. Thus for high-school aged children, who may require parental input for homework or for getting to extra classes after school, the negative effect of less parental time appears to balance the positive effect of greater parental income. For the young adults in the households, who are more independent, the parental time effect matters less. These young adults benefit from the income effect however, and are able to attend higher education to a larger extent. Given that (i.) university education is free in Georgia and (ii.) we consider university attending young adults still living at home, the channel through which income affects university attendance is likely via the opportunity cost of university – additional income enables young woman to pursue university education rather than work, either in the labor market or at home.

Our work contributes to three strands of literature.

First, we contribute to the literature on optimal public policy targeting. Specifically, our setting is one in which there is 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.

Additionally we add to the discussion on individual and household responses to cash transfers (Banerjee et al., 2017, see the discussion in). Although in a different setting, we align with the evidence of cash transfers increasing labor market participation in Brazil (Gerard et al., 2021). The results from our quasi-experimental setting add to the discussion regarding differential responses to windfall and earned income. Previous evidence from lab-experiments show how individuals take different decisions when income is awarded and when they have to work to gain it (Barber IV and English, 2019; Charité et al., 2022). Our main estimates represent a real world example of this – household that manipulate their eligibility status spend their additional “earned” income differently to non-manipulating households. Our set of control households below the 65,000 cutoff receive the same welfare income as successful manipulators, yet receive the income by falling below the threshold, not by “working” for it.

Finally, we add to the childhood skill investment literature investigating the return of different types of parental inputs at different stages of childhood. Making full use of both our administrative 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. Our findings of differential returns by age and gender of the children add to the existing findings within this literature.

³In section A.4, 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.

2 Institutional framework and data

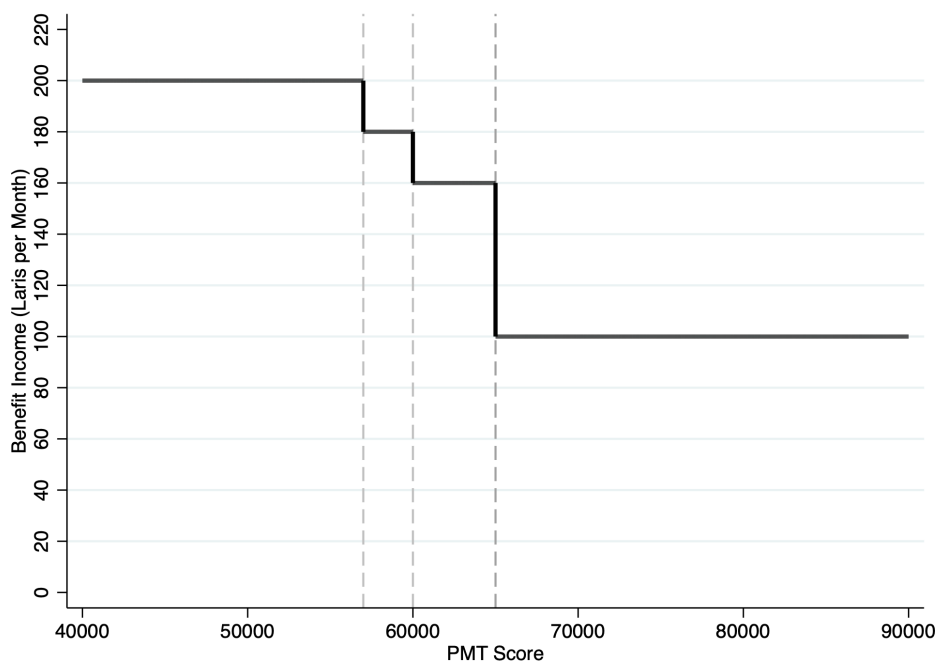
2.1 Georgia's TSA program

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 provides a full summary of the TSA benefit scheme.

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 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 household's location.⁶ In addition, if a household feels that their PMT score does not accurately represent

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

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

⁶The SSA has access to data from different governmental sources in order to follow the TSA beneficiaries. For

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-random⁷ 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 information from the Ministry of Labor database. 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

The analysis 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 we will show shortly, it is the only cutoff where we find clear evidence of PMT score manipulation. For this reason, we only use households with an initial PMT score in the range 60,000 - 70,000. We focus on households who have

example, births, deaths, children dropping out school, increases in formal labor market income, disability claims.

⁷We randomly surveyed households at specific parts of the PMT score distribution – these are the “blocks”.

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

⁹For more details of the questionnaire and sample selection in *Econometría* (2020)

children when initially assessed by the SSA as a matter of internal consistency – the household survey we conducted only interviewed households with children.

Furthermore, additional interviews play a key role in our analysis. For this reason we exclude households 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 who do not initiate the repeat interview. 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 they receive a different set of benefits from the SSA.

Our final administrative data sample, once we apply all relevant sample selection restrictions, contains 7,353 households. Our final survey data sample contains 383 households.¹⁰ Table 1 summarizes the main characteristics of our sample of analysis.

Table 1: Summary Statistics

	Mean	Standard Deviation
Household-Initiated Repeat Interview	.234	.423
Number of interviews ¹	2.87	.994
Household Composition		
Household size	4.55	1.53
Adults	2.71	1.2
Children	1.83	.852
Child Not in School	.0749	.263
Pensioner in the Household	.037	.189
Household Head Characteristics		
Age	50.9	15.3
Female	.392	.488
Single Mother	.0343	.182
Income and Expenditure		
Total income (Lari per Month)	287	272
Utility Bills (Lari per Month)	19.6	15.6
Housing Characteristics		
Number of Rooms	3.29	1.53
Good Quality Floor	.785	.411
Assets		
Owns any Estate	.626	.484
Owns a Car or Tractor	.0457	.209
Owns Agricultural Land	.616	.486
Owns any Livestock	.415	.493
Observations	7,353	

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.

In our sample, 23% 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 7.5% of households have at least one child not attending. The households in our working sample are poor – baseline income for these households is 286 Lari. 62% of households in our working sample own some form of estate (for example, garage, additional housing), 62% have agricultural land, and 41% have some livestock.

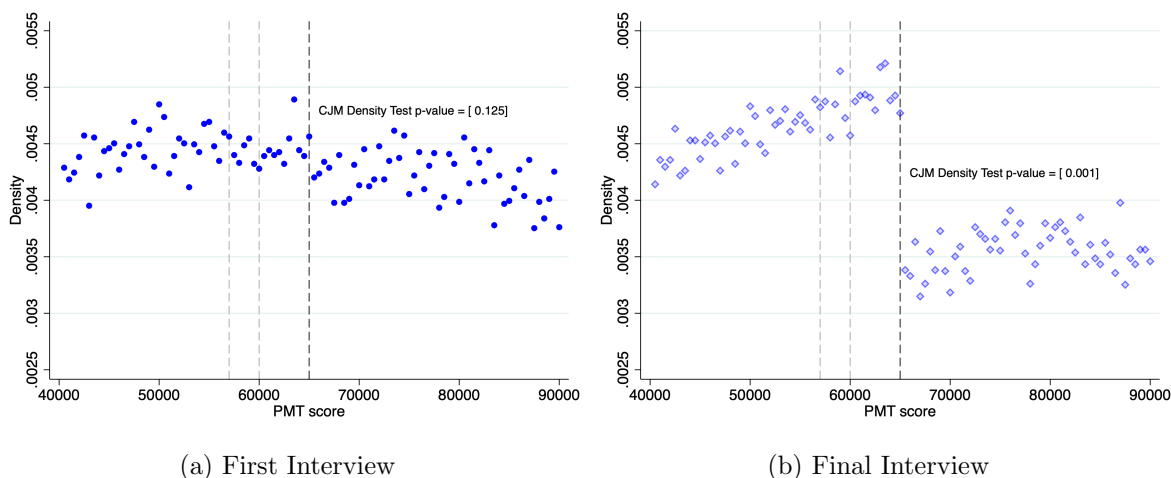
¹⁰The fact that the survey data is an order of magnitude smaller than the administrative data leads to sample sized-based power issues in some of our later analysis. Throughout our empirical work we balance the size and accuracy of the administrative data, with the richness of the survey data.

3 Evidence of Manipulation

We start by providing initial evidence of welfare manipulation, in order to motivate both the theoretical model, and 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 around 65,000 – our threshold of interest – in Figure 2(a). A visual inspection suggests that the distribution is smooth and continuous through the cutoff. This is confirmed by the associated p -value from a CJM density test (Cattaneo et al., 2020) of .125. In Figure 2(b) we present the analogous figure for the final PMT score distribution of households, allowing for household-initiated requests for reassessment. 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 .001.

Figure 2: PMT distribution for the first and last interview of each household



Notes: Bin size of 500. The CJM Density Test p -value is the p -value from the Cattaneo et al. (2020) manipulation test using households with scores between 60,000 and 90,000, a polynomial of order 2, and data driven bandwidths of: (a) First Interview: 2412.8 (below), 2980.7 (above) and (b) Last Interview 2475.8 (below), 2742.6 (above)

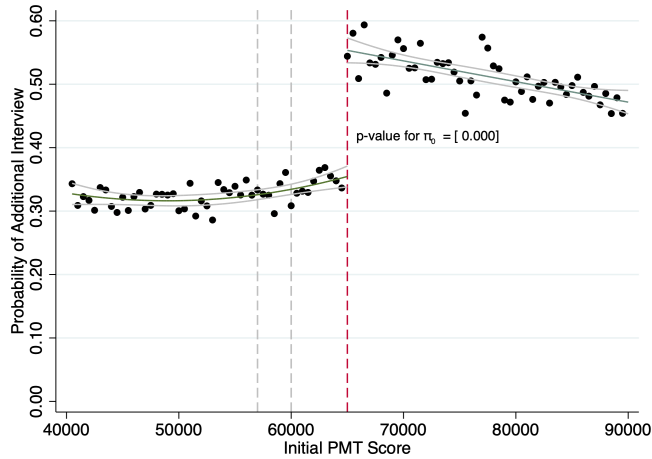
Households may be assessed multiple times for a variety of reasons and reassessments may be initiated by both households and the SSA. In Figure 3 we present evidence that is highly consistent with household-initiated requests for PMT score reassessment being the key driver of the discontinuity we document 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.

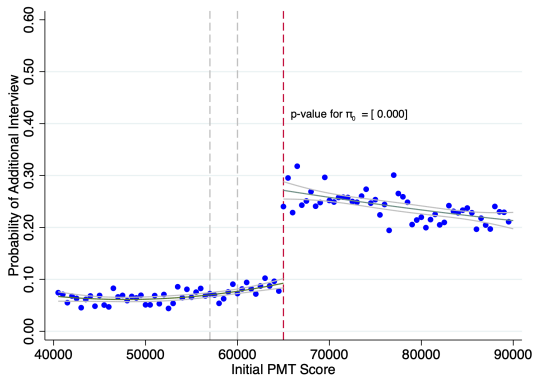
We then make use of the rich administrative data we have available to us, and separate between reasons for a reassessment. We plot the probability of a household initiated and non-household initiated reassessments respectively in Figure 3(b) and Figure 3(c). The discontinuity at 65,000 is driven solely 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.

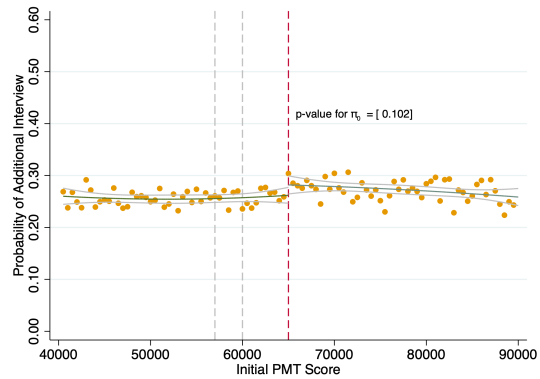
Figure 3: The Probability of an Additional Interview Jumps at 65,000 – an Effect Driven by Household-Initiated Reassessment Requests.



(a) All Cases



(b) Household-Initiated Reassessment



(c) Other Reason for Reassessment

Notes: In each graph we present the respective p -value for the parameter π_0 from a regression of the form $y_i = \pi_0 D_i + g_1^D(z_{0,i}) + \epsilon_i$, where $D_i = \mathbb{1}[z_{0,i} > 65,000]$ and $g_1^D(z_{0,i})$ is a polynomial of order 2 in $z_{0,i}$ above and below the cutoff.

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 exceeds that of accepting their initial benefit level:

$$E(V_R) > E(V_A). \quad (1)$$

When requesting a repeat interview, the individual may receive a higher benefit level B^+ with exogenously 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 requesting 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 $E(V_R)$:

$$E(V_R) = p(1 - q)U(B^+) + (1 - p)(1 - q)U(B^0) + qU(B^-) - C_i. \quad (2)$$

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

$$E(V_A) = U(B^0). \quad (3)$$

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:

$$p(1 - q)U(B_i^+) + (1 - p)(1 - q)U(B_i^0) + qU(B_i^-) - C_i = U(B^0). \quad (4)$$

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

$$p(1 - q)\Delta U^+ - q\Delta U^- > C_i. \quad (5)$$

4.2 Requesting a Reassessment and Empirical Specification.

We now map our theoretical model onto a specification that we will estimate with our data. ΔU^+ and ΔU^- are functions of a 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 $p(1 - q)\Delta U^+ - q\Delta U^- - C_i > 0$. 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:

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

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

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

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

$$R_i = \alpha_0 D_i + g_1^D(z_{0,i}) + \alpha_1 B_{0,i} + X_i' \alpha_2 + \mu_i \quad (7)$$

where $D_i = \mathbb{1}[z_{0,i} > 65,000]$ 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 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. We provide direct evidence that μ_i and D_i are indeed uncorrelated in Figure 2(a) – here we show that $z_{0,i}$ is continuous through the cutoff at 65,000, which means that D_i is randomly assigned. Random assignment of D_i rules out the possibility of a correlation with μ_i .

To measure how a welfare eligibility manipulation attempt affects a given outcome, for example, labor market participation, the second stage equation of the system is:

$$Y_i = \beta_0 R_i + g_2^D(z_{0,i}) + \beta_1 B_{0,i} + X_i' \beta_2 + \eta_i \quad (8)$$

where β_0 captures the effect of requesting a reassessment on Y .

Identification of β_0 relies on two core conditions. First we require that the distribution of our running variable $z_{0,i}$ is continuous around the cutoff, thus whether a household falls above or below the cutoff is as good as random. We have, in fact, already provided evidence in support of this first assumption, in Figure 2(a).

Second, in order to identify β_0 , we require that there is no other observable or unobservable variable that “jumps” around the cutoff and explain changes in Y_i . Figure A1 in Appendix A shows that the RDD continuity assumption holds for a large set of observable variables X , using information from the households’ initial PMT interviews. Additionally, given that (i.) the initial benefit level B_0 is determined by a function of the number of adults and children in the household, $z_{0,i}$, and D_i and (ii.) initial benefit levels will likely impact the outcomes we study, we include observed $B_{0,i}$ in all estimations.

Even having shown that the distribution of our running variable is continuous through the cutoff, and that, apart from $B_{0,i}$ which is included in all our estimations, there are no discontinuous jumps in observable characteristics at the threshold, 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 threshold 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 4, 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.

Manipulation-Induced Welfare Income Changes A welfare eligibility manipulation attempt should only affect outcomes if it leads to a change in the level of welfare income the household receives. We denote $B_{1,i}$ as the final benefit household i receives. We can write down the 2SLS system of equations as:

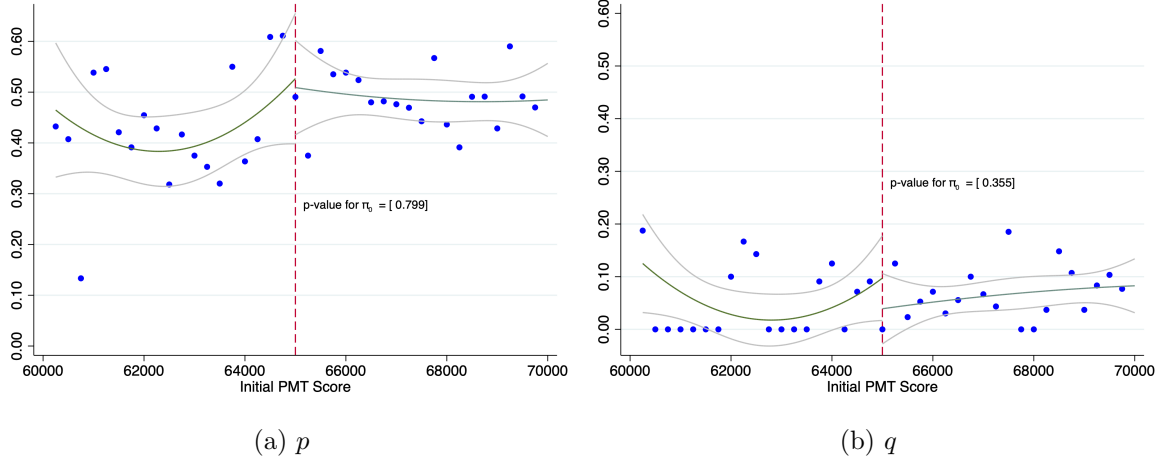
$$B_{1,i} = \alpha_0 D_i + g_1^D(z_{0,i}) + \alpha_1 B_{0,i} + X_i' \beta + \mu_i \quad (9)$$

$$Y_i = \beta_0 B_{1,i} + g_2^D(z_{0,i}) + \beta_1 B_{0,i} + X_i' \Gamma + \eta_i \quad (10)$$

In Section A.3 we show how one can move from an equation based on $B_{1,i}$ to an equation based on R_i . We do this for one key reason – when working with the survey data, we are confronted with small samples. Although the estimates from the 2SLS first stage equations are very similar when based on administrative

¹¹In order to bridge between Equations (6a) and (7) note that we parameterize $g_1(z_{0,i}, k)$ as $\alpha_0 D_i + g_1^D(z_{0,i}) + \alpha_1 B_0$.

Figure 4: The key theoretical parameters are continuous through the discontinuity



Notes: In each graph we present the respective p -value for the parameter π_0 from a regression of the form $y_i = \pi_0 D_i + g_1^D(z_{0,i})$, where $D_i = \mathbb{1}[z_{0,i} > 65,000]$ and $g_1^D(z_{0,i})$ is a polynomial of order 2 in $z_{0,i}$ above and below the cutoff.

or survey data, our first stage F -statistics are often problematically low for survey-based regressions. This is compounded by the fact that the realization of $B_{1,i}$ depends not only on a manipulation attempt R_i , but also on the response of the social security agency to the manipulation attempt, specifically whether or not the household manipulation attempt was successful (p_i), and whether or not the household was sanctioned for a fraudulent manipulation attempt (q_i). The SSA components (p and q) inject a level of randomness to the realization of $B_{1,i}$. Taken together, this results in weak instrument problems for us when using $B_{1,i}$ as the endogenous treatment variable of interest. What we show in Section A.3, is that what we find for R_i will also apply to what we would find for $B_{1,i}$. In the cases where we have sufficiently large samples, this is indeed what we find.

In Section A.4 we present first stage IV coefficients, sample sizes and first stage F -statistics for every data setting we encounter in this paper. As we show in Figure A2 and Figure A3, it isn't the first stage IV coefficient that is markedly different in each of the data settings. Rather it is the sample size. One can see the first stage F -statistic (Panel (C) of Figure A2 and Figure A3) moving in tandem with the sample size (Panel (B) of Figure A2 and Figure A3).

5 Results

We seek to answer two related questions regarding welfare eligibility manipulation. First, what type of households attempt to manipulate their welfare eligibility status. Secondly, what are the consequences – in terms of (i.) the labor market, (ii.) expenditure patterns and (iii.) outcomes for key household members – of manipulating welfare eligibility.

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 relative likelihood of having a given Bernoulli-distributed characteristic, x_{1i} , which we express as $P[x_{1i} =$

$1 | R_{1i} > R_{0i}] / P[x_{1i} = 1]$. For continuous characteristics, we binarize the variable.¹² R_{1i} and R_{0i} denote the potential outcomes of R_i when $D_i = 1$ and $D_i = 0$ respectively. We present a series of complier relative likelihoods in Figure 5(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. Complier households are more likely to own a workshop, granary or cattleshed, less likely to own a second home, more likely to have agricultural land, and more likely to own both livestock, and a car or tractor.

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 5(b) we present estimates for each characteristic from a regression where the dependent variable is household earned 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. Combining the information in Figure 5(a) and 5(b), we can better understand the baseline economic status of the compliant households – 5(b) informs us of the partial correlation between a given characteristic and baseline income, whilst Figure 5(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 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 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.

5.2.1 Administrative Data on Formal Labor Market Outcomes

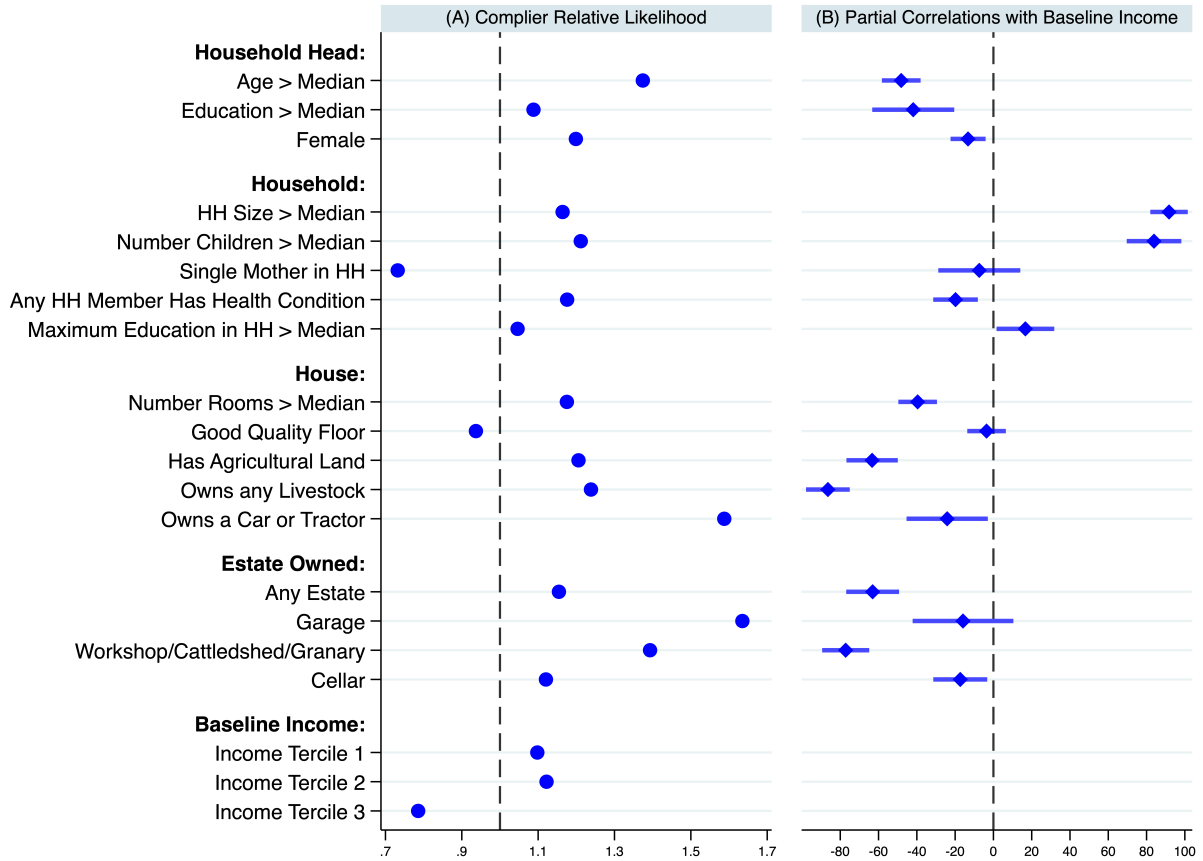
We first consider the impact of welfare eligibility manipulation on formal labor market outcomes. In Table 2 we provide evidence of the labor market consequences of attempted welfare eligibility manipulation.¹³ Although the welfare eligibility manipulation occurs at the household level, the heterogeneity in formal labor market responses by gender means it is instructive to consider the results by gender.

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 5(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

¹²We 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, $P[R_{1i} > R_{0i} | x_{1i} = 1] / P[R_{1i} > R_{0i}]$.

¹³In Table A2 we present analogous results using realized final benefits as the endogenous treatment variable.

Figure 5: 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 $P[R_{1i} > R_{0i} | x_{1i} = 1] / P[R_{1i} > R_{0i}]$ – yields the complier relative likelihood of a given characteristic, $P[x_{1i} = 1 | R_{1i} > R_{0i}] / P[x_{1i} = 1]$. 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.

of the outcome variable for those below the 65,000 cutoff (\bar{Y}_0 at the base of the table) we see those attempting to manipulate their scores earn 27% less in the formal labor market and are 18% less likely to be minimally attached to the formal labor force.

When we turn to the 2SLS results, we instead see a positive relationship between a welfare eligibility manipulation attempt and labor market outcomes. The effect is large and significant for labor income and being employed in all periods – our most stringent labor force participation measure. Those induced into a manipulation attempt due to falling just above the 65,000 cutoff are subsequently considerably more attached to the formal labor market.

The difference between the OLS estimates and the 2SLS is notable. Is this due to the OLS estimates being contaminated by selection bias? Or, rather, does this reflect treatment effect heterogeneity, whereby the compliant sub-population are so distinct that the average treatment effects for this group differs in sign from the population at large? In order to disambiguate between these two explanations, we use a complier reweighting approach analogous to that used by Bhuller et al. (2020).¹⁴

¹⁴This approach proceeds in several steps. We first predict the treatment variable with all key exogenous variables, and split the sample into 5 quintiles based on the predicted treatment index. For each quintile, we run

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
	Mean Income	Employed All Periods	Employed At Least Once	Mean Income	Employed All Periods	Employed At Least Once
OLS	-33.529***	0.007	-0.049***	-10.527***	-0.003	-0.029***
Repeat Interview	(9.851)	(0.009)	(0.014)	(3.981)	(0.007)	(0.011)
CW-OLS	-32.348***	0.008	-0.044***	-9.750**	-0.001	-0.025**
Repeat Interview	(9.127)	(0.009)	(0.014)	(4.022)	(0.007)	(0.011)
Reduced Form $\mathbb{1}[z_{0,i} \geq \kappa]$	98.846***	0.059***	0.046	14.116*	0.011	0.032
	(21.311)	(0.018)	(0.030)	(8.341)	(0.014)	(0.022)
2SLS	325.188***	0.193***	0.152	49.033	0.037	0.112
Repeat Interview	(75.740)	(0.061)	(0.096)	(29.891)	(0.047)	(0.077)
First Stage F-Stat	149.663	149.663	149.663	185.255	185.255	185.255
\bar{Y}_0	126.016	0.078	0.279	51.303	0.061	0.198
Observations	7,314	7,314	7,314	9,519	9,519	9,519

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

These results (which we present in Table 2 directly below the OLS results under the title of CW-OLS for Complier-Weighted OLS) are a useful bridge between the OLS and 2SLS results, as the CW-OLS estimates will reflect any treatment effect heterogeneity specific to the compliant sub-population, but will not circumvent endogeneity issues due to selection bias. The complier-weighted OLS results are as good as identical to the OLS results. We can conclude then that the fact that the 2SLS estimates are of different sign to OLS is due to selection bias contaminating the OLS estimates.

The results we present in Table A2, essentially mirror, and explain, those that we find for a welfare eligibility manipulation attempt. The labor market outcomes of welfare eligibility manipulators change after a manipulation attempt, as, on average, these attempts lead to higher welfare payments. For men, the OLS estimates document a negative relationship between final benefits and labor market outcomes. This statement remains true even if we re-weight our estimation sample to reflect the compliant sub-population. The 2SLS estimates are positive, significant for labor market earnings and being strongly attached to the labor market, and large in magnitude.¹⁵

In columns 4-6 of Table 2 we repeat our formal labor market analysis for women. The OLS estimates follow the same pattern as those for men, but are smaller in magnitude. The 2SLS estimates are all positive, but none are statistically significant. It is at this point where it is useful to consider the reduced form, at least to get the sign of the impact of falling above the 65,000 threshold on labor market outcomes. For formal earnings we see a positive and marginally statistically significant effect.

In Table A4 we present marginal treatment effect (MTE) analysis for the formal labor market outcomes. Given we have a binary instrument, we follow Brinch et al. (2017) and specify a linear MTE using the separate MTE estimation approach. We provide details of this in Section A.9. Our main reason to conduct MTE analysis was to explore the extent to which households select into manipulating their welfare eligibility status due to unobserved labor market gains of manipulation. We find no evidence that this occurs.

From the formal labor market analysis, we can conclude that in households induced to attempt to a separate first stage equation, in order to estimate the proportion of compliers in each quintile. The mechanics of this step is identical to that of the approach we discuss in Section 5.1, where we characterize compliers. We then re-weight our core sample so that the proportion of compliers in each quintile matches the proportion of the estimation sample for the quintile. Quintiles with few compliers will receive lower weights, and quintiles with a higher proportion of compliers will receive higher weights.

¹⁵We present the reduced form (where we regress our outcome on our standard set of control variables, and the instrument D_i directly) for the reader interested in understanding the magnitude of our 2SLS estimates.

manipulate their welfare status due to falling about the key threshold, there is a strong increase in formal sector earnings and labor supply from men and a more muted response from women, which appears to move in the same direction. We find no evidence of unobserved labor market gains to manipulation.

Interpreting the Formal Labor Market Results It is not immediately clear why we find an increase in formal labor market attachment, and consequent formal earnings, as a response to a manipulation-induced increase in welfare payments. In this section we propose three possible explanations for this, and assess whether there is any evidence that these explanations could be key drivers of the formal labor market results. It is important to note that under the rules of the TSA system in Georgia, gaining formal labor market employment does not preclude one from accessing welfare benefits if the increase in household income does not exceed 175 Lari per family member per month. The design of this rule was to enable those on benefits to work, without concerns of losing benefits in so-called “welfare traps”.¹⁶

The first potential explanation relates to job search behavior. Standard job search theory tells us that a *ceteris paribus* increase in non-labor income should lead to an increase in the reservation wage, and consequently a longer search duration and a higher wage on re-employment. Job searchers may also wait longer for a better match to arrive, which all else equal should result in a longer tenure spell with the newly matched employer.

In section A.8.1 we describe the steps we take to create the required data. In Table 3 we present the results of this analysis. For men, we document large increases in the first recorded post-unemployment income following a welfare manipulation attempt, which is suggestive that male workers may have increased their reservation wage in response to higher expected non-labor income. The findings for unemployment duration (measured not in months, but rather across the distinct time periods for which we observe the formal sector earnings and employment) are more nuanced. Whilst the OLS results point to shorter durations (and lower post-unemployment earnings) for those who attempt to manipulate their welfare status, the IV results find no difference between manipulators and non-manipulators. This suggests manipulating households converge to the job search behavior of (*ex-ante* better off) non-manipulating households. In terms of duration of employment spells post-unemployment – which is our best proxy of employer-employee match quality – we document a similar convergence in IV results, which again differ to the OLS estimates. The patterns we observe do not point to change in job search behavior as the singular channel through which manipulating households change their labor market outcomes, but are suggestive that job search behavior changes may be one such mechanism.

The pattern for women is considerably less clear cut, which echoes the more muted formal labor market response of women in manipulating households that we document in Table 2. The OLS estimates for women highlight significantly lower post-unemployment income, shorter unemployment durations and shorter post-unemployment employment spells. The IV estimates point to shorter unemployment durations, and some convergence in post-unemployment income and employment spells.

The second explanation relates to financial barriers to formal employment. Recall from Section 5.1 that our compliers are poorer households and more likely to live in a rural setting. Income constraints may mean that these individuals do not pursue formal labor market opportunities due to not being able to access these formal markets either physically – due to lack of transportation – or electronically – due to lack of means to use phones or email. Using the survey data we have available we can investigate such explanations. In Table A3 we present results from this ancillary analysis. The 2SLS estimates highlight that manipulating households spend more on communication, which differs from the negative OLS estimates for this expenditure. For transport expenditure we do not document any significant responses. We do find marginally significant reduced form evidence that households falling about the

¹⁶Rules explained in Resolution No.3, January 18th, 2019, that modify Resolution No. 145, July 28th, 2006 (accessed from matsne.gov.ge). Figure A5 in Appendix A.7 shows that most successful manipulators who have a formal job remain within the 175 Lari per month per family member rule.

Table 3: Welfare Eligibility Manipulation Leads to Changes in Aspect of Job Search and Matching

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
Treatment Variable	Income After Unemployment	Longest Unemployment Spell	Longest Post-Employment Spell	Income After Unemployment	Longest Unemployment Spell	Longest Post-Employment Spell
OLS						
Repeat Interview	-206.145*** (61.582)	-0.239** (0.097)	-0.321*** (0.090)	-48.057** (24.117)	-0.234** (0.117)	-0.370*** (0.099)
CW-OLS						
Repeat Interview	-197.612*** (56.939)	-0.247** (0.097)	-0.311*** (0.092)	-52.435** (25.075)	-0.243** (0.117)	-0.363*** (0.099)
Reduced Form $\mathbb{1}[z_{0,i} \geq \kappa]$	355.051*** (134.295)	-0.167 (0.201)	0.196 (0.168)	15.353 (58.406)	-0.733*** (0.242)	0.040 (0.189)
2SLS						
Repeat Interview	1456.014** (639.308)	-0.711 (0.853)	0.804 (0.752)	68.053 (255.781)	-2.511** (1.024)	0.178 (0.845)
First Stage F-Stat	23.219	14.496	23.219	18.125	15.255	18.125
\bar{Y}_0	709.373	2.167	1.907	374.516	2.013	2.026
Observations	1,488	1,142	1,488	1,393	908	1,393

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

65,000 PMT threshold are more likely to own a bicycle, which translates into a positive but significant 2SLS estimate. We find no effect of a manipulation attempt on motorcycle or car ownership, which is reassuring, as such expenditures are of a different order of magnitude to the financial gains manipulating households make. Taken together, we do not find any conclusive evidence that the increase in income of manipulating households leads to changes in transport expenditure. Given we do not know the purpose of the increased expenditure on communication, we conclude this exploratory analysis by acknowledging that the relaxation of financial constraints to travel are an unlikely driver of our formal labor market results.

Finally, we consider the possibility that lack of ability to fund childcare is a limiting factor on the formal sector employment of complier households. One finding from above already suggests this is not the case. If additional welfare income enables households to access childcare in order to increase labor supply, it is likely that this will impact women more than men. Yet in Table 2 we find a muted response on female labor supply, and a large response for men. Hence this is an unlikely channel. We further rule this explanation out in results we present later on in this paper (specifically Table 5), where we show no increase in childcare expenditure in response to welfare manipulation. We conclude that access to childcare is unlikely to be the key driver of our formal sector findings.

5.2.2 Survey Data on Labor Market Outcomes

If we were to only have access to data on formal labor market outcomes, we would be limit with what we could say about the labor market consequences of welfare eligibility manipulation in Georgia for two reasons. Firstly, the size of the informal sector in Georgia is non-negligible – at least 35% of individuals are estimated to supply at least part of their labor in the informal sector (DTDA, 2021). Hence, estimates based solely on the formal sector can present at best a limited perspective of the impact of welfare eligibility manipulation on labor market outcomes. Secondly, households may opt to supply labor in the informal sector as a response to the social security system in Georgia. Thus we may be missing part of the behavioral response of households to the discontinuity at 65,000.

In order to address these concerns, we now turn to survey data on labor market participation, which covers both the formal and informal sectors, as well as unpaid home production. In Table 4 we consider

both the extensive and intensive labor supply decisions of men (Panel A.)) and women (Panel B.)) separately.

Given that the survey data sample is almost an order of magnitude smaller than the administrative data sample, the key parameters are considerably less precisely estimated. It is reassuring to see that our instrument still has a sufficiently strong first stage when consider welfare eligibility manipulation attempts in this smaller sample.¹⁷ For men we document results that broadly mirror what we find with the administrative data. Specifically we see that (i.) all OLS estimates are small and statistically insignificant, (ii.) we find positive impacts of a welfare eligibility manipulation attempt on labor supply. We can separate waged work, self-employment and unpaid home work in the survey data. We see no response for labor supply decisions in the self-employed sector – the overall labor supply impacts are driven by waged work. In Table A6 we separate formal and informal waged work. Although the patterns are suggestive that the overall effects we document for waged work are driven by responses for those working in the formal sector, the estimates are too noisy for us to say anything more specific.

For women, there is no impact of a welfare eligibility manipulation attempt on labor market outcomes, in any of the sectors we consider.

The lack of statistical significance for many of the parameter estimates of interest is likely due to power issues that we face when working with the considerable smaller, but more detailed, survey data.¹⁸ Given this predicament, our task is to combine the evidence we find with the survey data and the larger, less detailed administrative data from the previous section.

The conclusions we arrive to from this triangulation process are that (i.) men supply more labor as a consequence of a welfare eligibility manipulation attempt, (ii.) this effect is driven primarily by labor supply in the formal sector of the labor market, (iii.) the informal sector response (be this informal waged work or along the self-employment margin) is limited, (iv.) labor income of men rises as a consequence of this increase in labor supply and (v.) there is very little response from women to a manipulation attempt. Bringing these points together, households that attempt to manipulate their welfare eligibility appear to be more attached to the labor market as a consequence of this manipulation attempt.

5.3 Household Expenditure Responses to Welfare Eligibility Manipulation

In the previous section we study the impact of welfare eligibility manipulation on labor market outcomes and find (i) a positive effect of welfare eligibility manipulation on employment and earnings for men and (ii) no significant impacts for women. From this we conclude that, on average, households that engage in welfare eligibility manipulation are better off for two reasons. Not only do these households receive higher welfare-related income (see Figure A4 in Appendix A.5), but those households that manipulate their welfare eligibility consequently supply more labor, and thus earn more.

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 5 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 (\bar{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. As we saw in the previous section, the 2SLS estimates have a different sign. What is most interesting about this table is where the extra expenditure

¹⁷That said, the instrument is weak when considering the outcome of a manipulation attempt – final benefits received. For completeness, we provide the results based on this treatment variable in Table A5. Given the rank condition does not hold, we cannot meaningfully interpret the 2SLS estimates.

¹⁸One can see this clearly by comparing the first-stage F -Statistic we present in Panel A of Table 2 with that of Table A5 – this is the same first-stage, based on the same set of covariates and fixed effects. The only difference is the sample size.

Table 4: We Document the Same Labor Market Patterns Using Final Benefits Instead of Repeat Interview as Our Treatment Variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Employed				Hours per Week		
	Total	For a Wage	Self-Employed	Unpaid Home Work	Total	For a Wage	Self-Employed
A. Men							
OLS:	0.011	0.017	-0.006	0.006	-0.225	-0.531	0.306
Repeat Interview	(0.042)	(0.032)	(0.041)	(0.005)	(1.679)	(1.462)	(1.317)
CW-OLS:	0.010	0.016	-0.006	0.006	-0.285	-0.607	0.321
Repeat Interview	(0.042)	(0.032)	(0.041)	(0.005)	(1.687)	(1.474)	(1.325)
Reduced Form:	0.133	0.146*	-0.012	-0.007	8.599**	6.744*	1.855
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.100)	(0.077)	(0.097)	(0.010)	(4.025)	(3.643)	(3.062)
2SLS:	0.507	0.553*	-0.046	-0.025	32.671*	25.624*	7.048
Repeat Interview	(0.397)	(0.318)	(0.360)	(0.044)	(17.646)	(15.007)	(11.371)
First Stage F-Stat	11.200	11.200	11.200	11.200	11.200	11.200	11.200
\bar{Y}_0	0.549	0.162	0.387	0.001	16.616	7.141	9.475
Observations	999	999	999	999	999	999	999
B. Women							
OLS:	-0.003	0.000	-0.003	-0.005	0.637	0.566	0.070
Repeat Interview	(0.037)	(0.030)	(0.027)	(0.039)	(1.389)	(1.324)	(0.594)
CW-OLS:	0.008	0.005	0.002	0.003	0.670	0.632	0.038
Repeat Interview	(0.039)	(0.032)	(0.027)	(0.040)	(1.499)	(1.460)	(0.545)
Reduced Form:	0.014	-0.035	0.048	-0.068	-0.589	-1.999	1.410
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.083)	(0.068)	(0.063)	(0.085)	(2.901)	(2.761)	(1.409)
2SLS:	0.056	-0.141	0.197	-0.277	-2.390	-8.116	5.726
Repeat Interview	(0.327)	(0.256)	(0.260)	(0.352)	(11.642)	(11.088)	(6.167)
First Stage F-Stat	12.262	12.262	12.262	12.262	12.262	12.262	12.262
\bar{Y}_0	0.299	0.149	0.150	0.513	8.740	5.673	3.067
Observations	1,173	1,173	1,173	1,173	1,173	1,173	1,173

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

is directed – not food, eating out, tobacco or alcohol, but on the children in the household. We document a 120 Lari increase in total expenditure on children, the lion’s share of which is on clothing, and 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 70% 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 25% of baseline expenditure. The fact that the first stage coefficient is .268 means that the 2SLS estimates are four times the size of the reduced form estimates. Irrespective of how these estimates are scaled or how we benchmark the effect sizes, the response of households to the increase in household income due to welfare eligibility manipulation is unambiguous – the majority of additional expenditure is directed towards children. In the next section we will push this finding one step further, and study the impact of household manipulation of welfare eligibility on child outcomes.

Before we do so, it is worth noting an additional piece of evidence we glean from Table 5. In Column 9 we see a small, but both statistically and economically insignificant rise in childcare expenditure. This

may explain the gendered labor market results from the previous section, where we see a large labor market response for men, but a considerably more muted response from women. The mean levels of the various outcome variables that we present in the penultimate rows of each panel of Table 4 also show men do essentially no unpaid home work.

Table 5: Household Level Survey – 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:	–33.458	–14.359	–0.598	–3.776	–3.390***	–3.332	–0.026	–2.805	–0.501
Repeat Interview	(49.303)	(12.479)	(0.408)	(3.803)	(1.164)	(3.172)	(2.296)	(1.708)	(0.651)
CW-OLS:	–21.788	–12.709	–0.640	–4.883	–3.073**	–5.697	0.101	–4.673**	–1.125
Repeat Interview	(57.837)	(14.944)	(0.411)	(4.039)	(1.345)	(3.883)	(2.483)	(2.278)	(1.016)
Reduced Form:	45.150	7.749	0.867	2.628	6.622**	31.973***	22.102***	8.464**	1.407
$\mathbb{1}[z_{1,i} \geq \kappa]$	(97.388)	(30.357)	(1.456)	(8.459)	(2.629)	(7.068)	(4.952)	(3.702)	(1.739)
2SLS:	168.555	28.929	3.235	9.811	24.721*	119.364**	82.512**	31.598*	5.255
Repeat Interview	(362.533)	(112.900)	(5.643)	(31.686)	(13.174)	(47.918)	(32.640)	(17.945)	(6.686)
First Stage F-Stat	10.338	10.338	10.338	10.338	10.338	10.338	10.338	10.338	10.338
\bar{Y}_0	471.484	158.610	0.983	21.917	7.247	32.864	24.582	7.034	1.249
Observations	900	900	900	900	900	900	900	900	900

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

5.4 The Impact of Welfare Eligibility Manipulation on Child Outcomes

In Section 5.2 we find that a manipulation attempt leads to a significant increase in labor supply for men, and no change for women. A consequence of this is an increase in labor income within the household. In Table 5 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., Forthcoming; Mullins, 2022).

5.4.1 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 6, which is large for the administrative 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 curtailing 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 patterns as a consequence of a manipulation attempt.

When we turn to the survey data, two key findings emerge. First, from the reduced form estimates, we see that there is a negative impact of falling just above the 65,000 threshold in terms of the number of health checks children under 6 receive. Second, although the reduced form estimates are too noisy to say anything with a strong degree of statistical certitude we see a substantial fall in total child-parent time for households above the 65,000 threshold.

Given the sample size limitations, it would be remiss to say too much based on the evidence from the survey data. That said, it does not appear that there are large and meaningful effects in either direction in terms of early childhood investments as a response to the discontinuity at 65,000.

5.4.2 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.¹⁹ 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. Some basic analysis of the university data highlights that ages 18-23 are core ages for university

¹⁹Post-compulsory education – both secondary and university education – is free in Georgia.

Table 6: There is no Strong Evidence of a Reduced Form Effect for Early Childhood Investments

	(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	0.028	0.004	0.004	0.482	0.010	-18.007**	-1.842
Repeat Interview	(0.024)	(0.026)	(0.076)	(0.684)	(8.435)	(8.371)	(2.995)
CW-OLS	0.033	0.011	0.006	0.405	-0.325	-18.996**	-2.184
Repeat Interview	(0.025)	(0.027)	(0.076)	(0.691)	(8.581)	(8.497)	(2.976)
Reduced Form	0.014	-0.001	-0.172	-2.896*	-16.197	-30.958	3.072
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.048)	(0.054)	(0.170)	(1.648)	(22.606)	(21.197)	(8.649)
2SLS	0.042	-0.004	-0.424	-7.141	-39.931	-76.324	7.574
Repeat Interview	(0.149)	(0.167)	(0.391)	(4.927)	(52.590)	(55.340)	(20.201)
First Stage F-Stat	42.447	42.447	7.415	7.415	7.415	7.415	7.415
\bar{Y}_0	0.207	0.291	0.814	5.462	63.874	71.307	15.785
Observations	2,014	2,014	364	364	364	364	364

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

attendance. Hence we focus on this age group, and create an indicator for university attendance or completion for those aged 18-23, in an analogous manner to the construction of our high school attendance indicator. We present our findings in Table 7.

Table 7: The Treatment Effect Estimates for High School and University Attendance Differ in Sign, Suggesting That Different Inputs may Matter More at Different Life Cycle Stages

	(1)	(2)	(3)	(4)	(5)	(6)
	High School			University		
	All	Males	Females	All	Males	Females
OLS	-0.012	-0.016	-0.011	-0.032*	-0.041*	-0.027
Repeat Interview	(0.016)	(0.026)	(0.022)	(0.019)	(0.025)	(0.031)
CW-OLS	-0.010	-0.015	-0.009	-0.038*	-0.040	-0.042
Repeat Interview	(0.016)	(0.026)	(0.022)	(0.019)	(0.025)	(0.032)
Reduced Form	0.018	-0.028	0.067	0.104***	0.073	0.144***
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.032)	(0.050)	(0.045)	(0.033)	(0.044)	(0.054)
2SLS	0.067	-0.111	0.265	0.559***	0.490	0.691**
Repeat Interview	(0.121)	(0.192)	(0.181)	(0.198)	(0.324)	(0.297)
First Stage F-Stat	80.350	37.756	32.012	45.086	14.692	23.413
\bar{Y}_0	0.807	0.776	0.840	0.282	0.213	0.360
Observations	4,365	2,246	2,093	4,047	2,128	1,884

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

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 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 distinctly different story – we find a strong positive effect of manipulation-induced benefit changes on university attendance, particularly for young women.

Why should welfare manipulation attempts lead to no impact on educational outcomes for high school aged children, but positive effects for young adults within the household? Our favored explanation for these results is that the skill production function inputs of (i.) parental income and (ii.) parental time have different returns at the different life-cycle stages, at least within the setting of Georgia. A very concise summary of the findings from previous stages is that welfare eligibility manipulation attempts lead to more income, and less available time, within the household as a whole. Combining this summary with the idea that the different inputs have different returns at different life-cycle stages enables us to reconcile the results we document in Table 7. For teenagers still living at home who may require parental input for homework or for getting to extra classes after school, the negative effect of less parental time cancels out the positive effect of greater parental income. For the young adults in the households, who are more independent, the parental time effect matters less. These young adults benefit from the income effect however, and are able to attend higher education to a larger extent. Given that (i.) university education is free in Georgia and (ii.) we consider university attending young adults still living at home, the channel through which income affects university attendance is likely via the opportunity cost of university – additional income enables young woman to pursue university education rather than work, either in the labor market or at home.

6 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 FRD-IV 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. Men in manipulating households work more, and earn more. The effect for women is more muted. In total, households that attempt to manipulated their welfare eligibility status are better off on average for two reasons – higher welfare income, and higher earned labor income. We document that these household spend more, with the majority of this increased expenditure child-focused.

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. For young children we document limited declines in health and time investments. For 15-18 year olds, we find no changes in high school attendance. For the oldest children in the household we find large increases in university attendance, particularly for young women.

The combined findings of our work provide a detailed, and morally nuanced picture of welfare manipulation in Georgia. The act of welfare manipulation is one that from an optimal welfare policy design we would like to avoid. And yet, those households who do engage in welfare manipulation end up spending working more in the formal sector of the labor market, increasing expenditure on children in the household, and end up being able to send the young women in their household to university in greater proportions.

References

- Alberto Abadie. Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal 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.
- Martin Eckhoff Andresen. Exploring marginal treatment effects: Flexible estimation using stata. *The Stata Journal*, 18(1):118–158, 2018.
- Abhijit V Banerjee, Rema Hanna, Gabriel E Kreindler, and Benjamin A Olken. Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs. *The World Bank Research Observer*, 32(2):155–184, 2017.
- Benjamin S Barber IV and William English. The origin of wealth matters: Equity norms trump equality norms in the ultimatum game with earned endowments. *Journal of Economic Behavior & Organization*, 158:33–43, 2019.
- 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.
- Christian N. Brinch, Magne Mogstad, and Matthew Wiswall. Beyond LATE with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039, 2017.
- 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.
- Adriana Camacho and Emily Conover. Manipulation of social program eligibility. *American Economic Journal: Economic Policy*, 3(2):41–65, 2011.
- 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.
- Jimmy Charité, Raymond Fisman, Ilyana Kuziemko, and Kewei Zhang. Reference points and redistributive preferences: Experimental evidence. *Journal of Public Economics*, 216:104761, 2022.
- 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.

- Thomas Cornelissen, Christian Dustmann, Anna Raute, and Uta Schönberg. From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41:47–60, 2016.
- 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.
- DTDA. Labour market profile georgia - 2021. Technical report, Danish Trade Union Development Agency, 2021.
- Econometría. Impact evaluation of targeted social assistance (tsa) in georgia. final report. Technical report, Econometría 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 Solé-Ollé. ‘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.
- Francois Gerard, Joana Naritomi, and Joana Silva. Cash transfers and formal labor markets: Evidence from brazil. CEPR Discussion Papers 16286, C.E.P.R. Discussion Papers, 2021. URL <https://EconPapers.repec.org/RePEc:cpr:ceprdp:16286>.
- Michael Geruso and Timothy Layton. Upcoding: evidence from medicare on squishy risk adjustment. *Journal of Political Economy*, 128(3):984–1026, 2020.
- James J. Heckman and Edward J. Vytlacil. Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. volume 6 of *Handbook of Econometrics*, pages 4875–5143. Elsevier, 2007.
- 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.
- 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 Salvanes, and Emma Tominey. Mothers working during preschool years and child skills: Does income compensate? *Journal of Labor Economics*, 0(ja):null, Forthcoming.
- 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.

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

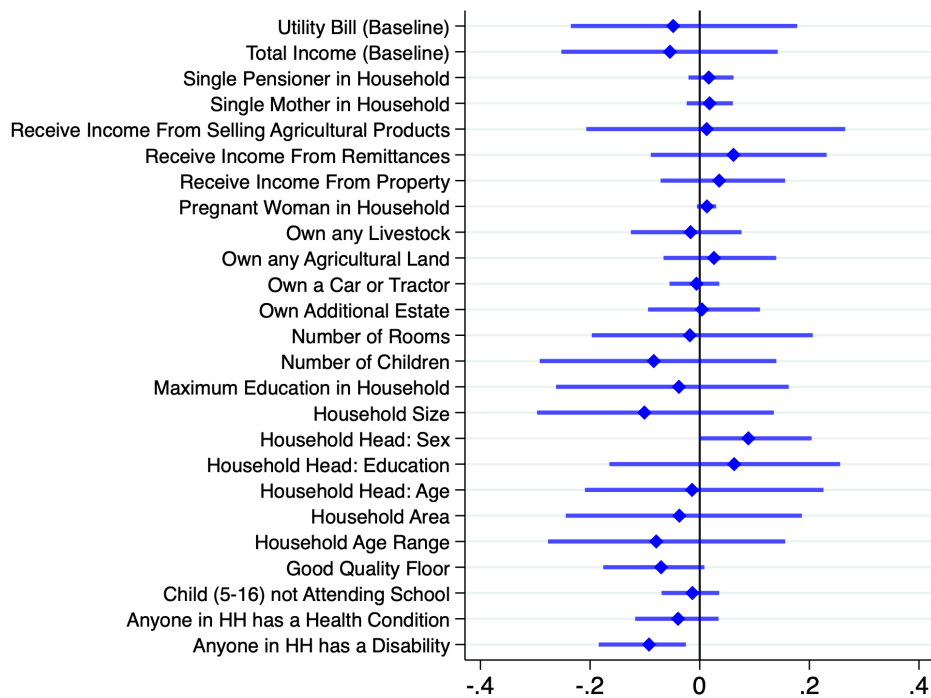
The key identifying assumption in a RD design is the continuity assumption, which states that the potential outcomes ($Y_{0,i}$ and $Y_{1,i}$) are smooth functions of the running variable $z_{0,i}$ through the cutoff, κ .

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

$$X_i = \lambda D_i + g_1^D(z_{0,i}) + v_i \quad (11)$$

where $g_1^D(z_{0,i})$ 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. 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 65,000 PMT Threshold



Notes: TBC.

A.3 Linking the two key Endogenous Regressors

It is instructive to see how one can move between equations based on final welfare income, $B_{1,i}$, and a welfare eligibility manipulation attempt, R_i . We can write the final benefit amount as function of the initial benefit amount and the process of reassessment:

$$B_{1,i} = R_i(p_i(1 - q_i)\Delta B_i^+ + q_i\Delta B_i^-) + B_{0,i} \quad (12)$$

where ΔB_i^+ is the increase in benefits if the reassessment is successful, and ΔB_i^- the reduction in benefits if the SSA sanctions the household for welfare fraud. Here, we restate Equation (10), with different parameters to distinguish this instance from the version in the body of the text:

$$Y_i = \theta_0 B_{1,i} + g_3^D(z_{0,i}) + \theta_1 B_{0,i} + X_i' \Theta + \zeta_i \quad (13)$$

Substituting in Equation (12) in Equation (13) and rearranging, we can connect it with Equation (8).

$$\begin{aligned} Y_i &= [\theta_0(p_i(1 - q_i)\delta B_i^+ + q_i\delta B_i^-)]R_i + g_3^D(z_{0,i}) + (\theta_0 + \theta_1)B_{0,i} + X_i' \Theta + \zeta_i \\ &= \Gamma_{0,i}R_i + g_3^D(z_{0,i}) + \Gamma_1 B_{0,i} + X_i' \Theta + \zeta_i \end{aligned} \quad (14)$$

A.4 Sample-Size and First Stage F -Statistics

For both endogenous treatment variables, the estimate first stage coefficient on the instrument is extremely stable across all data settings (Panel (A) of Figure A2 and Figure A3). What drives the variation in the first stage F -statistic across our various data settings is the sample size in the different settings. One can see the first stage F -statistic (Panel (C) of Figure A2 and Figure A3) moving in tandem with the sample size (Panel (B) of Figure A2 and Figure A3).

Figure A2: First Stage Estimates, Sample Size and Instrument Relevance for Repeat Interview

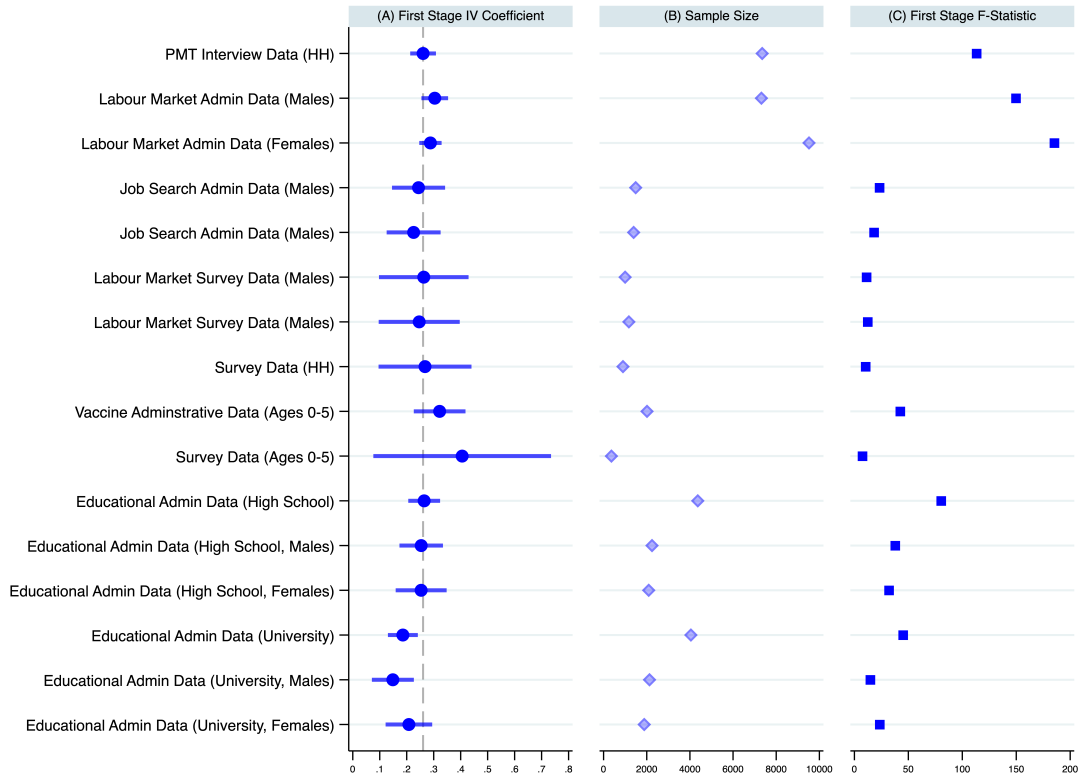
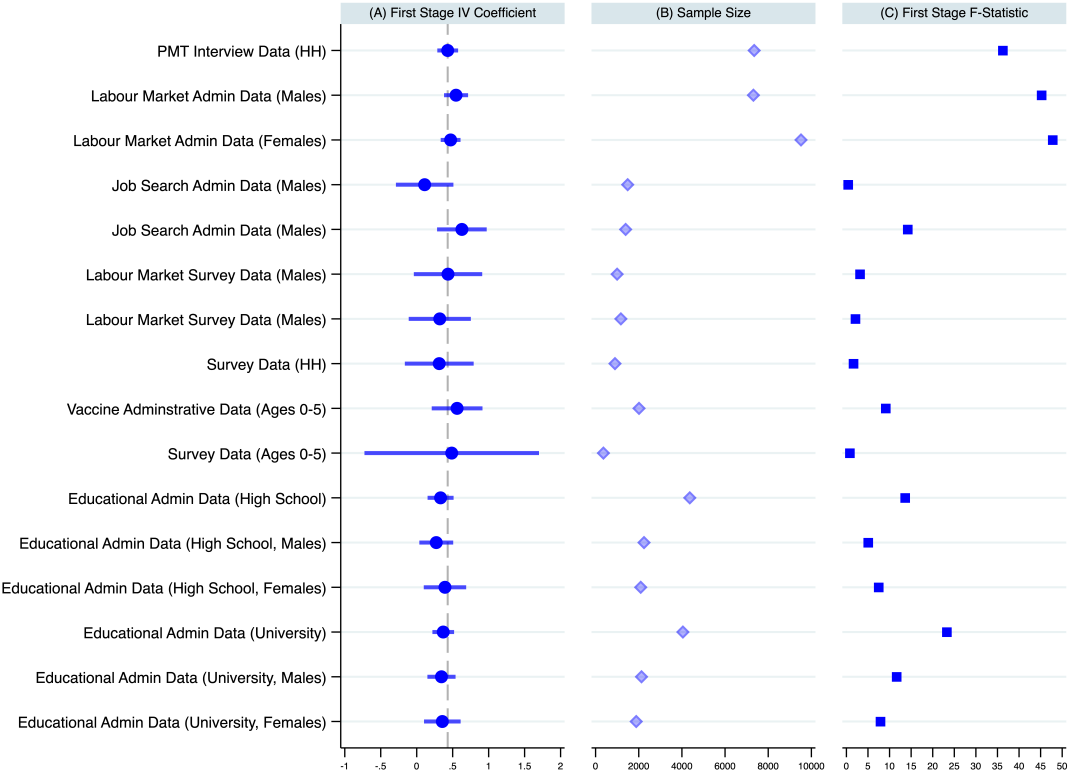
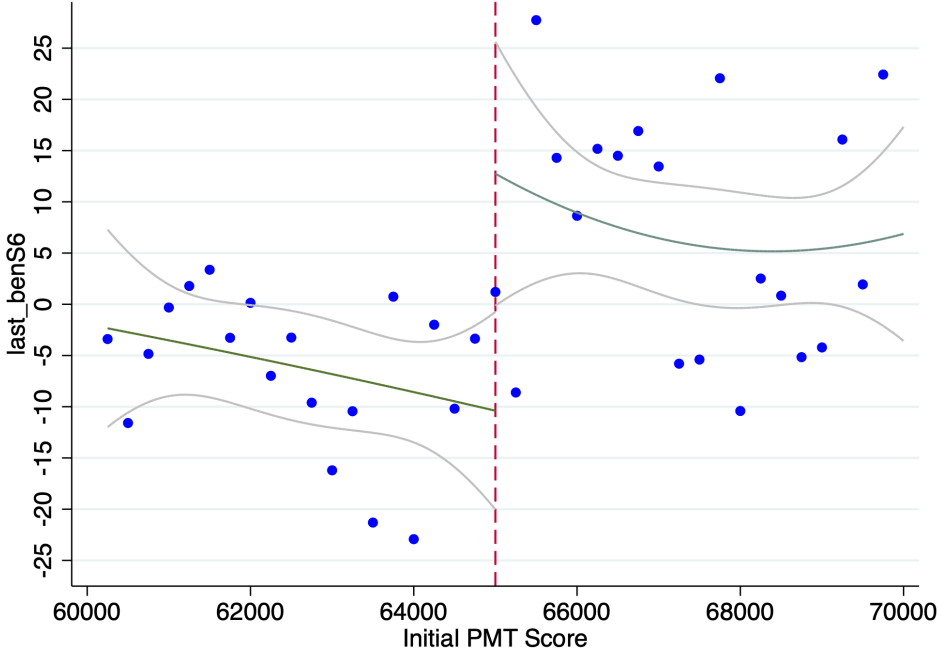


Figure A3: First Stage Estimates, Sample Size and Instrument Relevance for Realized Benefits



A.5 Linking Initial Welfare Income and Final Welfare Income Around the Cutoff

Figure A4: Conditional on initial benefits, final benefits are higher for those above 65,000



(a) Final Benefit

Notes: TBC.

A.6 Additional Formal Labor Market Results

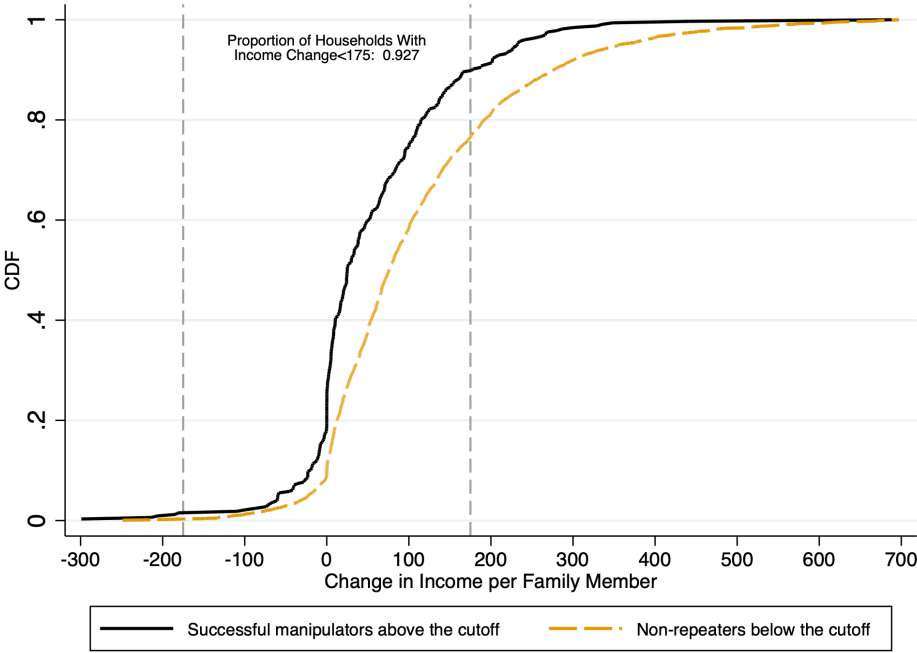
Table A2: Welfare Eligibility Manipulation Leads to Increased Formal Labor Market Engagement

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
	Mean Income	Employed All Periods	Employed At Least Once	Mean Income	Employed All Periods	Employed At Least Once
OLS						
Realized Benefit	-35.932*** (3.304)	-0.016*** (0.003)	-0.042*** (0.004)	-11.454*** (1.107)	-0.012*** (0.002)	-0.028*** (0.003)
CW-OLS						
Realized Benefit	-33.006*** (3.530)	-0.013*** (0.003)	-0.039*** (0.004)	-8.713*** (1.046)	-0.010*** (0.002)	-0.024*** (0.003)
Reduced Form						
$\mathbb{1}[z_{0,i} \geq \kappa]$	98.846*** (21.311)	0.059*** (0.018)	0.046 (0.030)	14.116* (8.341)	0.011 (0.014)	0.032 (0.022)
2SLS						
Realized Benefit	180.630*** (50.167)	0.107*** (0.037)	0.084 (0.056)	29.932 (18.952)	0.022 (0.029)	0.068 (0.048)
First Stage F-Stat	45.209	45.209	45.209	47.783	47.783	47.783
\bar{Y}_0	126.016	0.078	0.279	51.303	0.061	0.198
Observations	7,314	7,314	7,314	9,519	9,519	9,519

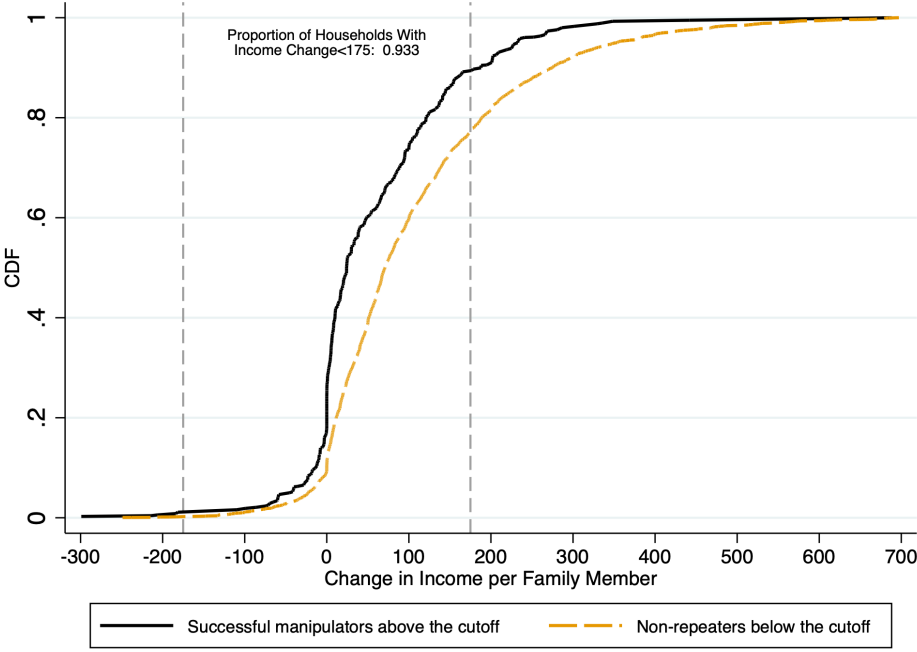
Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Realized Benefit is measured in 100s of Lari.

A.7 Manipulation and Subsequent Income Changes

Figure A5: The Formal Labor Income Increases of Successful Manipulators Fall Almost Exclusively 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.8 Potential Drivers of the Formal Labor Market Results

A.8.1 Job Search Behavior

In order to conduct the job search-related analysis we present in this section, we pool all available data we have on employment spells and formal income. The formal sector employment data we have provides us with four snapshots (periods 1-4) of individuals' formal sector employment status as well as formal sector earnings. We complement this data with the employment information in the most recent PMT interview prior to the first of our four formal sector snapshots (an additional period – period 0).

Using the five periods of data, we can construct workers' employment spells during up to five distinct periods²⁰ spanning up to 5 years.

²⁰In order to isolate the effect of the interview, we consider employment spells after the final interview only.

A.8.2 Transportation and Communication Barriers

Table A3: Transport Expenditure and Means of Transport

	(1)	(2)	(3)	(4)	(5)
	Expenditure on:		Owns a:		
	Communication	Transport	Bicycle	Motorcycle	Car
OLS:	-2.105*	-4.485	-0.001	-0.001	-0.037
Repeat Interview	(1.224)	(3.187)	(0.017)	(0.006)	(0.027)
CW-OLS:	-1.216	-5.327	-0.015	-0.005	-0.050*
Repeat Interview	(1.311)	(3.405)	(0.019)	(0.006)	(0.027)
Reduced Form:	6.635**	5.266	0.055*	0.008	0.048
$\mathbb{1}[z_{1,i} \geq \kappa]$	(2.799)	(8.176)	(0.033)	(0.014)	(0.059)
2SLS:	24.772*	19.659	0.204	0.031	0.180
Repeat Interview	(13.505)	(30.773)	(0.147)	(0.059)	(0.230)
First Stage F-Stat	10.338	10.338	10.338	10.338	10.338
\bar{Y}_0	15.423	25.882	0.033	0.006	0.112
Observations	900	900	900	900	900

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

A.9 Marginal Treatment Effects

In this section we estimate marginal treatment effects (MTEs) based on our binary instrument. We present a brief overview of the MTE approach in order to anchor what we do here, then proceed to discuss how we may identify the MTE when we have only a binary instrument. We follow the exposition of Andresen (2018), and additionally use his Stata command – `mtefe` – in order to estimate the MTEs.

The starting point for all MTE approaches is the generalized Roy model, which takes the form of:

$$Y_j = \mu_j(X) + U_j = X\beta_j + U_j \quad \text{for } j = 0, 1 \quad (15)$$

$$Y = DY_1 + (1 - D)Y_0 \quad (16)$$

$$D = \mathbb{1}[\mu_D(Z) > V] = \mathbb{1}[Z\gamma > V] \quad \text{where } Z = (X, Z_-) \quad (17)$$

Y_0 and Y_1 are respective outcomes in untreated (no manipulation) and treated (manipulation) states.

The unobservable, V , in (17) – which we view as the unobserved resistance to treatment status – plays a core role in the interpretation of MTEs. If V is continuously distributed, we can rewrite (17) as $D = \mathbb{1}[P(Z) > U_D]$, where $P(Z)$ is the propensity score and U_D represents quantiles of V and is uniformly distributed.

We follow Brinch et al. (2017), who show how to identify the MTE in the presence of a binary instrument using the separate MTE estimation approach developed by Heckman and Vytlacil (2007). This involves separately estimating the conditional expectation of the outcome variable for the treated and untreated samples separately, combined with an appropriate function for the conditional expectation of the error terms:

$$\mathbb{E}(Y_0 \mid X = x, D = 0) = x\beta_0 + E(U_0 \mid U_D > p) = x\beta_0 + K_0(p) \quad (18)$$

$$\mathbb{E}(Y_1 \mid X = x, D = 1) = x\beta_1 + E(U_1 \mid U_D \leq p) = x\beta_0 + K_1(p), \quad (19)$$

where $p = P(Z)$. Brinch et al. (2017) describe in detail the identification challenges posed when attempting to use a binary instrument to estimate MTEs, show how the separate approach with linear specifications for K_0 and K_1 enables identification of the MTE, and present a graphical representation of the geometry of the linear MTE. We leave the interested reader to view that paper for further detail.

A key output of most authors' MTE analysis is a plot of the MTE estimate along the distribution of U_D – the quantiles of the unobserved resistance to treatment. For an example, see the discussion of empirical applications in Cornelissen et al. (2016), where the slope of the MTE curve takes center stage in the discussion of the two papers' results. There is good reason for this – the slope of the MTE curve informs us about patterns of selection on unobservable gains to treatment. Where the MTE curve is downward-sloping, this indicates positive selection on unobservable gains – those with the highest marginal treatment effect are those with the lowest unobserved resistance to treatment, hence conditional on observables, the most likely to take up treatment. The converse applies.

In Table A4, we present the key outputs of our MTE analysis for formal labor market outcomes, considering three different functional form specifications for the choice equation, (17). Other dimensions on which authors often consider heterogeneity analysis – the orders of polynomials used in the specification of $K()$ and the type of MTE approach – are not permissible in our binary instrument case. Here we must use the separate approach, and $K()$ must be linear (Brinch et al., 2017). In lieu of 18 MTE curves, we extract the slope of the term in p , and the p -value of this term from each of our MTE specifications. This is a sufficient statistic for the key information contained in the linear MTE. As one can see from Table A4, there is no meaningful selection on unobservable gains to treatment for men nor women, for any of the outcomes we consider. What this means is that households are not manipulating welfare eligibility based on the (unobserved) labor market benefits of doing so.

Table A4: Marginal Treatment Effect Analysis Reveals That Households do not Manipulate Welfare Eligibility due to the Unobserved Labor Market Gains to Manipulation.

	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
	Choice Model Function:			Choice Model Function:		
	LPM	Logit	Probit	LPM	Logit	Probit
A. Mean Income						
ATE	73.844 (111.977)	63.875 (67.366)	65.912 (72.788)	5.170 (45.393)	-5.196 (29.576)	-7.691 (32.331)
ATT	273.822*** (75.673)	116.232** (50.229)	135.156** (53.457)	62.922* (33.476)	24.812 (21.228)	31.344 (22.742)
ATUT	14.018 (143.032)	49.100 (84.598)	46.237 (91.656)	-11.655 (57.338)	-13.514 (36.973)	-18.574 (40.545)
LATE	259.317*** (61.734)	152.809*** (44.949)	167.729*** (48.118)	44.077* (25.797)	13.473 (17.836)	18.209 (19.108)
<i>p</i> -value: Observable Heterogeneity	0	0	0	0	0	0
<i>p</i> -value: Essential Heterogeneity	0.073	0.272	0.229	0.242	0.328	0.246
Selection on Unobservable Gains	Positive*	Positive	Positive	Positive	Positive	Positive
Observations	7,187	7,187	7,187	9,421	9,421	9,421
B. Employed All Periods						
ATE	0.147 (0.119)	0.099 (0.071)	0.103 (0.077)	0.010 (0.093)	0 (0.059)	-0.007 (0.065)
ATT	0.142** (0.062)	0.081* (0.042)	0.087* (0.045)	0.031 (0.054)	0.014 (0.034)	0.021 (0.037)
ATUT	0.148 (0.152)	0.104 (0.089)	0.107 (0.097)	0.003 (0.118)	-0.004 (0.074)	-0.015 (0.082)
LATE	0.154*** (0.052)	0.111*** (0.040)	0.117*** (0.042)	0.027 (0.044)	0.013 (0.031)	0.017 (0.033)
<i>p</i> -value: Observable Heterogeneity	0	0	0	0	0	0
<i>p</i> -value: Essential Heterogeneity	0.954	0.990	0.994	0.786	0.756	0.617
Selection on Unobservable Gains	Positive	Negative	Negative	Positive	Positive	Positive
Observations	7,187	7,187	7,187	9,421	9,421	9,421
C. Employed At Least Once						
ATE	0.029 (0.176)	0.036 (0.110)	0.041 (0.118)	0.014 (0.144)	-0.019 (0.091)	-0.027 (0.100)
ATT	0.090 (0.108)	0.025 (0.072)	0.032 (0.076)	0.054 (0.087)	0.021 (0.058)	0.032 (0.062)
ATUT	0.010 (0.224)	0.039 (0.138)	0.044 (0.149)	0.002 (0.183)	-0.030 (0.114)	-0.043 (0.125)
LATE	0.109 (0.088)	0.069 (0.065)	0.075 (0.069)	0.071 (0.071)	0.041 (0.052)	0.048 (0.055)
<i>p</i> -value: Observable Heterogeneity	0	0	0	0	0	0
<i>p</i> -value: Essential Heterogeneity	0.693	0.952	0.957	0.781	0.657	0.560
Selection on Unobservable Gains	Positive	Positive	Positive	Positive	Positive	Positive
Observations	7,187	7,187	7,187	9,421	9,421	9,421

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

A.10 Additional Labor Market Results

Table A5: We Document the Same Labor Market Patterns Using Final Benefits Instead of Repeat Interview as Our Treatment Variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Employed				Hours per Week		
	Total	For a Wage	Self-Employed	Unpaid Home Work	Total	For a Wage	Self-Employed
A. Men							
OLS:	-0.033**	-0.043***	0.010	-0.003	-2.101***	-2.218***	0.116
Final Benefit	(0.014)	(0.011)	(0.013)	(0.002)	(0.541)	(0.470)	(0.414)
CW-OLS:	-0.034*	-0.034**	-0.000	-0.000	-2.115***	-2.132***	0.017
Final Benefit	(0.019)	(0.016)	(0.019)	(0.001)	(0.750)	(0.738)	(0.541)
Reduced Form:	0.133	0.146*	-0.012	-0.007	8.599**	6.744*	1.855
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.100)	(0.077)	(0.097)	(0.010)	(4.025)	(3.643)	(3.062)
2SLS:	0.307	0.334	-0.028	-0.015	19.752	15.491	4.261
Final Benefit	(0.294)	(0.269)	(0.218)	(0.027)	(15.156)	(12.596)	(7.160)
First Stage F-Stat	3.131	3.131	3.131	3.131	3.131	3.131	3.131
\bar{Y}_0	0.549	0.162	0.387	0.001	16.616	7.141	9.475
Observations	999	999	999	999	999	999	999
B. Women							
OLS:	-0.005	-0.020**	0.015*	0.008	-0.592	-0.717*	0.125
Final Benefit	(0.011)	(0.009)	(0.009)	(0.013)	(0.429)	(0.401)	(0.219)
CW-OLS:	0.004	-0.007	0.011	0.009	-0.025	-0.016	-0.009
Final Benefit	(0.017)	(0.012)	(0.013)	(0.017)	(0.583)	(0.537)	(0.334)
Reduced Form:	0.014	-0.035	0.048	-0.068	-0.589	-1.999	1.410
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.083)	(0.068)	(0.063)	(0.085)	(2.901)	(2.761)	(1.409)
2SLS:	0.043	-0.108	0.151	-0.213	-1.836	-6.233	4.398
Final Benefit	(0.253)	(0.203)	(0.216)	(0.305)	(8.950)	(9.128)	(5.449)
First Stage F-Stat	2.074	2.074	2.074	2.074	2.074	2.074	2.074
\bar{Y}_0	0.299	0.149	0.150	0.513	8.740	5.673	3.067
Observations	1,173	1,173	1,173	1,173	1,173	1,173	1,173

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

Table A6: Our Key Estimates for Waged Work Appear to be Driven by the Formal Sector – Repeat Assessment

	(1)	(2)	(3)	(4)	(5)	(6)
	Employed			Hours per Week		
	For a Wage	For a Wage (Formal)	For a Wage (Informal)	For a Wage	For a Wage (Formal)	For a Wage (Informal)
A. Men						
OLS:	0.017	-0.001	0.018	-0.531	-0.898	0.368
Repeat Interview	(0.032)	(0.027)	(0.018)	(1.462)	(1.311)	(0.768)
CW-OLS:	0.016	-0.002	0.018	-0.607	-0.991	0.384
Repeat Interview	(0.032)	(0.028)	(0.019)	(1.474)	(1.325)	(0.777)
Reduced Form:	0.146*	0.108	0.037	6.744*	4.262	2.482
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.077)	(0.068)	(0.045)	(3.643)	(3.284)	(1.963)
2SLS:	0.553*	0.412	0.141	25.624*	16.193	9.431
Repeat Interview	(0.318)	(0.266)	(0.166)	(15.007)	(12.607)	(7.094)
First Stage F-Stat	11.200	11.200	11.200	11.200	11.200	11.200
\bar{Y}_0	0.162	0.114	0.048	7.141	5.328	1.813
Observations	999	999	999	999	999	999
B. Women						
OLS:	0.000	-0.014	0.014	0.566	-0.276	0.843
Repeat Interview	(0.030)	(0.026)	(0.016)	(1.324)	(1.077)	(0.836)
CW-OLS:	0.005	-0.007	0.012	0.632	-0.285	0.916
Repeat Interview	(0.032)	(0.027)	(0.018)	(1.460)	(1.127)	(0.978)
Reduced Form:	-0.035	-0.037	0.003	-1.999	-1.177	-0.822
$\mathbb{1}[z_{1,i} \geq \kappa]$	(0.068)	(0.060)	(0.037)	(2.761)	(2.418)	(1.505)
2SLS:	-0.141	-0.151	0.011	-8.116	-4.779	-3.336
Repeat Interview	(0.256)	(0.224)	(0.144)	(11.088)	(9.153)	(6.673)
First Stage F-Stat	12.262	12.262	12.262	12.262	12.262	12.262
\bar{Y}_0	0.149	0.112	0.037	5.673	4.178	1.495
Observations	1,173	1,173	1,173	1,173	1,173	1,173

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

Table A7: Our Key Estimates for Waged Work Appear to be Driven by the Formal Sector – Final Benefit Level

	(1)	(2)	(3)	(4)	(5)	(6)
	Employed			Hours per Week		
	For a Wage	For a Wage (Formal)	For a Wage (Informal)	For a Wage	For a Wage (Formal)	For a Wage (Informal)
A. Men						
OLS:						
Final Benefit	-0.043*** (0.011)	-0.044*** (0.009)	0.001 (0.007)	-2.218*** (0.470)	-2.040*** (0.419)	-0.178 (0.257)
CW-OLS:						
Final Benefit	-0.034** (0.016)	-0.045*** (0.013)	0.011 (0.010)	-2.132*** (0.738)	-2.152*** (0.620)	0.020 (0.409)
Reduced Form:						
$\mathbb{1}[z_{1,i} \geq \kappa]$	0.146* (0.077)	0.108 (0.068)	0.037 (0.045)	6.744* (3.643)	4.262 (3.284)	2.482 (1.963)
2SLS:						
Final Benefit	0.334 (0.269)	0.249 (0.217)	0.086 (0.109)	15.491 (12.596)	9.790 (9.591)	5.702 (5.173)
First Stage F-Stat	3.131	3.131	3.131	3.131	3.131	3.131
\bar{Y}_0	0.162	0.114	0.048	7.141	5.328	1.813
Observations	999	999	999	999	999	999
B. Women						
OLS:						
Final Benefit	-0.020** (0.009)	-0.019** (0.008)	-0.001 (0.005)	-0.717* (0.401)	-0.672** (0.314)	-0.045 (0.279)
CW-OLS:						
Final Benefit	-0.007 (0.012)	-0.015 (0.010)	0.007 (0.009)	-0.016 (0.537)	-0.512 (0.379)	0.496 (0.440)
Reduced Form:						
$\mathbb{1}[z_{1,i} \geq \kappa]$	-0.035 (0.068)	-0.037 (0.060)	0.003 (0.037)	-1.999 (2.761)	-1.177 (2.418)	-0.822 (1.505)
2SLS:						
Final Benefit	-0.108 (0.203)	-0.116 (0.182)	0.008 (0.111)	-6.233 (9.128)	-3.671 (7.250)	-2.562 (5.340)
First Stage F-Stat	2.074	2.074	2.074	2.074	2.074	2.074
\bar{Y}_0	0.149	0.112	0.037	5.673	4.178	1.495
Observations	1,173	1,173	1,173	1,173	1,173	1,173

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