Centre of Excellence for Development Impact and Learning

February 2023

Research project paper 4

An empirically driven theory of poverty reduction

Sudhanshu Handa, Zhiyuan Liu, Gelson Tembo, Clement Adamba, Peter Mvula

London International Development Centre
UK aid from the British people
Oxford Policy Management

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL research project paper

CEDIL research project papers are the research outputs of the projects supported by CEDIL. They include impact evaluations, systematic reviews, evidence and gap maps and exploratory papers on research methods and approaches.

About this research project paper

This paper was submitted to CEDIL by the "An empirically driven theory of poverty reduction" L.191 team.

Please direct any comments or queries to the corresponding author, Sudhanshu Handa at shanda@email.unc.edu

Suggested citation: Handa, S., Liu, Z., Tembo, G., Adamba, C. & Mvula, P. (2023) 'An empirically driven theory of poverty reduction', CEDIL Research Project Paper 4. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford. Available from: https://doi.org/10.51744/CRPP4

This project was funded by the Centre of Excellence for Development Impact and Learning (CEDIL), supported by UK aid from the UK Government. The views expressed in this research project paper do not necessarily reflect the UK Government's official policies or CEDIL.

1. Contents

1. Introduction

The persistence of poverty, particularly in sub-Saharan Africa, means that public policy in this region continues to debate the right mix of supply- versus demand-side interventions that can move large groups of households out of extreme or ultra-poverty. Recent rigorous evidence on the microfinance model confirms that this approach has over-promised and under-performed in moving large numbers out of poverty. The next new innovation comes from Bangladesh through the NGO BRAC, who provide a big- push intervention designed to 'graduate' households from poverty.[1] Proponents of these 'graduation programs' claim that a big push at the household level, comprising a cash transfer, asset transfer, skills training and coaching can graduate households permanently into the middle-class. These interventions are expensive and complex, and have yet to be actually implemented by a national government, which is telling. Others have argued that rather than context and skills, decision-making under scarcity and the psychological toll of poverty are key barriers to economic growth.[2]

The reality is that there is unlikely to be a single approach that can transition all poor or ultra-poor households out of poverty. Some (but not all) households have the entrepreneurial talent such that a micro loan is the missing ingredient to graduation, other households, due to their talents and environment would instead benefit from extension services and input support, and still others could use an injection of cash to finance migration for wage work. And of course some households are unlikely to be able to move out of poverty at all (e.g. the elderly poor, those with disabilities, or living in extremely harsh environments). Given this reality, an unconditional cash transfer targeted to the ultra-poor can provide useful insights about what households themselves view as their most pressing constraints, and where they feel the returns to investment are highest given their circumstances and skills. In this paper, we use secondary evaluation data from four government unconditional cash transfer programs (UCTs) to identify high- and low-flyers, that is, those households that are able to use the income shock to significantly improve their living standards and those who aren't. We attempt to categorize the high- and low-flyers to create typologies based on their pre-shock characteristics. Then we look at their post-treatment behaviors to see what they did with the cash to improve (or not) their living standards. Putting together these different pieces of information (pre-treatment characteristics and post-treatment behaviors) can help us understand the different pathways out of poverty, and ultimately contribute to a middle-range theory of sustained poverty reduction.

The four programs, while all national cash transfer programs implemented by their respective governments, have some important differences that can help us understand the graduation potential of ultra-poor households. The Zambia and Ghana programs were targeted to ultra-

---

[1] Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry(2015) "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." Science 348, no. 6236 (May 14, 2015): 1260799–1260799.

[2] Haushofer, J., & Fehr, E. (2014). "On the Psychology of Poverty." Science 6186: 862-867.

poor households with very young children, under age 5 (Zambia) or one year of age or with a member who was pregnant (Ghana). In contract, the Zimbabwe and Malawi programs targeted ultra-poor labor-constrained households, those with high dependency ratios, which de facto leads to a much older beneficiary, mostly women, with very few able-bodied adults. This clearly limits the graduation potential of beneficiaries; these variations in the target population provide additional variation to help us understand how the poor spend money and the graduation potential of poverty-targeted cash transfer programs.

2. Data

We use impact evaluation data from four national UCTs. The countries and programs are: 1) Malawi Social Cash Transfer Program; 2) Zimbabwe Harmonized SocialCash Transfer Program; 3) Ghana Livelihood Empowerment Against Poverty; 4) Zambia Child Grant Program. All evaluations include one baseline and multiple follow-ups (except for Ghana with just one follow-up). In Malawi and Zambia the design is an RCT, in Ghana it is a discontinuity design using the proxy means test cut-off, in Zimbabwe the design is matched Wards (administrative units below the district) followed by the application of household targeting by the program so all households in the comparison Wards are future eligible households. In all four cases, treatment can be considered exogenous, and extensive baseline balance tests confirm the fidelity of the original designs. Table 1 summarizes the data sets and survey years, all initial surveys are baseline, years in bold are the ones used in the analysis. All impact evaluation reports are available at the Transfer Project website (https://transfer.cpc.unc.edu/) and report impact estimates across all major productive, social and family domains, typically over 100 indicators. The results from Ghana, Malawi and Zimbabwe are published in Handa et al (2021),[3] those from Zambia are published in Handa et al (2020).[4]

Table 1: Sample sizes and survey years (first year is pre-treatment)

	Treatment	Control	Survey years
Ghana LEAP	1,262	1,235	**2015**, **2017**
Malawi SCTP	1,730	1,800	**2013**, 2014, **2015**
Zambia CGP	1,260	1,252	**2010**, 2012, 2013, **2014**
Zimbabwe HSCT	1,029	1,034	**2013**, 2014, **2017**

[3] Sudhanshu Handa, Frank Otchere, Paul Sirma* on behalf of the Ghana LEAP, Malawi SCTP and Zimbabwe HSCT Evaluation Teams, 2022, "More Evidence on the Impact of Government Social Protection in Sub Saharan Africa," Development Policy Review Vol.40(3).

[4] Sudhanshu Handa, Luisa Natali, David Seidenfeld, Gelson Tembo and Benjamin Davis, 2018, "Can unconditional cashtransfers raise long-term living standards? Evidence from Zambia, "Journal of Development Economics, Vol 133(July): 42-65.

All survey instruments are comprehensive multi-topic questionnaires (also all available on the website), including the full consumption module taken from therespective national living conditions surveys in each country. The data contain psychological and behavioral measures such as affect, subjective well-being, optimism of the future, and time-discounting, as well as livelihoods and productive activity, crop production, livestock, credit, and domestic assets. We collected GPS points on households in each country which we used to bring in secondary data on the microenvironment of the household (land-use, Palmer dryness index, distance to district capital).

Our choice of variables is guided by theory as well as our prior published work on the overall impacts of these four UCTs. In the published versions of the impact results we have compiled broad domain indexes such as Income and Revenue, Finance and Debt, Food Security, Consumption, Schooling, Health and Nutrition, building each domain index with individual indicators representing different outcome areas within that domain. For this analysis, as the objective is to understand the drivers of changes in living standards (measured by consumption) rather than overall program impacts, we identify the five overarching domains of Finance and Debt, Income and Revenue, Assets, Psychological States and Environment. We also use individual characteristics of the household. The domains and associated indicators are presented in Table 2 for each country and wave. As indicators have different units of measure, we redefine them so that higher values are better, and convert each indicator to a z-score at each wave. Domain indexes are then constructed by taking the average of the z-scores of each individual indicator within that domain. Not all countries have all indicators at each wave (this is mainly an issue for the baseline). Baseline (pre-treatment) measures are used for predicting treatment effects.

While all four programs target the ultra-poor and are some of the best targeted programs in their respective countries, there are other important differences in eligibility that have implications for our analysis. The evaluation in Ghana is of the LEAP 1000 window of the overall LEAP, which specifically targets pregnant women and those with a child 12 months or younger. Similarly, the Zambian Child grant program targets families with a child under age 3 years. In these two programs, the household is at a much younger stage of the lifecycle, the recipient is relatively young and almost all are women. The Malawi and Zimbabwe programs on the other hand target 'labor-contrained' households, those with high dependency ratios, defined as the ratio of the number of elderly and children to able-bodied members. The typical recipient in these programs I over age 50, about 65 percent are women, with very few preschool children and many more teenagers and orphans. These distinct demographic profiles (Ghana, Zambia versus Malawi, Zimbabwe) add an additional layer of variation to the analysis. Finally, from an operational perspective a key parameter is the value of the transfer and its predictability. The Malawi and Zambia programs perform the best on this score, with larger relative transfer sizes (see the articles in footnotes 3 and 4 for further details about this) and consistent payments. The relative transfer value was lower in Ghana and Zimbabwe, and Zimbabwe in particular saw significant gaps in payments, so while households received the full amount due to them, there was a two year period during the four year study period where transfers were lumpy, consisting of double or even triple payments. In Ghana, transfers were paid on time the value was low. These operational features closely track the overall impacts of the programs as discussed in Handa et al (2022). Impacts were large and widespread in malawi and Zambia, and even generated large multiplier effects, while in Zimbabwe and Ghana they were much smaller and limited in range.

3. Identifying treatment heterogeneity using causal trees

The cash transfer in our data is used as an exogenous liquidity shock, and our main objective is to classify households into 'high flyers', those who were able to convert the cash transfer into large gains in consumption, and moderate and low (or perhaps negative) achievers. A simple but naïve approach would be to look at the endline consumption of treatment households and pick those in the top 10 percent (say) of the distribution as high-achievers, and then describe their pre-treatment characteristics. This approach has at least three weaknesses. First, we do not know the counterfactualof those in the top 10th percentile of consumption, perhaps they were always high achievers even before the program (even if we were to use consumption growth to identify high achievers, the issue of the missing counterfactual still exists). Second, we have a large potential pool of characteristics to useto describe these households—which are the most important or salient? Third, we do not how these characteristics may interact with each other to produce high achievers.

We resolve these issues using new methodological results from the intersection of the literatures in ML and causal inference. Athey & Imbens (2016)[5] lay out an approach that allows the data to define the relevant sub-groups for which heterogeneous treatment effects exist within a randomized control trial orother design where the treatment can be considered exogenous, as in our case. The approach uses classification regression trees (CART), a method that partitions or classifies the data based on the degree of homogeneity (or 'purity') of an outcome variable or target. For example, say the target is income and the characteristic is sex (male, female). The CART approach partitions the data into two groups (male, female) if the degree of homogeneity or purity of the outcome (income) in each of the sub-groups is greater than the degree of homogeneity in the overall sample, where purity is measured using the sum of the mean squared deviations for example. This is the intuition behind the CART, and the typical application is when there are a large number of characteristics or features in the data, the outcome or predictor is known, and we want to classify or partition the sample based on groups of characteristics that have the same value of the predictor. This is almost exactly our problem, except that our predictor is not the level of consumption but rather *the level of the treatment effect on consumption*. In other words, we need to solve the problem of the counterfactual by identifying households where the treatment effect is largest—this is the innovation introduced by Athey & Imbens.

We estimate the causal trees using the R Studio scripts provided by Athey on Github, modified appropriately for our data. In the typical run, half the sample (stratified by

[5] Susan Athey & Guido Imbens, 2016, Recursive Partitioning for Heterogeneous Causal Effects, PNAS, Vol.113(27): 7353-7360.

treatment status) is used as the trial sample to train the algorithm, and the remaining half used to estimate the treatment effect of each sub-group (represented by a final leaf) identified by the algorithm. Other diagnostic tests for CART are implemented as appropriate such as cross-validationand pruning. Ultimately, we estimate generalized causal forests using honest estimation and derive our estimates of the conditional (on a vector of observables) average treatment effect from this approach (CATE). Technical details are provided in the appendix.

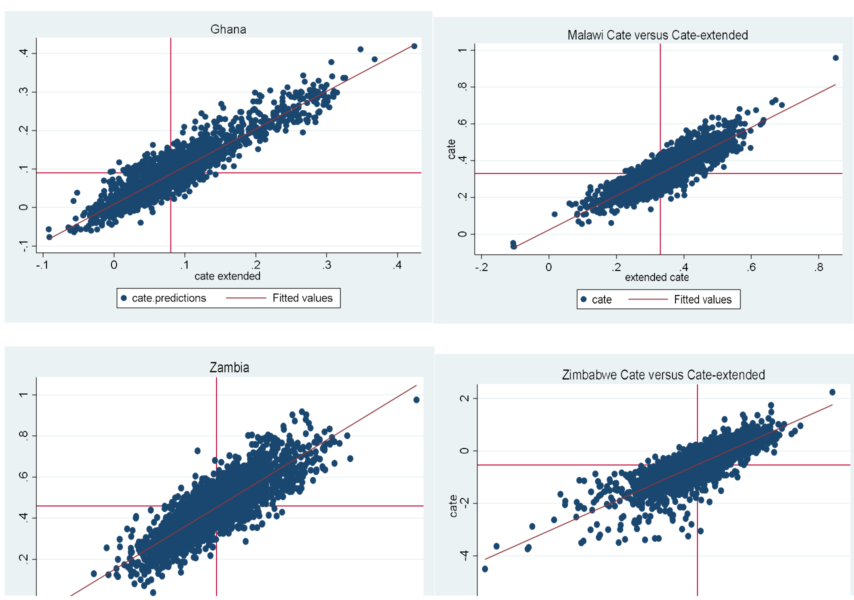
Table 2: Domains and availability of indicators by country and wave

	Baseline				Post-treatment			
Domains and Indicators	MLW	ZAM	ZIM	GHANA	SCTP	ZAM	HSCT	LEAP
Consumption (outcome)								
Overall per capital consumption expenditure	x	X	x	x	x	X	x	x
Assets								
livestock index	x	X	x	x	x	X	X	x
Agricultural asset index	x	X	x		x	X	X	
Domestic asset index	x	X			x	X	X	
Finance/debt								
Any savings	x	X	x	x	x	X	x	x
Amount saved	x	X	x	x	x	X	x	x
Whether household has new loan	x		x	x	x	X	x	x
Reduction in the amount owed	X		X	X	x	X	x	x
Reduction in the amount borrowed	X		X	X	x	X		x
Income & Revenue								
Value of harvest	x	X	x	x	x	X	x	x
Spending on agricultural inputs	x	X	x	x	x	X	x	x
Crop sales		X		X		X	X	X
Operating a non-farm enterprise (NFE)	x		x	x	x	X	x	x
Revenues from NFEs	x		x	x	x	X	x	x
Revenue from livestock sales			X				X	
Psychological state								
Marginal rate of substitution	x	X	X	X	x	X	X	X
Quality of life scale	X		x	X	X		x	X
Think life will be better in either 1, 2, or 3 years	x	X	x	x	x	X	x	x
Environment (included separately)								
Palmer dryness index	X	X	X	X				
Rainfall	X	X	X	X				
Land use/cover (% agriculture)	X	X	X	X				
Distance to district capital	X	X	X	X				
Household level (included separately)								
Age of head	X	X	X	X	X	X	X	X
Number of members	X	X	X	X	X	X	X	X
Schooling level of head	X	X	X	X	X	X	X	X
Sex of head (Ghana, Zambia 99% of recipients are female)		X		X		X		X

4. Results on conditional average treatment effects

We implement the honest causal tree analysis, construct associated generalized random forests and use the random forest to estimate the CATE for each unit, separately by country. The target is either (log) consumption per capita or the change in consumption per capita from baseline to post-treatment follow-up, results are stable and we use the target measured in levels; we also use an alternative measure of living standards, a wealth index composed of both productive and domestic assets. We begin with a smaller set of covariates using the domain indexes plus the household and environmental characteristics entered separately. We also run a specification where we use all the individual indicators instead of the domain indexes. Figure 1 shows the joint distribution of CATES estimated using the full set of individual features versus the domain indexes. The graphs, along with estimated Spearman rank correlation coefficients are close to 0.90 in all four cases suggest there is not much to be gained by using the individual feature list.

Figure 1: Distribution of CATES using parsimonious versus extended features



We now turn to the actual estimates of CATES for each country, which are depicted in Figure 2. Recall that the target is per capita consumption. Based on the full impact evaluations, the respective UCTS were most successful in Malawi and Zambia where there were large, positive impacts across a range of domains, and particularly large consumption impacts. Impacts were more scattered in Ghana and Zimbabwe due to among other things the low value of the transfer and inconsistent payments. This pattern is mirrored in figure 2: median impact is 0.40 and 0.35 SDs in Malawi and Zambia, but much lower in the other two countries. In Malawi and Zambia there were no negative impacts, but almost a fifth of the sample realized negative impacts in Zimbabwe. In that country overall impacts on consumption were not significant, but food security improved, and purchased consumption also improved significantly because the program triggered a reduction in the value of food gifts received by the household, thus muting overall consumption effects. *In the rest of the analysis for Zimbabwe we choose to use purchased consumption rather than total consumption to generate the CATES that we analyze*.[6] In Ghana there are fewer negative impacts and overall consumption impacts were small, though statistically significant.

To what extent do the distributions of CATES represent genuine heterogeneity in impacts? One suggested approach is to partition the estimated CATES into quartiles or quintiles and observe whether the mean of each quartile is progressively increasing.[7] Figure 3 shows box plots of the predicted CATES for each country by quartile. While all the means increase steadily by quartile (as they would by construction), there are clear differences in the degree of progressivity across the quartiles and countries. The dispersion of CATES is highest in Zimbabwe, where there is clear separation in the box plots across the quartiles—this is driven by the large left tail in Zimbabwe, and the fact that nearly all CATES in the first quartile are negative, while in the highest quartile the CATES are around 0.60 SD. The other extreme is Ghana, where there is very little to distinguish between the quartiles, and the CATES range from -0.10 to 0.25. In Ghana we can likely conclude that we are unable to establish treatment heterogeneity. In between these two cases lie Malawi and Zambia, where overall program impacts were quite large. In both countries the predicted CATES range from roughly 0.15 to 0.65, with a bunching of CATES in the middle two quartiles. In these two countries, there may be genuine treatment heterogeneity, particularly between the top and bottom quartiles.

[6] There was no negative impact on transfers received in the other three countries.

[7] Wager, Stefan & Susan Athey. 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. JASA Vol. 113(523): 1228-1242; Athey, Susan, Julie Tibshirani & Stefan Wager. 2019. Generalized Random Forests. The Annals of Statistics. Vol. 47(2): 1148-1178.

Figure 2: CATES by country

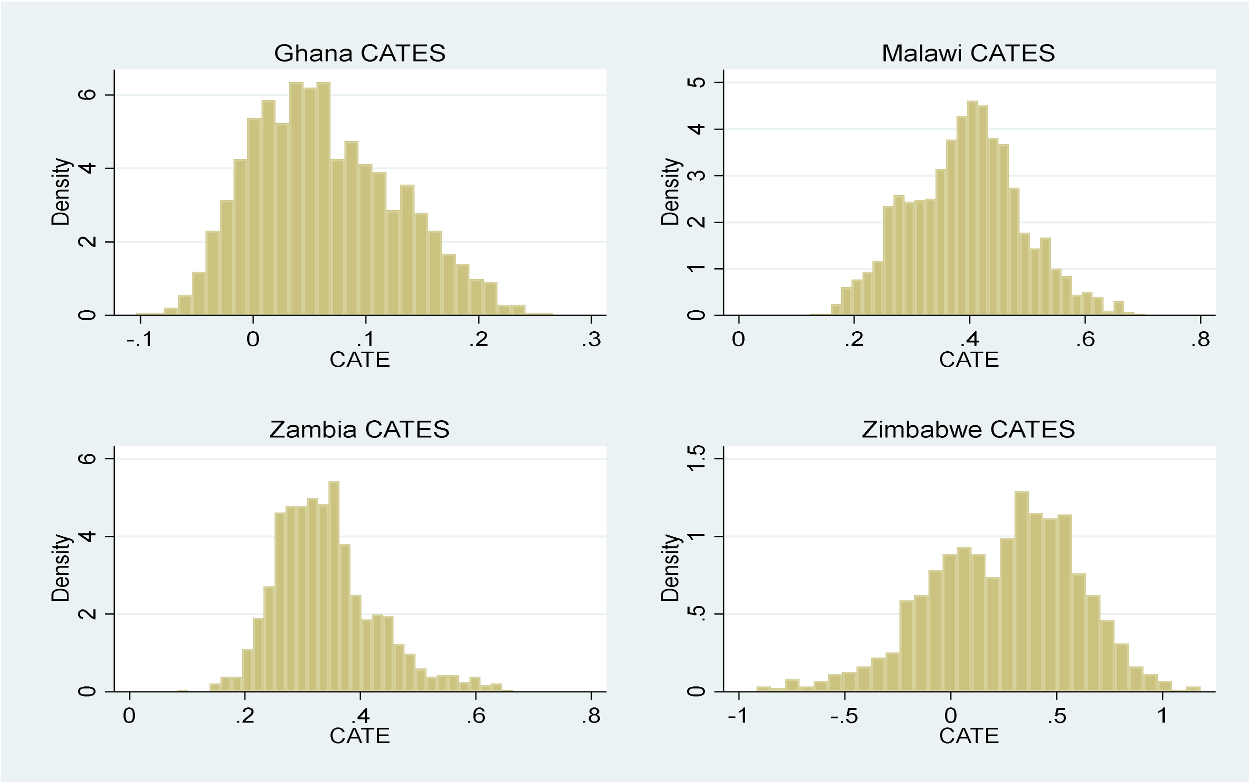
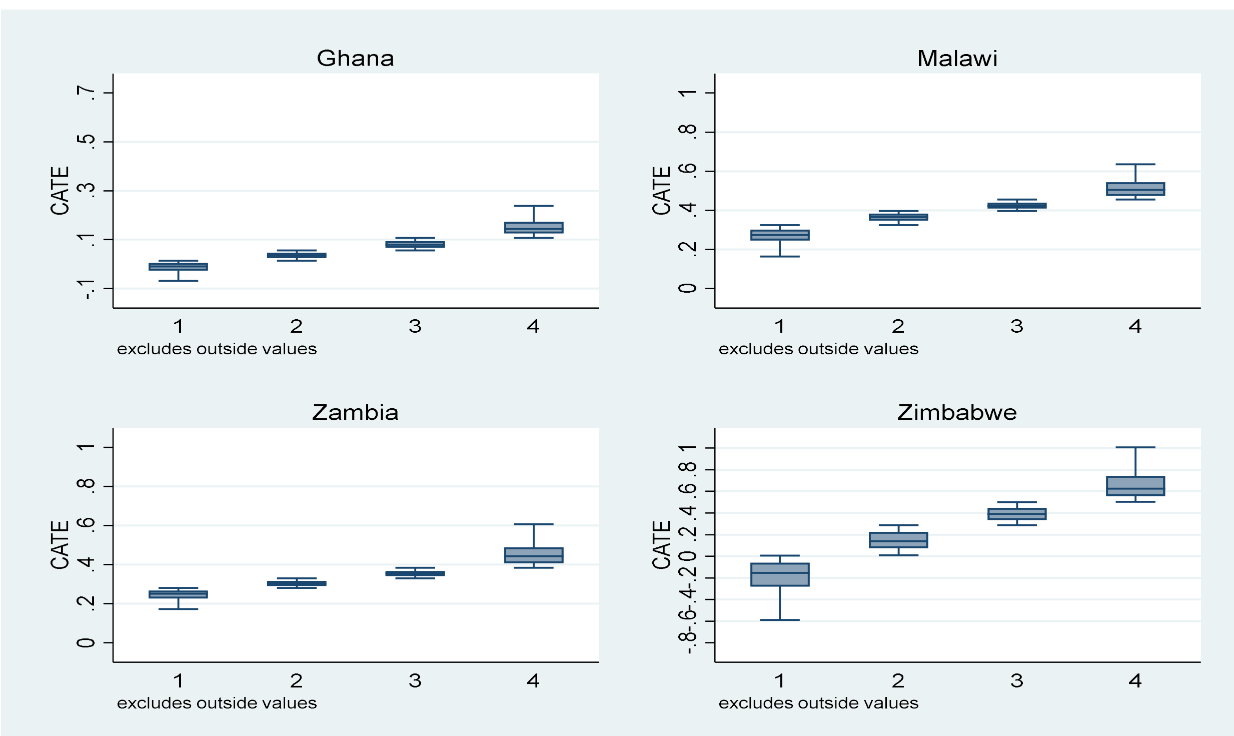
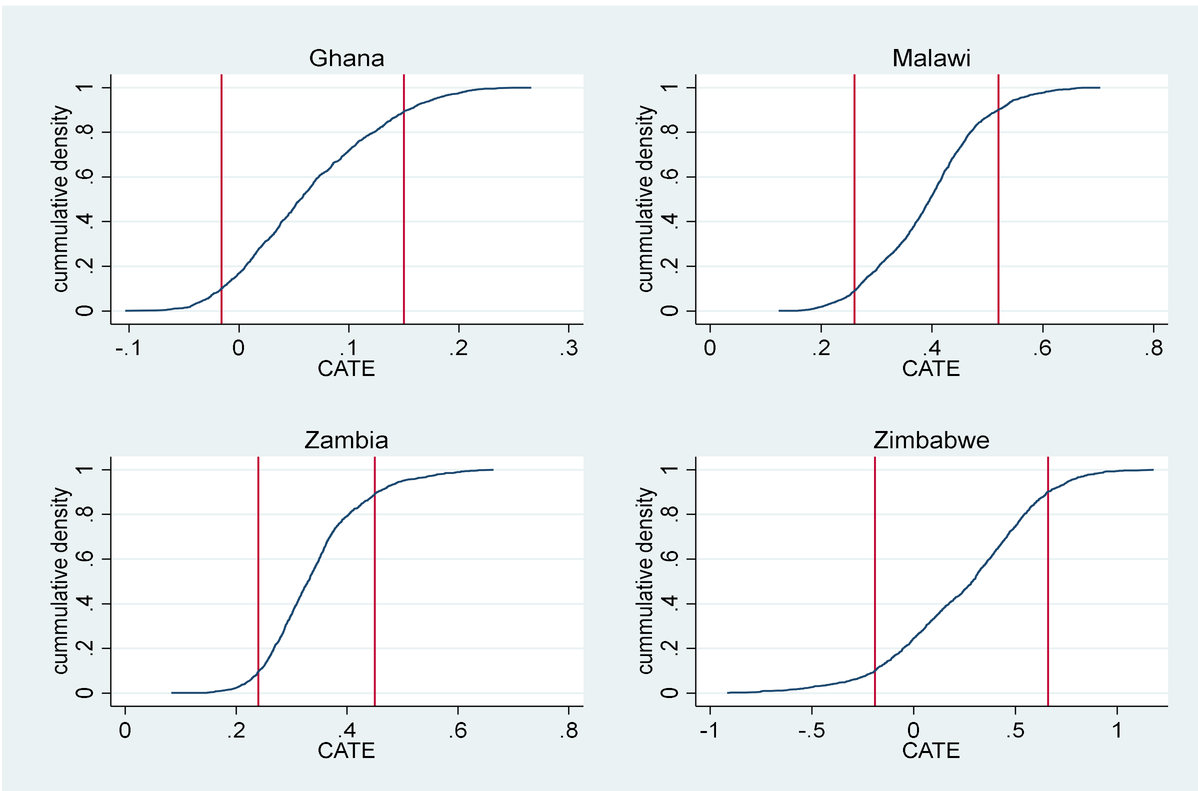


Figure 3: Box plots of CATES by quartile



We further investigate treatment heterogeneity by plotting the cumulative density function of the estimated CATES in Figure 4 with the vertical lines at the 10th and 90th percentile. The density for Ghana is flat and 80 percent of the predicted CATES are within a range of just 0.15SD, which is consistent with the box plots and suggests that we have not uncovered any meaningful treatment heterogeneity in Ghana. The density for Zimbabwe is also somewhat flat like in Ghana, but in this case the range of treatment effects is large, with many negative effects. The densitites are similar in Zambia and Malawi, a very steep cumulative density in the middle 80 percent indicating compression in the middle, but the difference at the top and bottom is quite large, suggesting some degree of treatment heterogeneity at the extremes of the distribution.

Figure 4: Cumulative density of predicted CATES by country



We examine more formally differences between high- and low-flyers in their baseline characteristics to see if these can help us understand why some were sensitive to the treatment and others not. In Table 3 we provide means of the feature list (measured at baseline, treatment group only) for households in the bottom quintile of the CATES distribution (low-flyers) versus the top quintile.[8] We can group the features into three broad categories: individual household characteristics, productive and psychological assets, and the environment. As there are a lot of comparisons to digest, we have summarized the statistically

[8] Appendix B shows variable importance graphs from the causal forests for each country—the percent of times a variable was used to split the tree. There are no clear features that stand-out consistently across the four countries.

significant differences in table 4, where red indicates the value of the feature is negatively associated with being a high flyer, green that it is positively associated, and no color means no significant difference. In Ghana LEAP 1000 and Zambia CGP all recipients are female, while in Ghana we do not have measures of productive assets at baseline—these are indicated with a 'not available' in the table. Though we have presented the results for Ghana, we will not comment on those because the prior analysis suggests that there is not significant heterogeneity in treatment effects.

Table 3a: Ghana Mean differences in baseline characteristics by low/high CATE

	Ghana (N = 596)		
	Low	High	P-value of diff
Per capita expenditure z-score	-0.15	-0.06	0.45
Head no formal schooling	0.87	0.85	0.69
Age of head	47.42	34.38	0.00
Household size	7.90	6.62	0.00
Children age 0-12	2.19	1.71	0.01
dependency ratio	0.74	0.59	0.01
Income and revenue index	-0.01	-0.08	0.58
finance and debt index	0.20	-0.35	0.00
Livestock index	0.19	-0.24	0.00
Psychological states index	0.19	-0.02	0.12
Mean rainfall (mm)	95.17	94.03	0.01
Distance (km) to district capital	7.29	16.64	0.00
Pct of area in Agriculture in 2015	0.59	0.63	0.06

Note: P-values are from tests of equality of means for each variable

Table 3b: Mean differences in baseline characteristics by low/high CATE – Malawi

	Malawi (N =)		
	Low	High	P-value of diff
pcexp_z	-0.02	0.48	0.00
Main respondent female	0.81	0.78	0.41
Main respondent ever attended school	0.33	0.26	0.18
Main respondent age	51.29	73.13	0.00
Main respondent widow	0.27	0.68	0.00
Household size	5.32	2.98	0.00
Children age 0-12 years	2.14	0.92	0.00
dependency ratio	2.78	2.21	0.00
productive assets index	0.31	-0.10	0.00
Income and revenue index	0.98	-0.43	0.00
finance and debt index	0.03	0.33	0.00
Livestock index	0.17	-0.15	0.00
Psychological states index	0.33	-0.36	0.00
rainfall (mm)	83.64	85.02	0.00
Distance (km)to the district capital	18.01	15.49	0.02
Pct of area in Agriculture in 2015	0.44	0.31	0.00

Note: P-values are from tests of equality of means for each variable

Table 3c: Zambia Mean differences in baseline characteristics by low/high CATE

	Zambia (N = 0)		
	Low	High	P-value of diff
Pc consumption expenditure	-0.24	0.21	0.00
Recipient has ever attended school?	0.57	0.86	0.00
Age of recipient	32.52	29.84	0.04
Household size	5.94	5.54	0.11
Children 0-12 years	3.27	2.98	0.09
dependency ratio	2.17	1.93	0.09
productive assets index	-0.02	0.00	0.76
Income and revenue index	-0.17	0.03	0.02
finance and debt index	-0.21	0.00	0.05
Livestock index	0.26	0.13	0.00
Psychological states index	0.09	0.05	0.75
rainfall (mm)	75.68	88.52	0.00
Distance (km) to the district capital	48.79	17.71	0.00
Pct of area in Agriculture in 2015	0.05	0.11	0.00

Note: P-values are from tests of equality of means for each variable

Table 3d: Zimbabwe Mean differences in baseline characteristics by low/high CATE

	Zimbabwe (N = 853)		
	Low	High	P-value of diff
Pc expenditure z-score	-0.02	-0.01	0.92
Ever attended school	0.70	0.28	0.00
Age of respondent (years)	57.54	66.39	0.00
Main respondent female	0.54	0.7	0.00
hhsize	6.18	3.06	0.00
Children age 0-12 years	2.50	1.07	0.00
dependency ratio	2.62	1.91	0.00
productive assets index	1.01	-0.90	0.00
Income and revenue index	0.94	-0.54	0.00
finance and debt index	0.09	0.17	0.46
Livestock index	1.00	-0.65	0.00
Psychological states index	0.45	0.09	0.00
rainfall (mm)	47.89	49.52	0.02
Distance (km)to the district capital	85.36	93.22	0.00
Pct of area in Agriculture	0.33	0.30	0.11

Note: P-values are from tests of equality of means for each variable

There is a clear pattern where high flyers tend to have lower dependency ratios more able-bodied members), fewer children and smaller households. Schooling is not significant in Malawi, positive in Zambia and negative in Zimbabwe—these ambiguous results are not surprising, there is no variation in the sample, the majority of recipients have had no education or have not completed primary. Initial consumption is significant in Malawi and

Zambia, suggesting a degree of state dependency, but looking down the table, other features that might also appear to be state dependent are not, the most surprising being livestock and productive assets (tools such as panga, hoe, pick and shovel); the income and revenue index is negatively associated with high flyers in Malawi and Zimbabwe but positively associated in Zambia.

The finance and debt index is interesting, it tends to be positively associated with high flyer households in all three countries, though not statistically significant in Zimbabwe. This is consistent with the theory of poverty traps that suggests many households require a big push to get them out of a low-level equilibrium; households in debt would require more than a small, periodic cash transfer to enable them to break out of the trap.

Among the agro-ecological variables only rainfall is consistently (and positively) associated with high flyer households; living closer to the district capital is helpful in Malawi and Zambia but not Zimbabwe, while the percent of local area (measured using a 2.5 kilometer buffer) is not consistently associated with high flyer households.

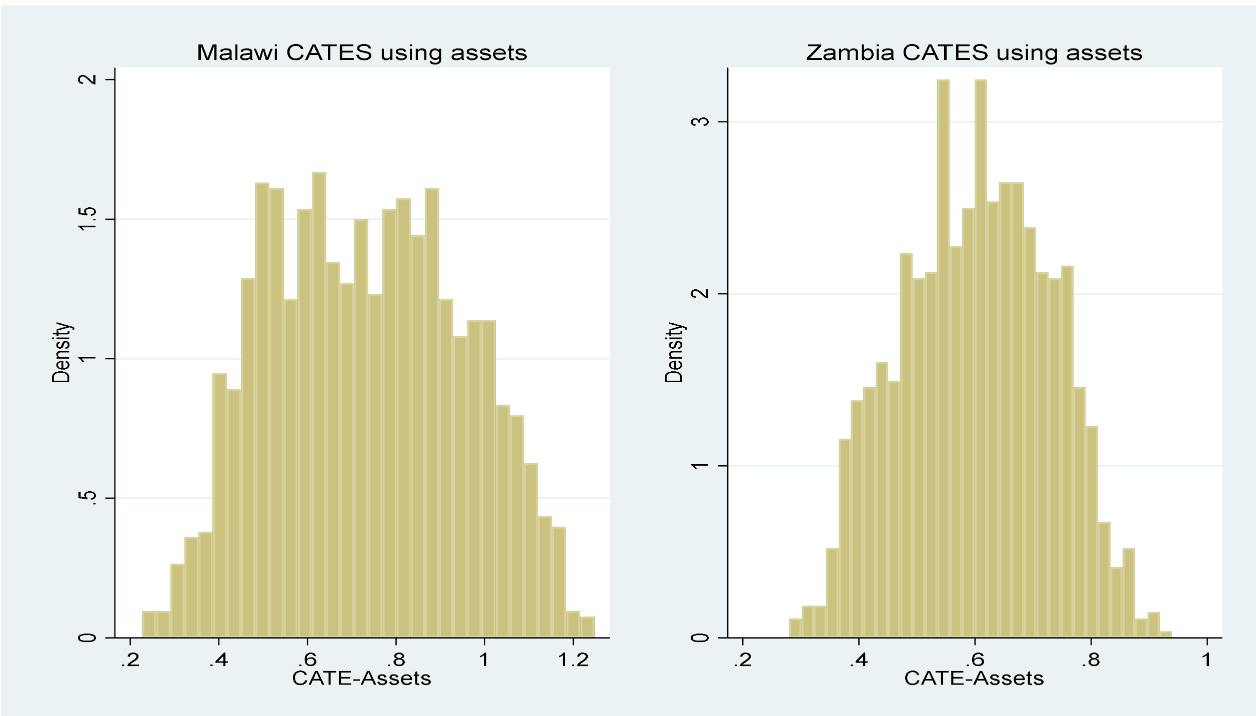
Table 4: Summary of significant differences between low- and high-flyers at baseline

	Ghana	Malawi	Zambia	Zimbabwe
Pc expenditure z-score				
Main respondent female	N/A		N/A	
Main respondent ever attended school				
Main respondent age				
Children age 0-12 years				
Dependency ratio				
hhsize				
productive assets index	N/A			
Income and revenue index				
finance and debt index				
Livestock index				
Psychological states index				
rainfall				
Distance (km)to the district capital				
Pct of area in Agriculture in 2015				

Red (green) indicates that the feature is negatively (positively) and statistically significantly associated with being a high flyer

Using assets as an alternative outcome: The analysis so far has used per capita consumption as the outcome under the common assumption that it is the best representation of well-being in low-income countries. Assets are an alternative measure of well-being in that they are more likely to indicate structural or long-term standards of living. Among our four countries, the most comprehensive measures of assets are contained in the Malawi and Zambia data sets, and comprise both productive and domestic assets. For those two countries we compute CATES from causal forests using assets at follow-up (in z-scores) as the outcome. Figure 5 shows the distribution of CATES for the two countries using assets as the target. The average treatment effects are higher than for consumption, about 0.50SD in each country, and all estimated treatment effects are positive.

Figure 5: Predicted CATES using assets



There is a rather wide range of CATES, and the box plots in Figure 6 show a very steep increase in mean CATES by quartile, suggesting that we are able to identify true heterogeneity in the data using assets as the target. However, the same households are not identified as high- and low-flyers, as shown in Figure 7, where there is a rather weak correlation between the predicted CATES for each household using the two alternative outcomes—the Spearman rank correlation is just around 0.20 in each of the countries for the two measures.

Figure 6: Box plot of CATES by quartile--assets

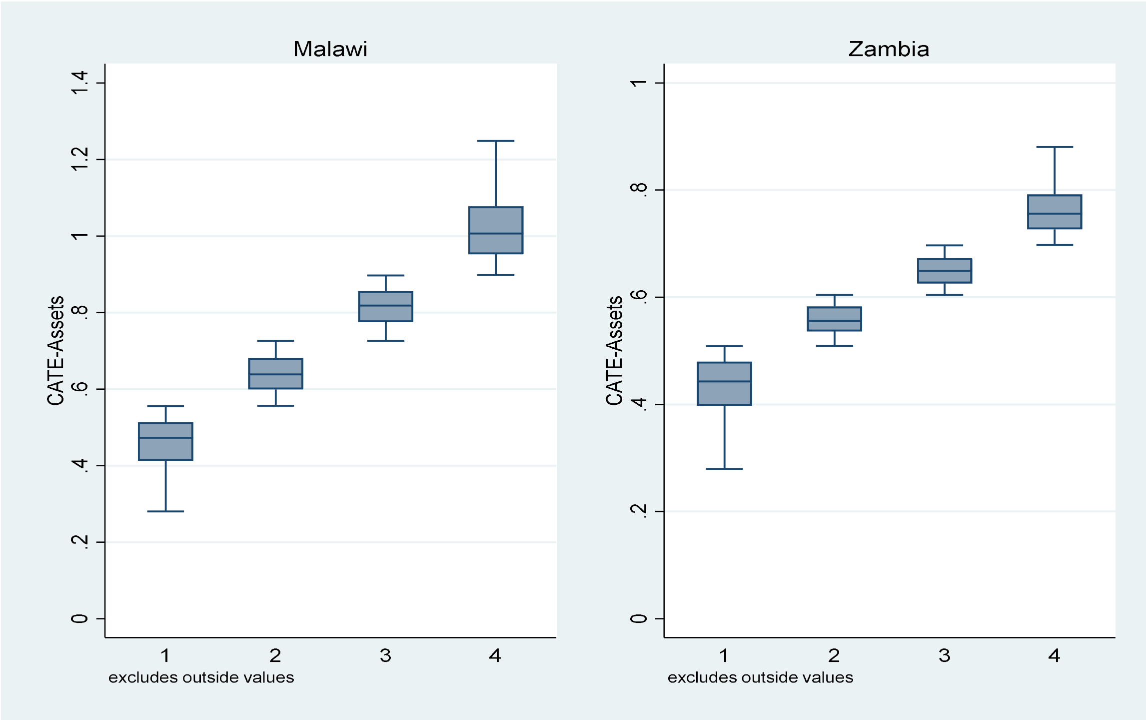
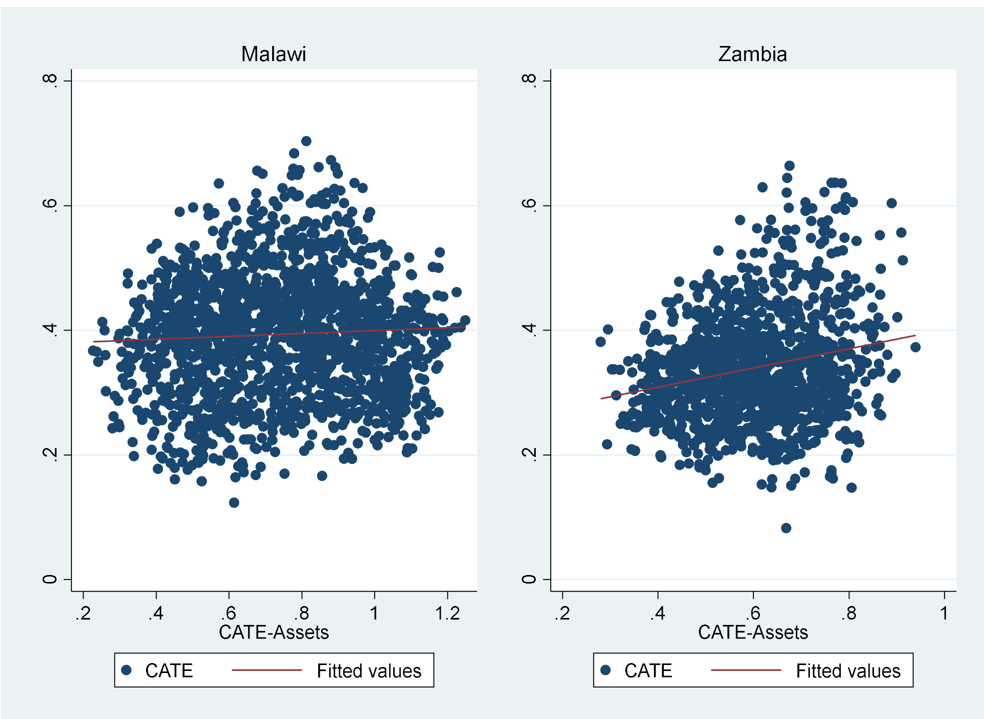


Figure 7: Comparison of CATES estimated with consumption versus assets



Following the previous analysis we compare the means of baseline characteristics between low- and high-flyers using the bottom and top quintile of the predicted CATE with assets as the outcome—results are shown in Table 5 and the summary provided in Table 6. There are important differences in the results from those using consumption—this makes sense given that the correlation in CATES within households is low. For example, high-flyer now tend to have larger households with more children and higher dependency ratios (though the latter is not statistically significant). The head of household is younger in Malawi and older in Zambia, which, given the profile of beneficiaries in those countries, puts them at roughly the same age, and heads have more schooling among high-flyers. The asset profile is also quite different, income and revenue are positive, livestock only negative for Malawi, and finance and debt non-significant. The one set of features that are consistent are those of the agro-ecological context, rainfall continues to be positively associated with high-flyers, and high-flyers live closer to the district capital in Malawi, with an associated lower share of land devoted to agriculture.

The contrast in the characteristics of high-flyers using assets versus consumption suggests that the type of households at baseline that opted to invest in assets is quite different from the type that realized large consumption gains. In both countries, the impact evaluation results showed significant impacts on both asset and consumption, so while the two can be viewed as substitutes to some extent, investment in productive assets at least can generate higher consumption. The fact that the two sets of high-flyers are quite different suggests that those who used the cash transfer to accumulate assets did not manage to realize large consumption gains through that process, suggesting either that the returns to those investments have not yet been realized or that they are simply not productive enough to be considered genuine candidates for graduation. Their distinct profile is also interesting, larger households with more children and higher dependency—here economies of scale in consumption may explain their relatively lower consumption, and the age profile of the head indicates these families are at an earlier stage of the life-cycle as well.

Table 5a: Malawi mean differences in baseline characteristics by low/high CATE - assets

	Malawi (N =)		
	Low	High	P-value of diff
pcexp_z	0.32	-0.35	0.00
Main respondent female	0.82	0.87	0.22
Main respondent ever attended school	0.23	0.53	0.00
Main respondent age	62.25	43.98	0.00
Main respondent widow	0.51	0.34	0.00
Household size	4.03	5.49	0.00
Children age 0-12 years	1.49	2.44	0.00
dependency ratio	2.53	2.85	0.10
productive assets index	0.31	-0.27	0.00
Income and revenue index	-0.13	0.39	0.00
finance and debt index	0.16	-0.05	0.06
Livestock index	0.20	-0.29	0.00
Psychological states index	0.16	-0.06	0.04
rainfall (mm)	83.66	86.53	0.00
Distance (km)to the district capital	18.56	9.43	0.00
Pct of area in Agriculture in 2015	0.60	0.27	0.00

Note: P-values are from tests of equality of means for each variable

Figure 5b: Zambia Mean differences in baseline characteristics by low/high CATE - assets

	Zambia (N = 0)		
	Low	High	P-value of diff
Pc consumption expenditure	0.12	0.02	0.45
Recipient has ever attended school?	0.65	0.84	0.00
Age of recipient	28.14	32.30	0.00
Household size	5.21	6.08	0.00
Children 0-12 years	2.87	3.41	0.00
dependency ratio	1.96	2.16	0.19
productive assets index	-0.02	0.02	0.62
Income and revenue index	-0.15	0.22	0.00
finance and debt index	-0.09	-0.02	0.54
Livestock index	0.18	0.19	0.83
Psychological states index	0.19	-0.00	0.11
rainfall (mm)	78.18	83.25	0.00
Distance (km) to the district capital	28.62	45.87	0.00
Pct of area in Agriculture in 2015	0.07	0.05	0.22

Note: P-values are from tests of equality of means for each variable

Table 6: Summary of significant differences between low- and high-flyers – CATE predicted with assets		
	Malawi	Zambia
Pc expenditure z-score		
Main respondent female		N/A
Main respondent ever attended school		
Main respondent age		
Children age 0-12 years		
Dependency ratio		
hhsize		
productive assets index		
Income and revenue index		
finance and debt index		
Livestock index		
Psychological states index		
rainfall		
Distance (km)to the district capital		
Pct of area in Agriculture in 2015		
Red (green) indicates that the feature is negatively (positively) and statistically significantly associated with being a high flyer		

5. What did high flyers do with the money?

Ultimately it is the combination of starting features and post-intervention actions that led some households to be more sensitive to the cash transfer and realize large consumption gains. Is there a systematic relationship between certain actions or combinations of actions and treatment effects? Figures 8-10 show local linear regressions (LOWESS) between selected features representing productive or investment type activity, and the CATE for each country. These bivariate relationships do not reveal any systematic, consistent relationship across the three countries (we exclude Ghana from this and subsequent analysis since we concluded that there was limited treatment heterogeneity in that country). However, in Zambia there seems to be a relationship between involvement in non-farm enterprise (NFE) and crop sales and higher predicted CATES; there is also a small indication of more debt reduction and predicted CATES across the countries. Even livestock, which in this context is the most obvious and easiest form of investment, does not seem to be consistently associated with higher CATES, even though there were significant impacts on livestock holdings in all three countries.

Figure 8: Malawi relationship between post-treatment behaviors and CATES

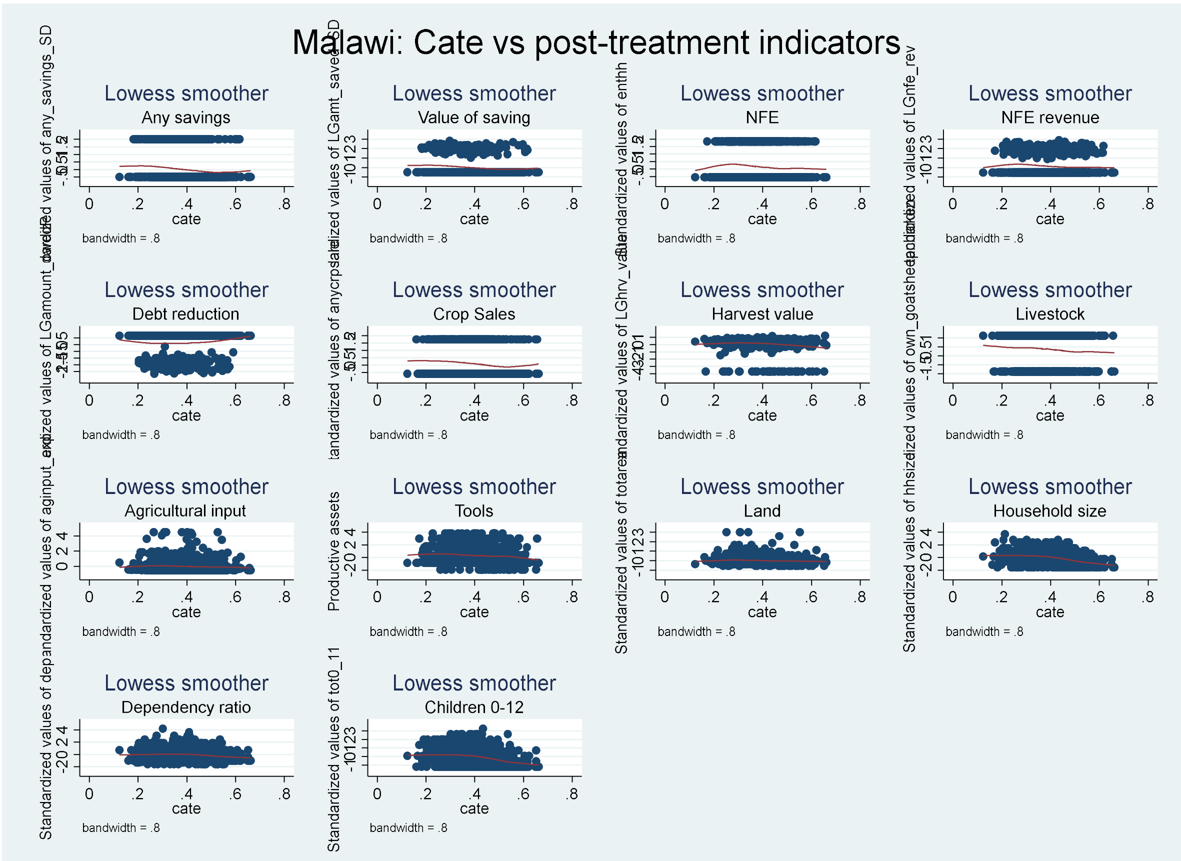


Figure 9: Zambia relationship between post-treatment behaviors and CATES

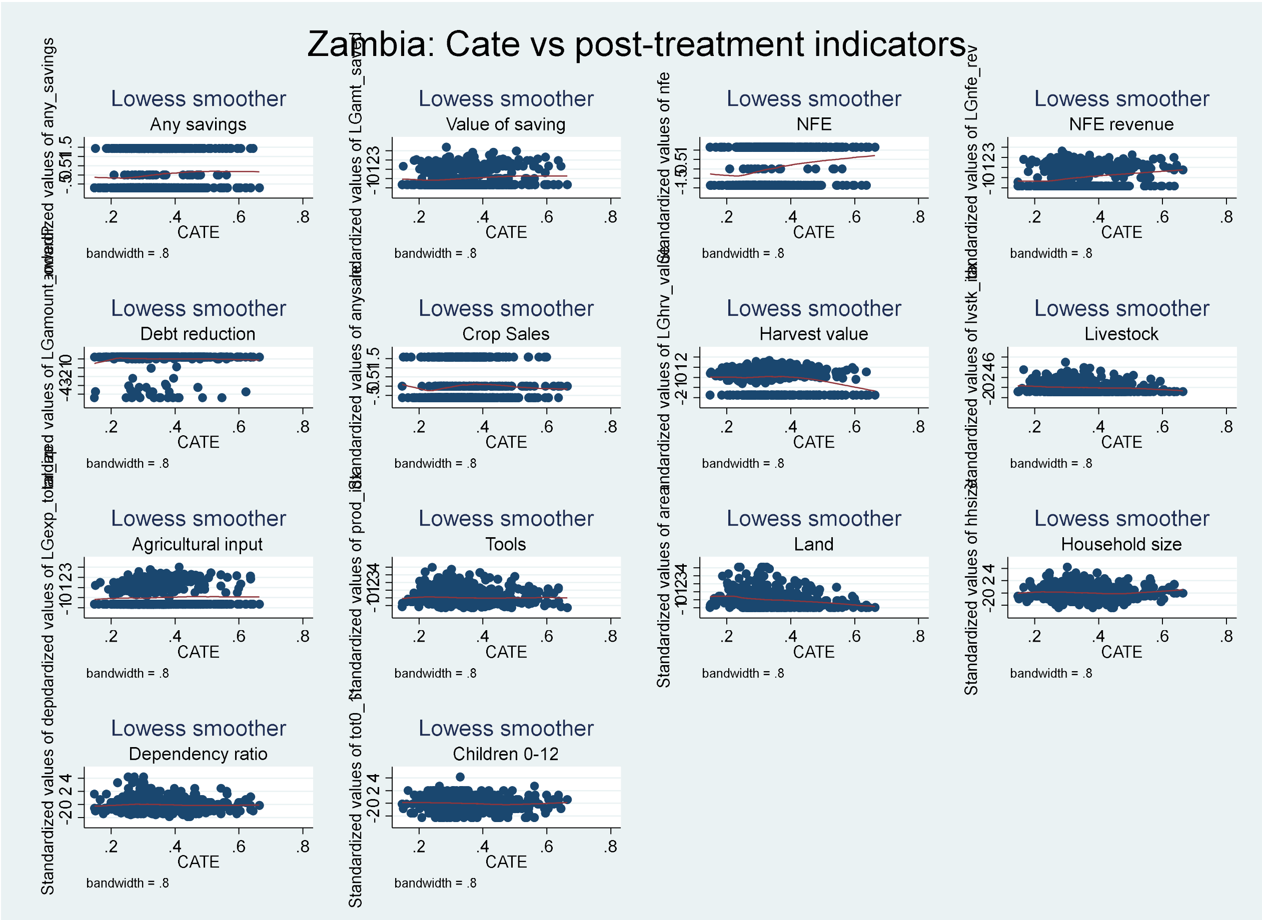
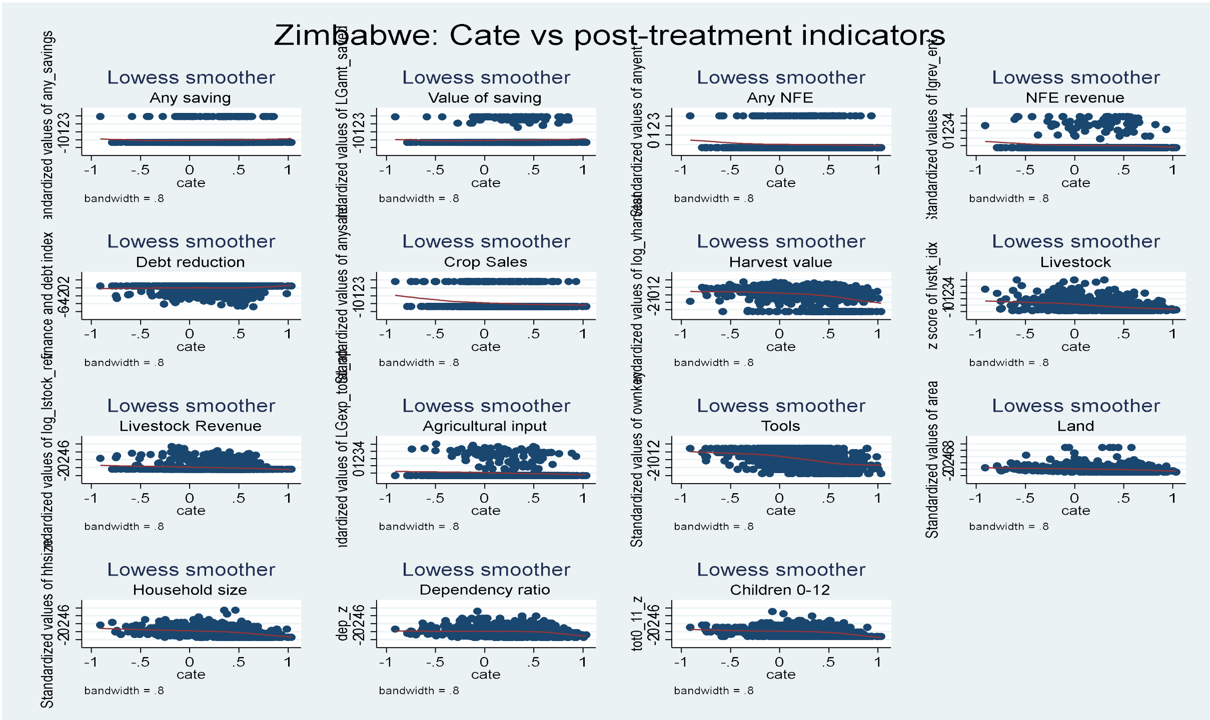


Figure 10: Zimbabwe relationship between post-treatment behaviors and CATES



In tables 7-9 we focus on households in the bottom and top quintile of the CATE distribution and compare means across a set of post-intervention indicators of investment and productive activity. Only in Zambia do we observe a clear pattern whereby high CATE households appear to be more engaged in NFE and crop sales and have higher savings. In the other two countries on the other hand there are no systematic increases in investment or productive behavior to suggest longer term graduation potential. This pattern of results may well be linked to the type of beneficiaries targeted in these three countries. In the Zambia CGP, households are younger with two able-bodied adult heads of household, while in Malawi and Zimbabwe, beneficiaries are labor-constrained, and recipient tend to be elderly, often disabled or chronically ill, with few babe-bodied members. In those programs, high consumption households are likely those choosing present consumption over future consumption as their time horizon and overall productive capacity is low; in Zambia on the other hand, time horizon and productive capacity is high, and indeed high-flyers appear to be those who have increase their engagement in the market through NFE and agricultural commercialization.

Table 7: Mean differences in post-intervention characteristics by low/high CATE - Malawi

	Malawi (N =)		
	Low	High	P-value of diff
Productive assets	0.58	0.22	0.03
Land area	0.03	-0.05	0.08
Livestock	0.48	0.26	0.05
Any crop sales	0.23	-0.09	0.01
Value of harvest	0.22	-0.10	0.01
Engage in NFE	0.27	0.04	0.07
NFE revenue	0.32	0.03	0.02
Debt reduction	0.09	0.30	0.04
Any savings	0.21	-0.17	0.00
Amount saved	0.22	-0.16	0.00
Agricultural inputs	0.05	-0.11	0.08

Note: P-values are from tests of equality of means for each variable

Table 8: Mean differences in post-treatment characteristics by low/high CATE - Zambia

	Zambia (N = 251)		
	Low	High	P-value of diff
Productive assets	0.05	-0.06	0.36
Land area	0.29	-0.29	0.00
Livestock	0.08	-0.13	0.06
Any crop sales	-0.22	0.01	0.03
Value of harvest	0.05	-0.28	0.01
Engage in NFE	-0.32	0.36	0.00
NFE revenue	-0.32	0.33	0.00
Debt reduction	-0.01	-0.00	0.94
Any savings	-0.17	0.16	0.01
Amount saved	-0.18	0.22	0.00
Agricultural inputs	-0.07	0.15	0.09

Note: P-values are from tests of equality of means for each variable

Table 9: Mean differences in post-treatment characteristics by low/high CATE - Zimbabwe

	Zimbabwe (N =)		
	Low	High	P-value of diff
Productive assets	0.67	-0.61	0.00
Land area	0.22	-0.22	0.00
Livestock	0.38	-0.45	0.00
Any crop sales	0.35	-0.16	0.00
Value of harvest	0.23	-0.16	0.00
Engage in NFE	0.33	-0.50	0.00
NFE revenue	0.04	-0.07	0.27
Debt reduction	0.07	-0.08	0.15
Any savings	-0.01	0.11	0.22
Amount saved	-0.07	0.00	0.45
Agricultural inputs	-0.10	-0.03	0.41
Productive assets	0.07	-0.14	0.04

Note: P-values are from tests of equality of means for each variable

CATES predicted with assets: We repeat the above analysis using the CATES predicted with assets rather than consumption, and as before, results do change. In Malawi, households that are high-flyers based on assets now display higher levels of productive activity in the post-treatment period, more livestock, engagement in NFE and productive assets. The reverse occurs in Zambia however, where high-flyers based on this definition no longer show higher levels of productive engagement and investment activity using this definition. There is clearly an important difference in results when using assets as the measure of living standards or graduation potential versus consumption.

Table 10: Mean differences in post-treatment characteristics by low/high CATE assets - Malawi

	Malawi (N =)		
	Low	High	P-value of diff
Productive assets	0.23	0.36	0.37
Land area	-0.04	0.03	0.17
Livestock	0.24	0.54	0.01
Any crop sales	-0.03	0.08	0.31
Value of harvest	0.04	0.27	0.01
Engage in NFE	-0.23	0.57	0.00
NFE revenue	-0.20	0.58	0.00
Debt reduction	0.35	-0.21	0.00
Any savings	-0.17	0.19	0.00
Amount saved	-0.18	0.22	0.00
Agricultural inputs	-0.22	0.15	0.00

Note: P-values are from tests of equality of means for each variable

Table 11: Mean differences in post-treatment characteristics by low/high CATE assets - Zambia

	Zambia (N =)		
	Low	High	P-value of diff
Productive assets	-0.14	0.15	0.01
Land area	-0.13	-0.04	0.36
Livestock	-0.16	0.02	0.10
Any crop sales	-0.30	-0.04	0.01
Value of harvest	-0.17	-0.00	0.17
Engage in NFE	-0.03	0.10	0.30
NFE revenue	-0.02	0.12	0.29
Debt reduction	0.10	-0.03	0.23
Any savings	-0.01	0.04	0.68
Amount saved	-0.06	0.05	0.39
Agricultural inputs	0.05	0.05	0.96

Note: P-values are from tests of equality of means for each variable

Of course, these bivariate analyses do not reveal the potential combination of activities, such as diversification, which might be the key to consumption gains. Because the combinations are quite large, and, so far, the evidence we have looked at has not pointed to any clear combinations or pathways, we resort again to ML to sort the data for us. The ML approach we propose is K-means or hierarchical clustering, which essentially identifies discrete groupings based on a set of covariates.[9] In our application, the covariates are behavioral choices the household made in response to the cash transfer, such as livestock production, non-farm enterprise, crop diversification, overall crop production, use of fertilizer or improved inputs, wage work, or crop yield (essentially the features we used to predict the CATES, but now measured post-treatment). The algorithm groups together cases for which the 'distance' in their values across all the covariates is minimized. Hierarchical clustering approaches allow the data to define the optimal number of clusters, in our application we have used the Calinski/Harabasz Pseudo-f statistic to define the stopping rule, or the optimal number of clusters. Note that this is an 'unsupervised' approach because there is no outcome or predictor *per se*; rather units are being grouped based on their values over a set of covariates, though we do include he predicted CATE as one of the features.

As a first step to assess the potential of k-means clustering to provide answers to our research question, Figure 11 shows the average CATE for each of the clusters (ranging from six to ten) generated by the algorithm in each country. Our hope is that features are clustered in low- and high-flyers. In other words, that high-flyers chose a set of common actions which may be too complicated for the researcher to discern, but which the algorithm is able to identify and to thus cluster together. The figure suggests that in Malawi and Zambia there is no clear set of actions that lead high or low-flyers to cluster together, however in Zimbabwe households in cluster two have a mean CATE that is almost four times that of cluster six, suggesting perhaps a common set of post-treatment activities among low- and high-flyers.

[9] Max Kuhn & Kjell Johnson, 2013, Applied Predictive Modeling, Springer.

Figure 11: Average CATES per clusters generated by k-means algorithm

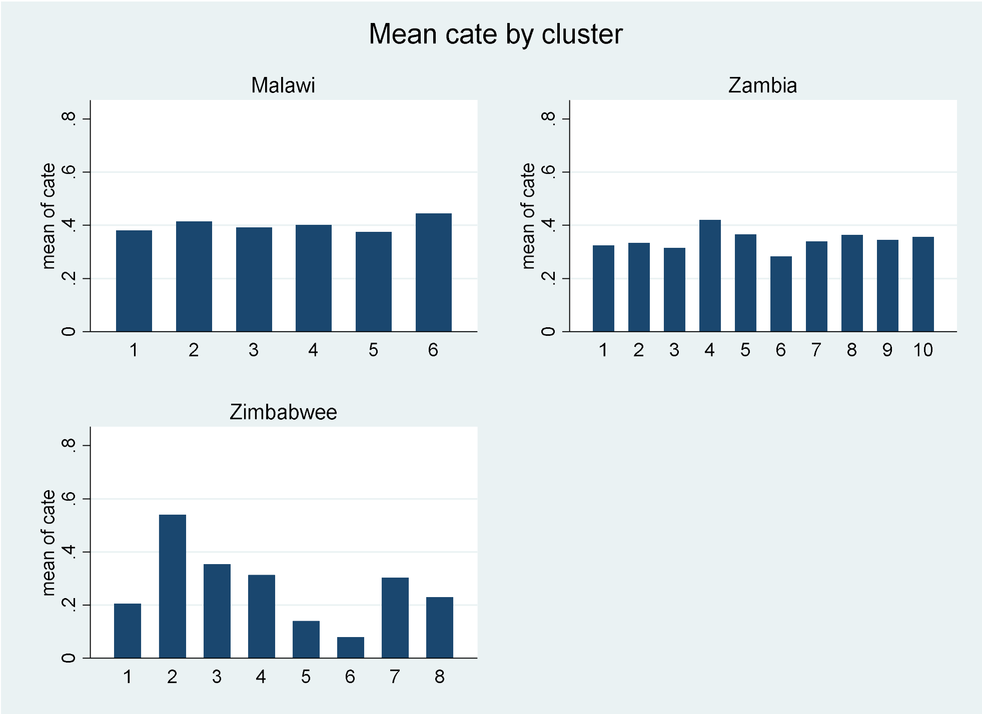


Figure 12-14 shows the means of the post-treatment productive and investment indicators for the low-CATE and high-CATE clusters from the k-means analysis. Recall that only in Zimbabwe did there appear to be any real difference in average CATE across these two clusters. In fact, the results are similar to those from the previous tables where we reported means by top and bottom quintile of the CATE distribution. In Malawi and Zimbabwe, households in high CATE clusters seem to have lower productive and investment activity, while in Zambia we see the same pattern of higher engagement in NFE and more crop sales and savings. One interesting result is that households in high CATE clusters do have positive debt reduction. Overall though, the k-means clustering approach does not seem to provide any further information from what we have already observed.

Figure 12: Mean of post-treatment behaviors by high/low clusters - Malawi

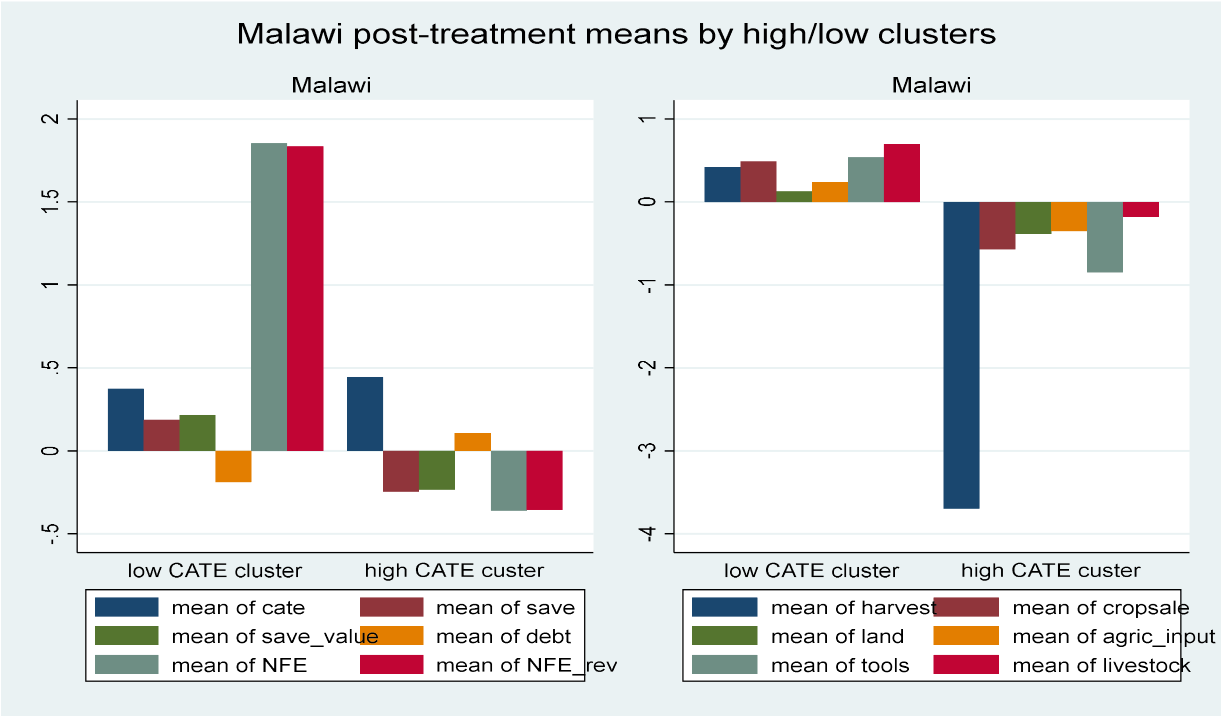


Figure 13: Mean of post-treatment behaviors by high/low clusters - Zambia

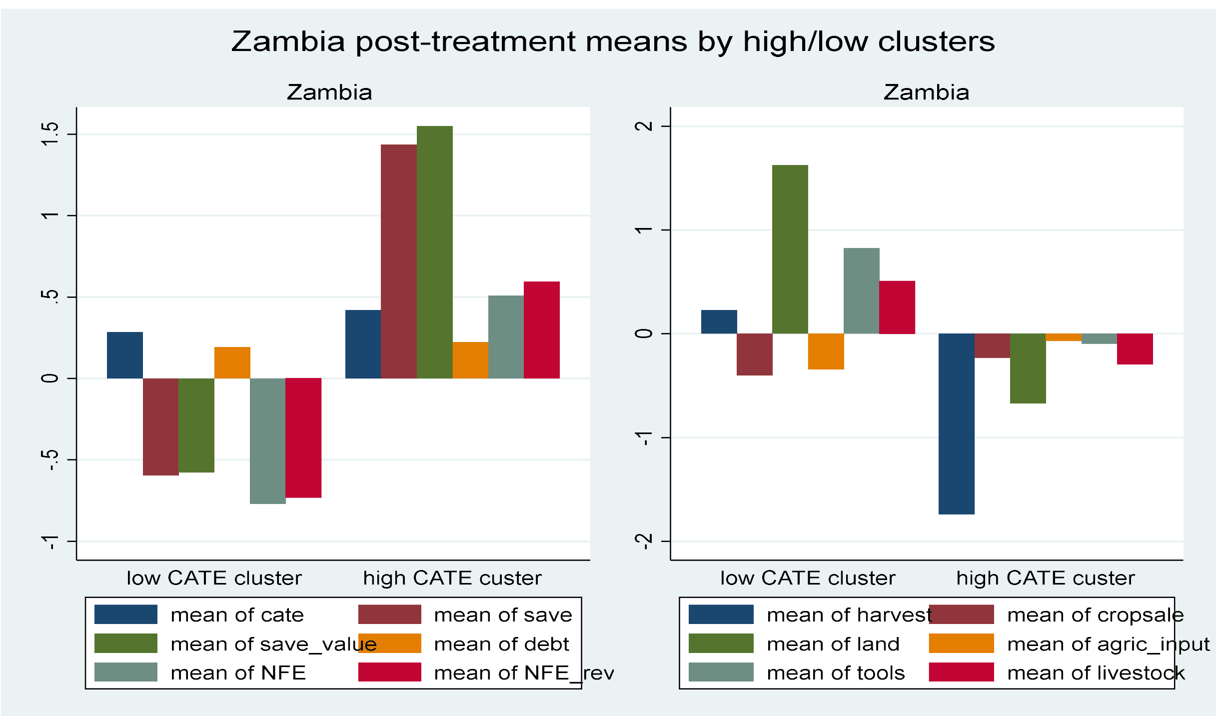
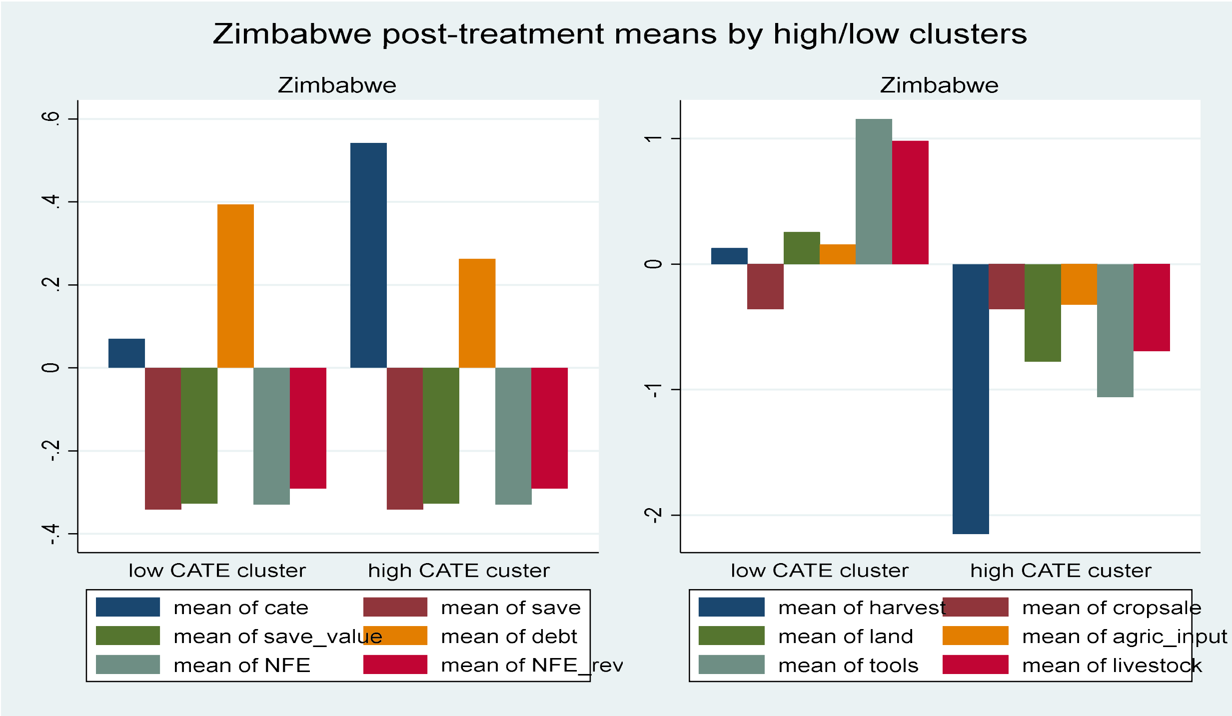


Figure 14: Mean of post-treatment behaviors by high/low clusters - Zimbabwe



6. Using existing theory to predict high flyers

The promise of machine learning is that, with minimal structure, we can allow the data to decide what features of the data are important or not. The standard approach is to use existing theory and prior evidence to build a model that includes the appropriate variables. How would we proceed if we were looking or heterogeneous treatment effects in cash transfer programs, or trying to predict the graduation potential of households? Treatment heterogeneity has been explored in the suite of impact evaluations undertaken by the Transfer Project (see evaluation reports on the website), but those cover a narrow range of topics and are driven primarily by government interest rather than theory per se (e.g. looking at female-headed households, or poorest households at baseline). In countries such as Ghana where the transfer is relatively low, higher impacts are reported among the poorest households, below the median of baseline consumption, an intuitive result since the intensity of the treatment is higher for them.

In terms of graduation itself, Ghatak (2015) provides a useful theory of graduation out of poverty and distinguishes between structural features of the environment, what he refers to as frictions, versus internal constraints to the individual, such as scarcity driven choices that can perpetuate poverty. His theory suggests that UCTs can alleviate the scarcity constraint by raising purchasing power, but if market or external frictions exist, that will not be enough to move people out of poverty through a UCT alone. In other words, policies that address both external frictions and internal constraints are complementary. A recent article by Balboni, Bandiera et al (2022) have tested the theory of poverty traps using an RCT where ultra-poor households received a big push of cash plus complementary asset and training. They find support for the idea of a poverty trap—households that were initially just below a critical point in the wealth distribution, and for whom the intervention put them over the threshold, were more likely to be better off or to have 'graduated' four years later.

Both articles point out that UCTs, because they do not provide a big push, are unlikely to lead to graduation, as consumption support alone, unless it is large enough--at least 100 percent of baseline consumption and possibly up to 300 percent). In national programs, monthly transfers are from 10-25 percent of baseline consumption, hardly near the level required to provide a big push. But the theory developed in these two papers provides information that we might exploit to see if we could predict high-flyers based on these theories. The complementarity of market friction and internal constraints suggests that households with fewer external constraints may be more likely to be high-flyers (living closer to the district capital, more favorable climate). Second, those already close to or above a critical threshold in terms of asset or wealth may be more likely to switch occupations and become high-flyers. Third, those able to overcome certain behavioral scarcities, for example with lower discount rates and more propensity to save and invest, would be candidates for high-flyers. In terms of post-treatment behaviour, the articles talk generally about switching occupations, specifically moving away from low-productivity casual labor to either wage work or NFE, or increasing investment in their own farm in order to raise productivity to possibly sell goods. In this sense, we do observe these exact behaviors among high-flyers in Zambia, and as we indicated earlier, the target group for the program in Zambia is younger and thus a longer time horizon, so in a better position to exploit the UCT to escape poverty permanently.

The above discussion suggests a suite of variables that might predict high flyers or high consumption gains: 1) those closer to the district capital or exposed to higher rainfall; 2) those with lower discount rates or already displaying a propensity to save; 3) those with higher wealth or assets. Of course, we used all these variables in our CATE prediction as well. In terms of post-treatment behaviors, we would expect high-flyers to engage in more diverse livelihood activities, more livestock, crop sales and NFE and wage work—we have already observed this in Zambia, where high-flyers diversify into NFE and commercial cropping.

Table 12-14 show coefficient estimates of regressions where post-treatment consumption is regressed against baseline consumption (the so-called ANCOVA model), an indicator for treatment status, and then in separate models, the treatment indicator is interacted with indicators that represent the three concepts described above. We also control for baseline household size, the dependency ratio and number of children age 0-12. In each country baseline consumption is highly significant and is highest in Zimbabwe, indicating stronger state dependency in that country. Treatment effects are also statistically significant in all three countries in column (1) of each table.

Turning now to the interaction effects, we measure behavioral constraints using time discounting, higher values representing higher myopia—none of the interaction effects are statistically significant. Market friction or contextual constraints are measured by rainfall, the interactions are statistically significant and positive in Malawi and Zambia, indicating complementarity as the theory predicts, but the coefficient is negative and significant in Zimbabwe. Finally, the interaction of treatment and assets is statistically significant in Zambia and Zimbabwe but of opposite sign—this is in fact what the analysis using CATES also showed, that those with high predicted CATES actually have lower baseline assets.

Table 12: Heterogeneous treatment effects - Malawi

baseline consumption	0.184***	0.184***	0.183***	0.185***	0.183***
	-0.0145	-0.0145	-0.0145	-0.0146	-0.0146
Treated	0.353***	0.353***	0.362***	0.353***	0.347***
	-0.0255	-0.0255	-0.0257	-0.0255	-0.0278
Treated * myopic		0.000			
		-0.0184			
Treated*rainfall			0.0494**		
			-0.0194		
Treated*asset index				-0.0102	
				-0.0274	
Treated *(high assets)					0.0264
					-0.0426
N=3,303. *** (**) indicates significance at 1 (5) percent					

Table 13: Heterogeneous treatment effects - Zambia

baseline consumption	0.253***	0.253***	0.230***	0.258***	0.261***
	-0.0196	-0.0196	-0.0193	-0.02	-0.02
Treated	0.311***	0.310***	0.308***	0.310***	0.341***
	-0.0358	-0.0358	-0.035	-0.0358	-0.0386
Treated * myopic		0.0108			
		-0.0261			
Treated*rainfall			0.257***		
			-0.0246		
Treated*asset index				-0.04	
				-0.0367	
Treated *(high assets)					-0.122**
					-0.0594
N=32,519. *** (**) indicates significance at 1 (5) percent					

Table 14: Heterogeneous treatment effects - Zimbabwe

baseline consumption	0.370***	0.370***	0.362***	0.377***	0.376***
	-0.0214	-0.0214	-0.0214	-0.0215	-0.0215
Treated	0.104***	0.104***	0.102***	0.102***	0.134***
	-0.0365	-0.0366	-0.0364	-0.0365	-0.0384
Treated * myopic		-0.014			
		-0.0209			
Treated*rainfall			-0.098***		
			-0.0196		
Treated*asset index				-0.082***	
				-0.0274	
Treated *(high assets)					-0.125**
					-0.051

N=3,567. *** (**) indicates significance at 1 (5) percent

7. Discussion and Conclusions

We have used causal forests, a relatively new machine learning algorithms, to identify heterogeneous treatment effects in four national UCTs in Africa. The advantage of this approach is that it allows the data to identify organically the high-flyers, households who are most sensitive to the intervention. In one country, Ghana, we are not able to identify genuine treatment heterogeneity. In the other three countries, high-flyers have lower dependency ratios, fewer children, and lower debt at baseline, and in two cases, favorable agro-ecological environments. We then compared the post-treatment behaviors of high and low-flyers, and also used an unsupervised MLA—kmeans clustering-to look for systematic behaviors that could explain large consumption gains. In Zimbabwe and Malawi, it was difficult to identify any key behaviors in the productive sphere that would lead to large consumption gains. On the other hand, in Zambia we see a clear pattern of livelihood diversification and a move towards the market, with more engagement in NFEs and crop sales. The pattern of these results can be understood by the type of beneficiaries targeted in these three countries. In the Zambia CGP, households are younger with two able-bodied adult heads of household, while in Malawi and Zimbabwe, beneficiaries are labor-constrained, and recipient tend to be elderly, often disabled or chronically ill, with few able-bodied members. In the latter two programs, high consumption households are likely those choosing present consumption over future consumption as their time horizon and overall productive capacity is low; in Zambia on the other hand, time horizon and productive capacity is high, and indeed high-flyers appear to be those who have increase their engagement in the market through NFE and agricultural commercialization.

In two countries with comprehensive asset information, we use assets rather than consumption to identify high-flyers. We find that high-flyers are different when we assess treatment heterogeneity using assets, indicating that the type of households at baseline that opted to invest in assets is quite different from the type that realized large consumption gains. To some extent investment in assets requires sacrificing current consumption, but it could lead to higher future consumption, or at least less variance n consumption (i.e. better ability to withstand shocks). The fact that the two sets of high-flyers are quite different indicates that those who used the cash transfer to accumulate assets did not manage to realize large consumption gains through that process, either because the returns to those investments have not yet been realized or they are simply not productive enough to be considered genuine candidates for graduation. Their distinct profile is also interesting, larger households with more children and higher dependency—here economies of scale in consumption may explain their relatively lower consumption, and the age profile of the head indicates these families are at an earlier stage of the life-cycle as well.

We can put these two sets of results together along with the profile of beneficiaries to shed some light on the graduation potential of ultra-poor households. First and foremost, ultra-poor households are quite diverse as shown by the stark contrast in the profile of beneficiaries across the four counties, and as a result, their response to the income transfer is different. In Zambia, where households are at an earlier stage of their lifecycle, high-flyers engage in livelihood diversification and a shift towards the market, and do so productively so that they are able to raise their standard of living significantly after four years.[10] In Zimbabwe and Malawi where beneficiaries are elderly households with fewer able-bodied members, there is less scope for graduation as households are further along in the lifecycle and time horizon is shorter. Within this group, high-flyers are those with lower dependency ratios, fewer kids and younger heads of household. Thus, across the three programs, high-flyers are characterized by younger households with fewer dependents. In Zambia we see a clear set of activities that involve livelihood diversification and market integration, but we do not see these behaviors in the other two countries. On the other hand, in all three countries, high-flyers have less debt at baseline, and in Malawi and Zimbabwe there is a clear post-treatment choice of debt reduction, but not in Zambia. This is consistent with a shorter time horizon of beneficiaries in those two countries.

We compare our results with predictions from theories of poverty reduction. The poverty trap theory predicts that households can escape poverty with a big push, in order to overcome technological indivisibilities and other market frictions. Behavioral constraints due to scarcity can also present actions that can lead to long-term graduation. We explicitly tested these theories via standard regression analyses, and did not identify statistically significant treatment effects across the dimensions of myopia or pre-program assets, but did find larger treatment effects among households with better agro-ecological conditions, a proxy for market friction, even after controlling for the dependency ratio. However, the regression framework is somewhat limited in flexibility, and the particular strength of the regression—that it can isolate partial correlations—is not necessarily an advantage when it is the combination of characteristics that matters. In our example, it is the combination of a particular set of baseline household characteristics plus the post-treatment behaviors of those particular households that helps us understand graduation potential.

[10] This aspect of the Zambia UCT was featured on National Public Radio in the United States on August 9, 2017 on Goats & Soda.

The above-mentioned theories of poverty predict that UCTs alone are not enough to enable graduation. Even if a UCT was to eliminate behavioral constraints by alleviating scarcity, technological indivisibilities and market frictions would continue to be barriers to graduation. Our results suggest that UCTs in the hands of a particular group of the ultra-poor, those households at the younger side of the lifecycle, and provided over a long time period (in our Zambia case four years) can lead to significant improvements in consumption via a shift in livelihoods—which is the key to graduation in any theory of poverty. Practically speaking though, it must be remembered that poverty-targeted UCTs are primarily meant to provide consumption support, and often target households who have limited productive capacities. What we show is that among this wider group, there is a specific subset for who a long-term UCT can enable a productivity enhancing shift in livelihoods that may boost long-term living standards.

Appendix A: Estimating causal trees and deriving conditional average treatment effects

We analyze conditional average treatment effects (CATEs) with pre-treatment data. The effect (i.e., outcome) and the cause (i.e., treatment) are usually analyzed via causal effect analysis.To understand the relationship between caus and effect, a common approach is to partition the population in feature space into subpopulations. Each subpopulation has a specific sensitivity to the treatment. Essentially, the common variation between subpopulations is related to common social and economic environment. The individual variation, on the other hand, reveals the different sensitivity that we are interested in. The corresponding decomposition also results in the factors for each variation component. For example, the factors for the individual variation of sensitive households give insights on the key properties of escaping poverty. Also, the factors provide guidance of making policies.

The first stage is based on causal effect estimation. This stage aims to establish key covariates that lead to maximally heterogeneous treatment effects. Meanwhile, we are able to identify the high flyers that show more frequently in the maximal heterogeneity subgroup. Specifically, we start from a population of households, some of which are treated while others are not. Per capital expenditures of these households are taken as the outcome, which indicate the relative living standard of the household. We hypothesize that the households behave differently due to the treatment and such differences lead to the discrepancy between the treated and control group. Following Athey et al. we use $\tau(X) = E(Y_1 - Y_0 \mid X = x)$ to denote the CATE between the treatment and control group, where Y_1 and Y_0, respectively, denote the outcome of the treated and control group and where X denotes the covariates of households. We simplify the notation of CATE as τ in the following text. The distribution of conditional average treatment effects can thus be written as $p(\tau)$. We model the distribution via training a causal forest, which allows us to estimate the conditional treatment effect $\tau(Y_1 - Y_0 \mid X = x)$ for an arbitrary household x. From the estimation, we identify high flyers.

Let T_i denote the indicator of the intervention on the i^{th} subject (e.g., households); $T_i = 1$ means the i^{th} subject is treated (e.g., financially aided) while $T_i = 0$ means this subject is not treated. As the consequence of the intervention, subjects present different outcomes that are measurable. In the example of poverty reduction, an important measurable outcome is expenditure (per household) as the consequence of financially aid from government. We use Y^i to represent the outcome of the i^{th} subject, and the outcome of the i^{th} subject varies due to the intervention. Not every subject is associated with a treatment indicator or an outcome variable. In fact, it is not clear for many subjects whether they should be treated although this is desired for a policy maker. Nonetheless, every subject provides the same number of features or covariates. The subjects with definite treatment indicators and outcomes are taken as training data, while others are taken as test data. Typically, it is difficult to obtain the treatment effect of an arbitrary subject, as discussed in the next section. However, the treatment effect of a subject can be approximated with the average treatment effect of a subpopulation.

Approximate Treatment Effect of A Subject in A Subpopulation: A population can be partitioned into subpopulations, and the response of the intervention varies from one subpopulation to another. An average quantity of a subpopulation is representative to subjects within the subpopulation. The treatment effect, for example, quantifies the difference between a subject (e.g., a household) being treated (e.g., financially aided) and not being treated. Typically, this difference can not directly be measured because a subject can only be either treated or not treated. Therefore, an approximation of the treatment effect of a subject is needed.

Assuming a population consists of treated subjects and controls (i.e., subjects not being treated) with the same number of features, we can partition the population into subpopulations by maximizing the dissimilarity between subpopulations. A partition Π of a population map a set of all subjects into subsets according to features or covariates. The dissimilarity of treatment effects between subpopulations can be defined as between-subpopulation variance of the treatment effects. This dissimilarity is also referred to as *heterogeneity*. Thus, the partition that maximizes the between-subpopulation variance is also known as *heterogeneous treatment effects estimation*.

In addition to maximizing heterogeneity of treatment effects, a good partition also minimizes within- subpopulation variance. In total, the partition is driven by the combination of maximizing between-subpopulation variance and minimizing within-subpopulation variance.[11]

As the feature space is partitioned via Π as above discussed, the estimated average treatment effect of a subject is naturally conditioned on the covariates; namely, it depends on in which subpopulation the subject fall. This Conditional Average Treatment Effect (CATE) can thus be written as

$$\tau(x) = \mathrm{E}(Y_1 - Y_0 \mid X = x) \tag{1}$$

where $\mathrm{E}(\cdot)$ denotes the expectation; Y_1 and Y_0 are, respectively, the outcome of subjects being treated and controls. The symbol X denotes the feature space of all possible covariates, while x is a feature tuple of a given subject.

CATE provides a good estimation of treatment effect for a given subject. The estimation of τ relies on robust assign- ment of the subject to a subpopulation. There exists a huge volume of research on clustering a population such that each subpopulation presents homogeneous properties while different subpopulations present significant heterogeneity. For example, Sano et al.[12] applied k-mean to cluster on the poverty data of provinces all over Indonesia. Athey et al. started a trend of partition populations with the honest causal tree. This trend continues with a variety of studies on recursive partition of a population.[13][14]

[11] Hyun-Suk Lee, Yao Zhang, William Zame, Cong Shen, Jang-Won Lee, and Mihaela van der Schaar. Robust recursive partitioning for heterogeneous treatment effects with uncertainty quantification. *arXiv preprint arXiv:2006.07917*, 2020.

[12] Albert V Dian Sano and Hendro Nindito. Application of k-means algorithm for cluster analysis on poverty of provinces in indonesia. *ComTech: Computer, Mathematics and Engineering Applications*, 7(2):141–150, 2016.

[13] Christopher Tran and Elena Zheleva. Learning triggers for heterogeneous treatment effects. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5183–5190, 2019.

[14] Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021

Honest Estimation of Causal Tree: Honest estimation address the generalizability in fitting causal trees. A causal tree is a generalized decision tree aiming for the estimation of causal effect in response to intervention (i.e., treatment).Unlike a decision tree, a causal tree adopts different splitting rules to maximize the heterogeneity within subpopulation. Athey et al. noted that the estimation could easily overfit the training data and thus lose the generalizability to unseen data. Therefore, they proposed an honest estimation of a causal forest as explained below.

The honest estimation relies on independent partitions of estimation data and tree-growing data. The estimation process can be understood as a two-stage algorithm: the first stage uses the tree-growing data to initialize a tree that is optimal w.r.t. these data, while the second stage uses the estimation data to further refine the tree such that it is more generalizable. Specifically, given a set of data X, a procedure of growing tree is invoked on a subset $X_{tr} \subset X$, resulting in the initial tree T. The complementary data $X_{tr} \subset X$ are propagated through the tree. Due to the random difference between X_{tr} and X_{tr}, this propagation yields a different configuration of leaves. According to this configuration, the tree T is refined in the two senses: (a) the leaves with very few data are regarded as a characterization specific to X_{tr} and not generalizable and (b) the average treatment effect in a leaf is updated according to the new configuration.

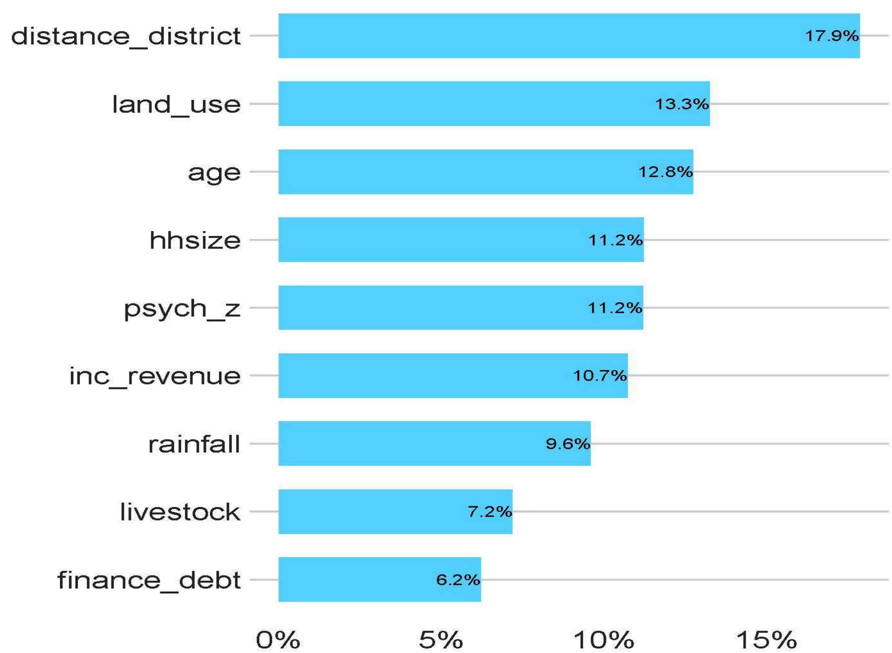
Generalized Random Forest with Honest Causal Trees: GRF (Generalized Random Forest) [17] provides functions of fitting randomly selected data with causal trees in an honest manner.[15][16] The randomness of selecting the data in training, estimation and cross-validation makes the fitted model more generalizable to unseen data. Given a new subject, GRF predicts the CATE for the subject with every causal tree in the forest. The average CATE is taken as the predicted CATE for the subject.

[15] Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.

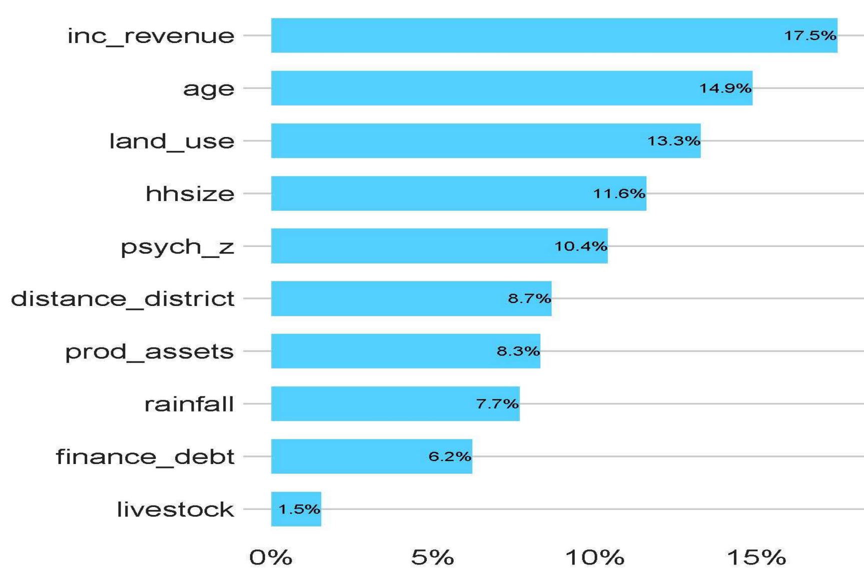
[16] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Appendix B: Variable Importance in the Causal Forest Analysis

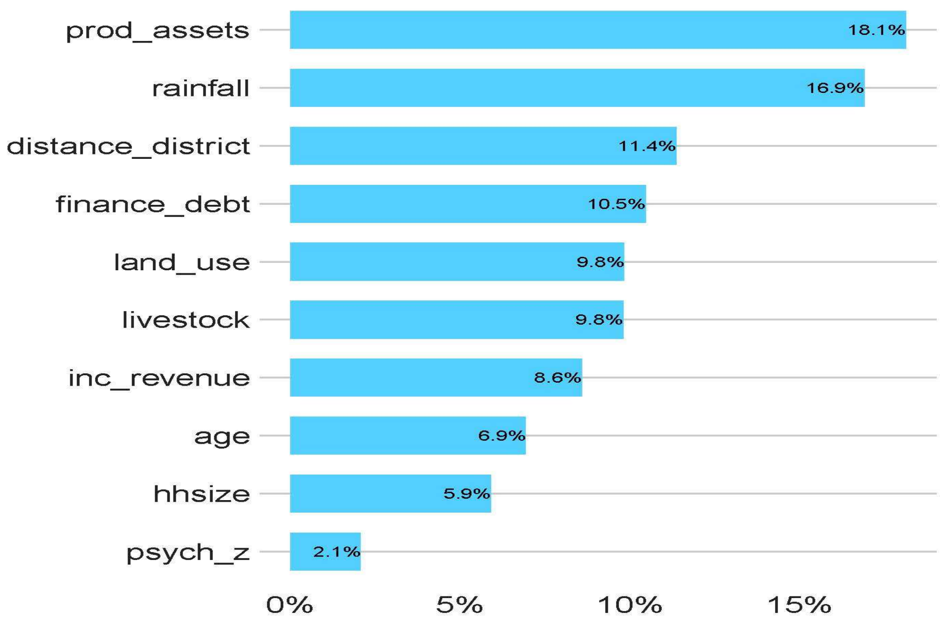
Ghana Variable Importance



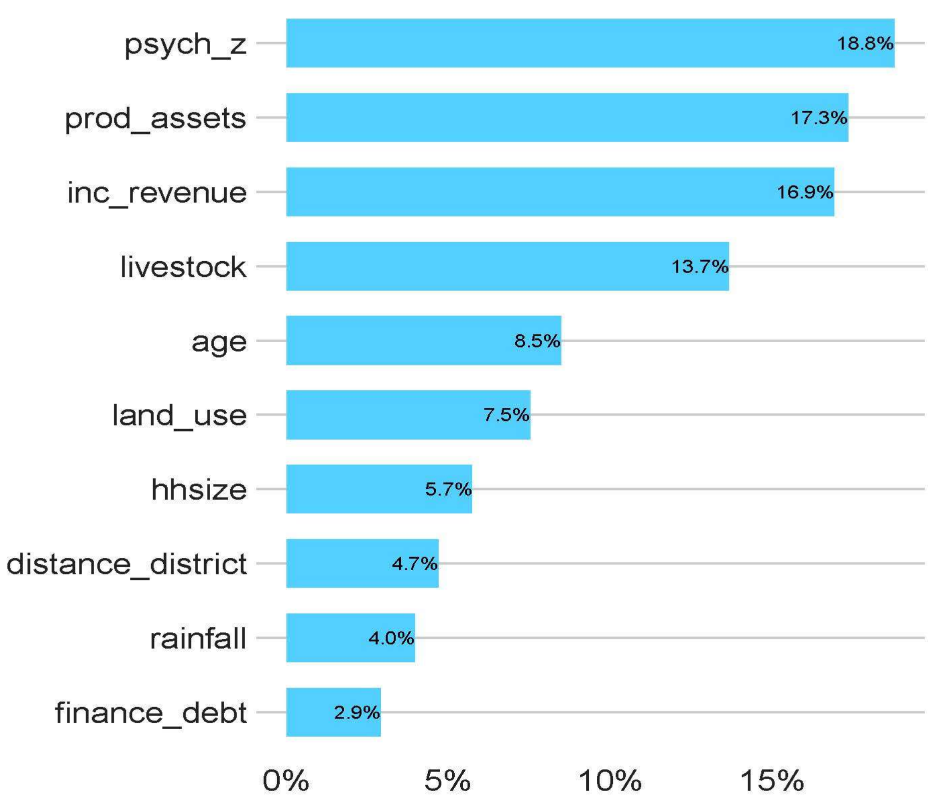
Malawi Variable importance



Zambia Variable Importance



Zimbabwe Variable Importance



Centre of Excellence for Development Impact and Learning

www.CEDILprogramme.org