Environment for Development

Discussion Paper Series June 2020 ■ EfD DP 20-19

Aid, Collective Action and Benefits to Smallholders

Evaluating the World Food Program's Purchase for Progress Pilot

Dambala Gelo, Edwin Muchapondwa, Abebe Shimeles, and Johane Dikgang

Central America
Research Program in Economics and Environment for Development in Central America Tropical Agricultural Research and Higher Education Center (CATIE)

Chile
Research Nucleus on Environmental and Natural Resource Economics (NENRE)
Universidad de Concepción

China
Environmental Economics Program in China (EEPC)
Peking University

Colombia
The Research Group on Environmental, Natural Resource and Applied Economics Studies (REES-CEDE), Universidad de los Andes, Colombia

Ethiopia
Environment and Climate Research Center (ECRC), Policy Studies Institute, Addis Ababa, Ethiopia

Ghana
The Environment and Natural Resource Research Unit, Institute of Statistical, Social and Economic Research, University of Ghana, Accra

India
Centre for Research on the Economics of Climate, Food, Energy, and Environment, (CECFEE), at Indian Statistical Institute, New Delhi, India

Kenya
School of Economics
University of Nairobi

Nigeria
Resource and Environmental Policy Research Centre, University of Nigeria, Nsukka

South Africa
Environmental Economics Policy Research Unit (EPRU)
University of Cape Town

Sweden
Environmental Economics Unit
University of Gothenburg

Tanzania
Environment for Development Tanzania
University of Dar es Salaam

Uganda
EfD-Mak, School of Economics and Department of Agribusiness and Natural Resource Economics, Makerere University, Kampala

USA (Washington, DC)
Resources for the Future (RFF)

Vietnam
University of Economics
Ho Chi Minh City, Vietnam

Contents

Aid, Collective Action and Benefits to Smallholders: Evaluating the World Food Program's Purchase for Progress Pilot

Dambala Gelo, Edwin Muchapondwa, Abebe Shimeles, and Johane Dikgang*

Abstract

Smallholder farmers often face prohibitive transaction costs in agricultural commodity markets in developing countries. Consequently, they are only partly integrated into these markets. Although cooperative institutions such as Farmers' Organizations (FOs) may reduce transaction costs and revitalize agricultural production and commercialization, they rarely have been successful in fully delivering on these promises. Against this backdrop, the World Food Programme (WFP) has recently implemented a multi-year and multi-country pilot to increase smallholder participation in commodity markets. The projects involved investing in physical and human capacities of FOs to aggregate commodities and add value, as well as locally purchasing food aid from FOs. The combination of interventions was expected to increase the relative price of agricultural products, particularly staple crops. In this study, using Ethiopian panel survey data, we causally estimate the income and investment effects of the Ethiopian P4P intervention among smallholders, using an entropy balancing model and a semi-parametric difference-in-difference model. We find that P4P has increased participating smallholders' per capita consumption, investment in child schooling, and asset holding, relative to comparison farmers. However, we also find an increased share of expenditure on food (which is a measure of vulnerability) and evidence of elite capture by FO managers.

Keywords: P4P; welfare effect; distributional bias; semi-parametric DID

JEL Codes: D23; Q02; D02

* Dambala Gelo (corresponding author: dambala.kutela@wits.ac.za), School of Economics and Finance, University of the Witwatersrand, South Africa. Edwin Muchapondwa, Luleå University of Technology, Sweden. Abebe Shimeles, African Development Bank, Cote d'Ivoire, and School of Economics, University of Cape Town, South Africa. Johane Dikgang, School of Economics, University of Johannesburg, South Africa.

1. Introduction

Agricultural commodity markets in most low-income countries, particularly in Africa, often operate in an environment of prohibitive transaction costs arising from inadequate provision of physical (i.e., road, electricity, telecommunication) and institutional infrastructure (i.e., effective legal mechanisms to enforce contracts, standardization and certification services, market information services) (Gabre-Madhin, 2001; Barrett, 2008; Francesconi and Heerink, 2010).

Transaction costs related to searching, screening and enforcement have considerably limited smallholder participation in agricultural commodity markets (Tadesse and Shively, 2013 and Gelo, 2020). Transaction costs are also to blame for poor integration of geographic markets and imperfect competition within each of them (Gabre-Madhin, 2001; Barrett, 2008), which, in turn, appear to constrain smallholder market participation, via price volatility and higher mark-up by merchants with monopsonistic market power, respectively (de Janvry et al., 1991). Consequently, a large share of smallholders are bound to engage in subsistence or semi-subsistence agriculture and are thus unable to benefit from market liberalization reforms (Jayne et al., 2002). Barrett (2008) observes that subsistence production, often characterized by low specialization, rudimentary technology (low productivity), and thus lower income, has caught smallholders in a low-income equilibrium (poverty trap) across eastern and southern Africa. Renkow et al. (2004) and Barrett (2008) argue that public investment in physical and institutional infrastructure is expected to remedy the situation via reducing transaction costs and stimulating smallholder market participation, thereby raising net returns to agricultural production.

One such institutional alternative is organizing agricultural marketing cooperatives, commonly dubbed Farmer Organizations (FOs).[1] By enabling farmers to integrate down the marketing chain, FOs reduce transaction costs in the following major ways. As a form of vertical integration, they reduce information asymmetry and the associated opportunism by taking the position of the buyer (Bijman and Wollni, 2008). Moreover, by their ability to practice contingency pricing through patronage refunds (Valentinov, 2007), introducing quality grades and standards (Bijman and Wollni, 2008), and counterbalancing the market power of traders (Cook, 1995; Staatz, 1987), the FOs allow their members to receive a higher price for their products. Finally, through scale economy, they also reduce initial fixed costs needed to have access to the markets, which include storage costs, transport costs, and bargaining and searching costs (Kirsten and Sartorius, 2002). As such, by reducing transaction costs, and equivalently increasing effective prices, the FOs increase their members' income. Apart from transaction cost effects, they further increase

[1]In fact, recent years have seen a proliferation of optimism within policy and donor circles that FOs could overcome smallholders' marketing constraints (World Bank, 2003).

their members' income by reimbursement of the surplus gained at another level of the value chain. Furthermore, some cooperatives are also involved in processing agricultural and food products in order to add value to members' products and to extract a greater share of returns along the value chain.

In Ethiopia, our study context, there is increasing empirical evidence pointing to the positive effects of cooperatives on a range of smallholder outcomes (see Abebaw and Haile, 2013; Biggeri et al., 2018; Gezahegn et al., 2019; Zhou et al., 2019). [2] For example, Biggeri et al. (2018) found a positive effect of cooperative membership on the share of production sold by farmers to pasta makers through cooperatives in Ethiopia. This is consistent with findings by Bernard and Spielman's (2009) study in Ethiopia, which found that, although non-members did benefit from positive spill-overs generated by cooperatives, their benefits were limited relative to those accruing to members. Biggeri et al. (2018) suggest that the Agricultural Value Chains Project in Oromia (AVCPO) benefitted land-rich and land-poor farmers equally, as compared to non-participant landless poorer farmers. Abebaw and Haile (2013) confirmed that cooperative membership has a strong positive and heterogeneous impact on fertilizer adoption among Ethiopian smallholders.

In fact, FOs provide a special form of integration in that they involve not only vertical coordination, but also horizontal coordination, i.e., agents horizontally coordinate (form a club) to accomplish vertical integration (Sexton, 1986). This suggests that cooperatives face different governance costs and incentive structures compared to vertically integrated investor-owned firms.

For the most part, though, FOs have failed to fully deliver on their promises, in spite of the common expectation that they would revitalize agricultural production and commercialization (Bernard et al., 2012).[3] Although they are faced with growing competition, and hence require substantial investment and commitment to survive the competition, many FOs appear to be severely resource-constrained and have seen a declining commitment of their members. Both theoretical and empirical evidence from the organization economics literature confirms that FOs often suffer from various incentive problems, lack of clearly defined property rights, and

[2] In the context of Ethiopian smallholders, a large part of transaction costs relates to information asymmetry of price and product quality. Compared to traders, smallholders have little or no information about prevailing central market prices for the quality and quantity of products they are supplying to the market, in which case they either face considerable cost in searching for a better price or receive a below-market price for their produce (Tadesse and Shively, 2013).

[3] In particular, the recent decade has witnessed revived interest and optimism within policy and donor circles to promote farmer organizations as a natural avenue for stimulation of agricultural commercialization owing to reduction in the transaction costs of accessing input and output markets and improved bargaining power of smaller farmers vis-à-vis large buyers or sellers (Kherallah and Kristen, 2001).

management inefficiencies, which erode members' incentives to invest in FO resources (Cook, 1995; Cook and Iliopoulos, 2000).[4]

Normally, one would expect cooperative members to invest in their organization to earn residual income in the form of dividends and higher prices for their produce. However, if an open membership policy is adopted and FO services are freely accessible to non-members too, free riding on FO resources is the optimal strategy of each member, a realization that traps the organization in the equilibrium of underinvestment. This problem is real among Ethiopian FOs, as they follow an open membership policy and provide services for non-members in their respective villages (Bernard et al., 2012).[5]

As a solution, donors (NGOs) or government entities often finance the needed investments to leverage FOs' capacities.[6] Recent interventions by the World Food Program (WFP) through the Purchase for Progress (P4P) initiative is a case in point. Since 2009, WFP has implemented five-year P4P pilot initiatives in 20 countries to strengthen the cooperatives so that they can become a reliable provider of the food commodities demanded by WFP. The present research attempts to evaluate the welfare and investment impacts of this multi-faceted P4P intervention.

Our study was motivated by the following major empirical gaps in the related literature. First, extant literature on food aid has long documented the indirect impacts of food aid on smallholder outcomes, largely through a channel of food-grains price distribution (Barrett 2008). For the most part, this literature confirmed that imported food aid reduces smallholders' incentives for production and market participation, as it drives down local prices of food grains through increasing market supply of these commodities (Tadesse and Shively, 2010, Barrett, 2008, Fitzpatrick and Storey, 1989). Moreover, another strand of studies has confirmed the direct impact of food aid on household welfare (Gilligan and Hoddinott, 2007), child nutrition (Gebreheiwot and Castilla, 2018 and Debela et.al, 2015) and technology adoption (Alem and Broussard, 2018).

[4] Some of the major incentive problems in traditional cooperatives are the free riding problem (gains from cooperative action can be accessed by individuals who did not fully invest in developing the gains), the horizon problem (residual claims that do not extend as far as the economic life of the underlying asset), the portfolio problem (the organization's investment portfolio may not reflect the interests of any given member), the control problem (control of the manager by the members) and influence costs (decisions affect the wealth distribution among members) (Cook and Iliopoulos, 2000).

[5] Ethiopian FOs are engaged in a broader portfolio of service provision, including improved seed and fertilizer, credit, agricultural services, price information, consumption services, literacy training, HIV prevention and provision of public infrastructure for both members and non-members (Bernard et al., 2012). While some of these activities complement core commercialization activities of the FOs, others such as consumption services and public goods provisions are unrelated to their primary purpose.

[6] As external (donors') commitment is often unlikely to remain a sustainable source, complementing FOs with establishment of well-defined property rights may sustainably avoid FOs' serious resource constraints.

However, there exists only anecdotal evidence that links food aid programs and smallholder benefits when such programs are designed to target collectively organized smallholders in the form of cooperatives. A few extant studies that evaluated P4P along this line of inquiry, including Krieger (2014a, b, c) and Lenz and Upton (2015; 2016), have produced inconclusive evidence concerning the impact of this program on a range of smallholder outcomes. However, these studies employed a standard difference-in-differences (DID) estimator, which is based on a strong identification assumption of a parallel trend, raising concern about possible bias in their reported estimates of treatment effects.

Second, cooperative organizations are often prone to incentive and governance problems. Cooperative governance structures are such that members, as formal owners of a cooperative, delegate control to management teams and exercise ownership rights through voting and influencing activities (Cook, 1995; Iliopoulos and Hendriksen, 2009). Divergence of interests between members (principals) and management teams (agents) in a cooperative organization constitutes an agency problem in the governance of these organizations. Agency problems exist, to a certain degree, within any organization in which there is separation of ownership and control rights. However, the absence of a market for exchanging equity shares, and the lack of equity-based management incentive mechanisms and formal Boards of Directors, exacerbate this problem in cooperatives (Cook, 1995). Particularly, the control rights of cooperative management teams provide scope for patronage and rent seeking (elite capture) (Banerjee et al., 2001).

The presence of elite capture undermines the success of cooperative organizations in the commercialization of products because it erodes members' incentives and trust in investing in the organization's resources. In fact, many studies have reported the existence of elite capture within FOs. Among others, Bernard and Spielman (2009) found that, within many cooperatives in Ethiopia, decision-making tends to be concentrated in management committees that are less inclusive of the poorest members of the organization.[7] However, this evidence is anecdotal, context dependent and largely qualitative, suggesting an application of alternative rigorous quantitative analysis to test the elite capture hypothesis within FOs.

The present study sets out to respond to these apparent lacunae in the literature. Drawing on panel data analysis, we provide cogent causal evidence of whether smallholder participation in

[7] Cayers and Dercon (2012) found that Ethiopian rural households that are vertically connected to those in power within the kebele (*village*) have a 12 percentage point higher probability of obtaining food aid. Moreover, Bernard and Spielman (2009) found a lower tendency to participate in marketing FOs among poorer Ethiopian smallholders. The authors also uncovered that, when they do participate, they are often excluded from decision-making processes, supporting our prior belief that control rights and the attendant management decisions rest with the elite group.

P4P programs raised households' per capita consumption and spurred investments.[8] We also provide comprehensive empirical evidence of heterogeneity of these effects arising from possible elite capture. We used the Ethiopian P4P dataset for the analysis, which confirmed that P4P raised per capita consumption and food consumption score per annum, supporting the hypothesis that the return to the P4P intervention is positive. We also document strong evidence that the distribution of this effect is heterogeneous and appears to be biased toward the FOs' management team. Our additional analyses show that the P4P program increased the food consumption score and spurred investments in child education and asset holding.

The rest of the article is organized as follows. Section 2 gives background on the institutional context and assignment mechanism (rollout design) of the P4P program in Ethiopia. Section 3 presents the estimation strategy, while Section 4 presents the main empirical results of the article. Section 5 concludes the paper.

2. The Program and its Assignment Mechanism

In the Ethiopian context, there are two types of FOs: in the first instance, smallholder farmers belong to Primary Cooperatives (PCs). PCs in turn belong to the Cooperative Union (CU). In providing smallholders with better markets, WFP used the CU as the entry point in the sense that it targeted the CU for both procurement and capacity building. The CU has a greater capacity for aggregating the commodity, and both the market and capacity stimulus provided at the CU level were expected to be transmitted all the way down to the PCs and their smallholder members (Krieger, 2014).

The P4P intervention was three-pronged. First, it was aimed at addressing resource constraints of FOs via investing in their physical and human capital. It provided FOs with support such as skills training, including organization management, farming techniques, quality control, and post-harvest handling, and it equipped them with storage infrastructure.[9] These interventions were expected to reduce FOs' resource constraints and help aggregate commodities, add value (e.g., achieve WFP's and its partners' quality standards to fetch a better price), and identify and sustainably access markets (Krieger, 2014b; Erin and Upton, 2015).[10] Second, at the upstream level (demand side), P4P provided an additional source of demand, possibly a larger one, via

[8] These items include all food consumption; non-food consumption items were restricted to direct consumables (matches, soap, linen, and clothes), and school and health expenditures, as well as taxes and other contributions.
[9] This is referred to as supply-side (downstream level) intervention.
[10] Aggregating commodities amounts to attaining economies of scale owing to sharing storage, processing and marketing facilities among a large number of smallholders. We thank the editor for helping us elaborate on this concept.

purchasing food for aid locally from these FOs. Under P4P, demand-side intervention involved the purchase of maize and beans for aid (Erin and Upton, 2016).

The WFP procurement of these commodities has taken the form of direct contracts, forward delivery contracts, and "soft" competitive tenders, although this appears to vary across intervention countries (Erin and Upton, 2015).[11] In 2015, WFP in Ethiopia purchased 30 000 tonnes of maize through P4P guaranteed contracts. It is also noteworthy that, although all P4P FOs received the supply-side intervention, the accompanying demand-side interventions were occasional and didn't include all the P4P FOs. Moreover, purchases quantities have been irregular, in some cases insubstantial, over years across these FOs.

It is expected that these demand-side interventions, such as local procurement or contracting with farmers, will increase demand for products that are often of higher value or higher grade than those normally sold (Lentz and Upton, 2016). This demand shock increases the equilibrium price of P4P commodities in respective local markets. Likewise, it increases the price of close substitutes of these commodities – for example, maize and beans of lower grade normally sold in local markets – through general equilibrium effects. Moreover, guaranteed sales under a forward delivery contract are likely to hedge against price fluctuations of commodities.

Thus, the combination of the increase in prices and reduction of the risk of fluctuation of at least some of the prices, in turn, spurs a range of behavioural responses, both intended and unintended, among P4P program smallholders. However, the effect of price increases cannot be limited to P4P FOs and smallholders. As non-program FOs and farmers use the same markets, they are equally exposed to the increased equilibrium price of P4P commodities and the related substitute commodities. The resultant behavioural responses among non-P4P smallholders in terms of consumption, production, investment and the welfare outcomes thereof are commonly referred to as spill-over effects, which render causal inferences from program evaluation considerably more complex. More tellingly, in the presence of spill-over effects, Rubin's Stable Unit Treatment Value Assumption (SUTVA) is violated, leading to biased estimates of the demand-side program's impacts.

An additional concern about evaluating the P4P program relates to its non-random placement and the consequent selection bias. WFP selected FOs for P4P intervention on the basis of location, prior existence of storage capacity, and access to rural financing, and then identified comparison FOs from agro-ecologically similar areas (Krieger, 2014b; Lentz and Upton, 2015). This means that the P4P roll-out did not follow a random assignment mechanism at either FO or

[11]According to WFP (2016), the results on how much of the final product is sold using guaranteed contracts were mixed, dependent on various factors related to market structure, and country specific.

smallholder levels. In other words, the exposure to the program among FOs and their respective members was driven by observed and unobserved heterogeneities. Consequently, the distribution of some or all of these covariates may not be balanced across the P4P and non-P4P farmers, raising concern about possible selection bias in the program's impact estimates.

3. Methodology

3.1 Data and Descriptive Statistics

Our analysis was based on Ethiopian household panel data that spanned three waves collected from P4P FOs members and non-P4P FOs members. The Ethiopian WFP country office commissioned data collection from samples of P4P FOs/CUs and PCs in every year of the five-year pilot and samples of non-P4P FOs/CUs and PCs in year 1 (the baseline), year 4, and year 5.[12] These surveys also generated a panel data set of the PCs, including information on initial and current capacities (managerial and marketing) of FOs/CUs and PCs, their respective size, marketing experience, location, services provided to members, storage capacity, marketing activity, and credit utilization, among others. At the household level, data was collected on household characteristics, including gender of household head, age of household head, education, membership in FO management, production practices, marketing activity, credit utilization, income from crops, livestock and off-farm sources, and consumption.

Our treatment variable is defined as a smallholder's participation in an FO subject to the P4P intervention.[13] We use per capita household consumption expenditure as the outcome variable. In studies like ours, consumption rather than income is widely used, for several reasons (Skoufias and Katatyama, 2011).[14] First, consumption is commonly believed to provide better evidence of the standard of living than income. Second, an income survey may not capture informal, in-kind or seasonal income, and, thus, may be more susceptible to under-reporting. Third, due to consumption smoothing, consumption expenditure exhibits less fluctuation than income in the short run. Fourth, consumption expenditure provides information about the consumption bundle that fits within the household's budget, although credit market access and household savings affect that. Fifth, income is much more difficult to measure, in particular in a rural environment, where

[12] Note that we are using CUs and FOs interchangeably.
[13] The major intervention in the Ethiopian P4P took the form of WFP's direct investment in increasing CUs' storage capacity, suggesting that increased access to storage can be interpreted as a major part of the P4P treatment.
[14] Compared to household size data, age data of household members is largely imprecise – it contains a lot of missing observations, causing the adult equivalent index to suffer from measurement error – which also translates into measurement error of per capita consumption. To avoid a possible attenuation bias, we therefore simply divided total expenditure by household size to compute per capita consumption.

it is common for farm households to consume what they produce.[15] Our additional outcome variables include the expenditure share on food, expenditure on children's education, asset holding index (measured through principal component analysis), and total cost of selling maize (which measures the transaction cost of participating in maize markets).

Our descriptive statistics analysis (see Table 1) suggests that distribution of FO and household-level baseline covariates were not balanced across treatment and non-treatment groups. These covariates include the FO's working capital size, distance from the WFP warehouse, and membership size; source of market price information (whether a smallholder obtains price information from SMS/cell phone or television/radio); and household head's education and age. The lack of balance raises concerns about selection bias, alluded to in the preceding section.

Table 1: Descriptive Statistics of the Data

Variable	P4P		Non-P4P			
	Mean	S.E	Mean	SE	difference	P_value
Outcome variables						
Per capita expenditure (ETB)	2025.456	591.026	1246.917	75.3073	778.539	0.225
Food expenditure share	0.4434	0.0116	0.4281	0.0116	-0.01535	0.351
Total expenditure on child education (ETB)	560.174	36.5961	463.082	25.368	-97.091	0.0337**
Per-capita expenditure on child education (ETB)	278.969	21.865	203.677	11.5063	75.293	**0.001****
PCA index of asset holding	0.1050	0.0659	-0.1195	0.0636	-0.2245	0.015**
Total cost of selling maize (ETB)	5.615	0.7853	8.281	1.8835	2.665	0.1665
Covariates used for PS/weight estimation						
FO's membership size	877.88	86.94	1173	855.45	295.14	0.030**
FO's working capital (ETB)	2059955	80617	798198	256089	1261755	0.15
Capacity of long-term storage (m^3)	9933.9	1988.91	6294.61	983.66	3639.28	0.07**
FO's distance from WFP warehouse	270.806	1.9610	280.288	2.3211	9.482	0.002***
FO collection point distance from the market	29.633	1.785	27.352	2.047	-2.281	0.399
Price info from direct contact with traders	0.8683	0.0112	0.8661	0.0123	-0.0022	0.894
Price information from agric information board	0.1012	0.0111	0.1073	0.0124	0.0061	0.730
Price information from SMS/mobile phone	0.06987	0.00863	0.04329	0.0065	-0.026575	0.0131***
Price information from radio/television	0.09163	0.0097	0.06597	0.0079	-0.02565	0.0405**
FOs as source of price information	0.5581	0.0172	0.518	0.019	-0.0393	0.125
Household head's education in years	2.141	0.0737	-0.191	0.064	- 0.1917	0.053 **
Households head's age	41.1089	0.6569	43.183	0.708	2.074	0.032 **
Household's land size owned	1.2095	0.03258	1.246	0.0327	-0.03724	0.421
Livestock holding	2.381	0.04924	2.3103	0.04674	-0.071134	0.294
Covariates interacted with the treatment						
Household head's gender (Male=1)	0.7778	0.012	0.821	0.012	0.0437	0.012***
FO's management team membership(yes=1)	0.288	0.013	0.137	0.011	-0.151	0.000***
Subsamples						
Smallholders subsample	1,157		970			
FOs subsample	428		207			

*** p<0.01, ** p<0.05, * p<0.1

[15] We would like to thank the first reviewer for suggesting this point.

3.2 Empirical Strategies

Our data set presents two major econometric challenges. First, assignment to the P4P program, both at FO and smallholder levels, is not randomized. Non-random assignment means that there are differences in the distribution of pre-intervention observable and unobservable characteristics across treatment and control groups. The presence of such differences leads to biased estimates of the program's causal effects. More worryingly, if FOs' and smallholders' characteristics are unbalanced across the treatment and control groups and those differences are associated with the dynamics of the outcome variable, then they will affect the difference between the two groups in the outcome's time trends. In that case, the parallel trend assumption does not hold, which leads to biased impact evaluation estimates of the conventional DID analysis.

Second, the program effect can be heterogeneous across individual or group characteristics, such as gender, in which case pre-treatment differences in observed characteristics can lead to non-parallel outcome dynamics. To overcome the problem of potential selection bias arising from non-random placement into the program, we first estimated a conditional difference-in-differences that combines propensity score matching and difference-in-differences (DID).

A limitation of using a conditional DID is that slight misspecification of the propensity score model can result in substantial bias of estimated treatment effects (Imai and Ratkovic, 2013; 2014). In fact, lack of balance in relevant covariates after matching on estimated propensity scores may signal misspecification of the propensity score equation (Caliendo and Kopeinig, 2005). Because the true propensity score is unknown, the problem of model uncertainty is real, in both matching-based conditional DID and standard PSM models. Faced with such model uncertainty, researchers often estimate the propensity score using an iterative search over propensity score models, which often fails to jointly balance out all of the covariates and in some cases even counteracts bias reduction when balance on some covariates decreases as a result of the pre-processing (Diamond and Sekhon 2006).

To address these shortcomings, we extended our analysis to the entropy balancing (EB) method proposed by Hainmmueller (2012). The EB method is a significant improvement over PSM. Compared to matching procedures, it constructs a weight for each control observation such that the sample moments of observed covariates are identical between the treatment and weighted control groups (Imai and Ratkovic, 2013). The EB method has been shown to detect balance problems (in observed covariates) that matching procedures may not detect (Oyenubi, 2018).

Specifically, it addresses PSM limitations by involving a reweighting scheme that directly incorporates covariate balance into the weighting function that is applied to the sample units.[16]

However, the EB model is limited by a restrictive assumption of conditional independence (CIA). Specifically, when applied to pooled panel data, it cannot allow any kind of dependence between selections for treatment and individual-specific time-invariant unobserved heterogeneity. Neither would it account for treatment effect heterogeneity.

In responding to these econometric challenges in order to provide cogent causal evidence, we further extended our analysis to the implementation of a semi-parametric estimator DID based on Abadie (2005). Unlike a conditional DID estimator, this model relaxes the parallel trend assumption (Abadie, 2005).[17]

To be specific, consider the standard two-period simple DID specification in (1) below;

$$Y_{ipt} = \alpha + \gamma d_t + \lambda t + \tau(d_t * t) + \mu_{ip} + \varepsilon_{ipt} \quad (1)$$

where Y_{ipt} is per capita household consumption of smallholder i in FO p at time t, d_t is the dummy for P4P program participation by smallholder i in FO p, t is the time dummy for the post-intervention period, λt is the time-series difference across periods, μ_{ip} is a cross-sectional difference across treated and control groups, and ε_{ipt} is an individual-transitory shock that has mean zero at each period, with the possibility of correlation in time. Our parameter of interest in (1) is τ, which represents a difference-in-difference treatment effect estimate of the P4P program. Identification of τ relies on the untestable assumption of parallel trend, which implies that $E(\varepsilon_{ipt}|d_t * t) = 0$. The lack of an explicit test for this assumption, however, raises the concern about the acceptability of τ as an exact program effect.

To overcome these challenges, we first estimated (1) using conditional DID. Following Bertrand et. al (2002), we clustered standard errors of the latter at FO level to account for a group-level variation. We then expanded our analysis to the implementation of the entropy covariate balancing (EB) model to identify τ within CIA. Finally, to relax the parallel trend assumption of both conventional DID and CIA, we implemented the semi-parametric DID as alluded to in the previous paragraphs.

[16] The challenge of model uncertainty doesn't apply here because the researcher imposes a potentially large set of balance constraints, such that the covariate distributions across treatment arms in the pre-processed data match exactly on all pre-specified moments. Once these constraints of pre-specified desired level of covariate balance are imposed, entropy balancing searches for the set of weights that satisfies these constraints (Hainmmueller,2012).

[17] The semi-parametric DID estimator is constructed in three steps. First, the differences in outcomes over time for the treated and untreated units are computed. Second, the probability of receiving treatment (propensity score) is estimated and used to weight each observation. Finally, the weighted changes of the outcome over time across treated and non-treated groups are compared to identify the ATT.

As argued by Abadie, (2005)[18], this estimator relaxes the parallel trend assumption of the conventional DID model by allowing the distribution of both observed and unobserved factors to differ between the treated and untreated. Essentially, identification of τ in this model obtains, provided that the effects of unobservable factors on the outcome don't vary with time (*ibid*).

Moreover, this estimator enables the use of covariates to describe how the average effect of the treatment varies with changes in observed characteristics. In addition, the estimator accommodates multi-level treatment variables – i.e., different treatment intensities (Abadie, 2005). The fact that this model allows for treatment effects to vary among individuals is of great interest in our study, as this feature helps us unpack the treatment effect for different groups of sample households (gender groups and FO management team). To estimate the propensity scores that we used for PSM identification in conditional DID and for weighting in the EB and semi-parametric DID models, we used a broader set of baseline covariates, including storage size (storage capacity of long-term storage), FO membership size, FO's initial capacity, FO's distance from the WFP warehouse, FO membership size, FO collection point's distance from the nearest market, source of market price information (whether smallholders obtain price information from the FO, through direct contact with traders or the price information board of the Department of Agriculture, etc., SMS/mobile phone or television/radio), household head's age and education, and household's land holding and livestock holding.

4. Results and Discussion

In this section, we present results of P4P's welfare and investment impacts, from different empirical strategies. Before presenting the results from the conditional DID and EB analyses, we briefly discuss the results of tests aimed at assessing the covariates' balances in these models. With regard to the former, we followed propensity score matching (PSM) kernel matching algorithms, which reduced our sample to 520 program farmers matched to 531 control farmers. Estimates of the logit model (Table 2.1) suggest that the following covariates are predictors of the probability of participating in the P4P program: (1) FO-level covariates, including FO's capital size, FO's distance from WFP warehouse, FO's distance from nearest market, FO's membership size; and (2) household-level covariates, including usage of SMS/cell phone, access to price information, and household head's education. These results broadly tally with descriptive statistics reported in Table 1 and suggest that the program assignment mechanism is non-random.

[18] In the interest of brevity, we refer the interested reader to Abadie (2005) for an excellent treatment of the identification strategy of this estimator.

Table 2.1: Logit Model Estimation of probability of P4P Participation

Covariates	Marginal effect
land_owned	-0.164
	(0.126)
distance_wfp	0.000755***
	(0.000165)
orgsize	3.95e-05**
	(1.78e-05)
captlsiz	1.90e-07*
	(1.08e-07)
storage_capacity	1.25e-05*
	(6.51e-06)
hh_age	-0.000497
	(0.00143)
hh_edu	0.0321**
	(0.0129)
live_stock	0.00104
	(0.00469)
direct_info_trader	-0.0107
	(0.0456)
info from FO	0.0749**
	(0.0345)
gender	-0.0919
	(0.0787)
info_board_MOA	-0.0836
	(0.0606)
SMS_info	0.137***
	(0.0382)
FO_distance_mkt	0.00331***
	(0.000748)
tv	-0.0171
	(0.0385)
Observations	1,013
Log pseudolikelihood	-677.218
Pseudo_R^2	0.0510

As is common in PSM analysis, we performed a t-test to assess covariates' balance across the two treatment arms. We find that, in the post-match sample, there are no statistically significant mean differences of the covariates, as shown in Table 2, Columns 4 and 5. However, for most of the covariates, the standardized difference in mean (a key balance test statistic) far exceeds the conventional threshold of |.1|, suggesting a poor covariate balance in the PSM method (see Figure

1). This is not surprising, as it seems very difficult for PSM to jointly balance all covariates when many of these covariates follow skewed distributions (these are depicted in the subsequent analysis of the entropy balancing model).

Figure 1: Standardized Mean Difference of Covariates in PSM Covariate Balance Assessment

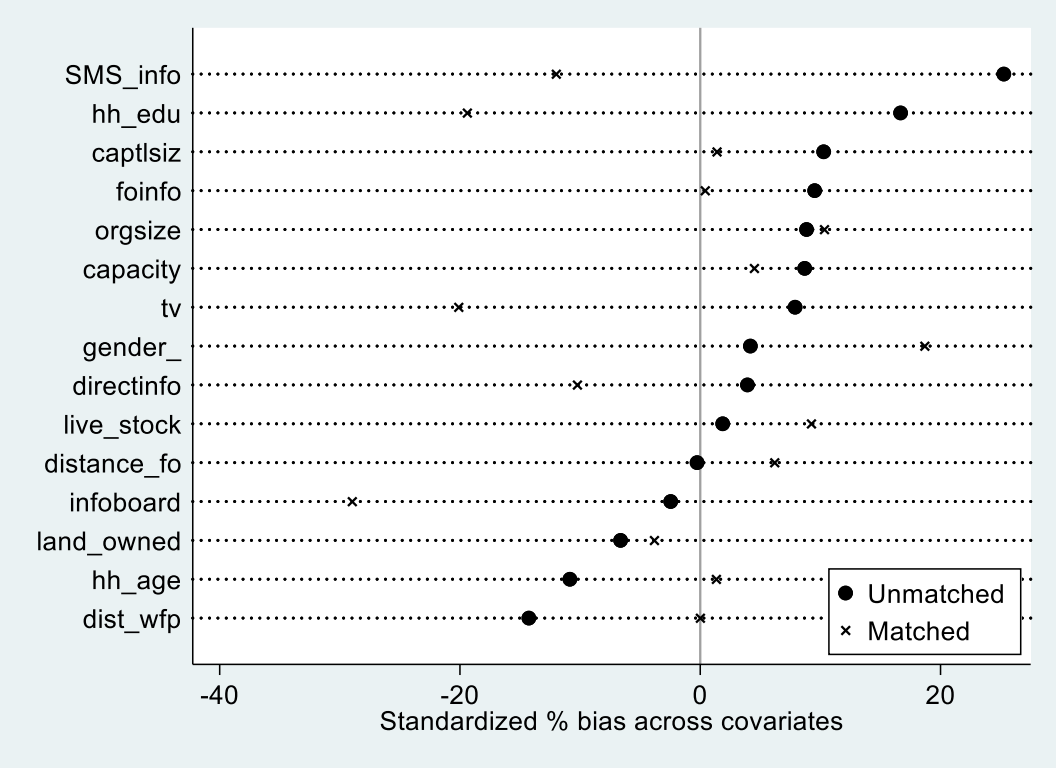


Table 2.2: Covariate Balance Assessment of PSM

	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5
Weighted variables	mean_control	mean_treated	difference	t_statistics	P_value
FO working capital (ETB)	2.7e+05	2.8e+05	8044.452	0.81	0.4207
Capacity of long-term storage (m^3)	4650.000	4771.914	121.914	0.64	0.5230
Membership size of FO	840.000	836.636	-3.364	0.19	0.8471
FO's distance from WFP warehouse	268.000	268.0049	0.0049	0.16	0.8703
FO collection point distance from the market	17.00	17.344	0.344	0.83	0.4086
Price info from direct contact with traders	0.870	0.865	-0.005	0.16	0.8703
Price information from agric information board	0.044	0.040	-0.004	0.22	0.8261
FOs as source of price information	0.385	0.401	0.017	0.36	0.7183
FO collection point distance from the market	17.00	17.344	0.344	0.83	0.4086
Household head's education in years	2.503	2.434	-0.069	0.48	0.6337
Household's head age	42.655	42.596	-0.059	0.05	0.9590
Household's land size owned	1.007	1.010	0.003	0.30	0.7646
Household head's gender (Male=1)	0.020	0.018	-0.002	0.13	0.8968
Livestock holding	2.562	2.616	0.054	0.28	0.7795
Price information from SMS/mobile phone	0.444	0.458	0.014	0.30	0.7658
Price information from radio/television	0.380	0.406	0.025	0.54	0.5876
Matched sample	520	531			

As expected, entropy balancing provided a much higher level of covariate balance than the PSM adjustment techniques. As shown in Table 3, the covariate balance is markedly improved, such that the reweighted control group now has identical means compared to the treatment group on all covariates and the standardized mean differences are all below a tolerable threshold level |0.1| (see Table 3, Column 7). Moreover, after entropy balancing, the variance and skewness of some of the covariates are also very similar, suggesting balanced distribution of the covariates in terms of higher moments. Overall, a high degree of balance in our data set is delivered by entropy balance (according to standard metrics), suggesting superior performance of this approach in randomizing the treatment assignment on the basis of the observed covariates.

Table 3: Covariate Balances Assessment of Entropy Balancing (EB)

	Treatment			Control			
Covariates weighted	mean	variance	skewness	mean	variance	skewness	stdiff_post
FO's working capital (ETB)	6974	2.71e+08	12.32	6971	4.44e+08	9.662	0.0001893
Capacity of long-term storage (m^3)	271580	8.07e+12	14.77	2712	4.60e+12	7.773	0.000146
Membership size of FO	1251	2.32e+07	14.46	1251	1108381	2.493	-0.0013151
FO's distance from WFP warehouse	272.8	4787	2.042	272.9	8431	0.8724	3.75E-06
FO collection point distance from the market	29.88	3373	5.119	29.89	3923	6.093	-0.0002773
Price info from direct contact with traders	0.9817	0.01801	-7.182	0.982	0.0180	-7.182	-0.0000354
Price information from agric information board	0.006873	0.006833	11.94	0.007	0.0068	11.94	-0.0000347
FO's as source of price information	0.9095	0.0824	-2.855	0.910	0.0824	-2.855	-0.0003218
Household head's education in years	2.05	0.6098	2.138	2.05	0.4364	2.264	-0.0004523
Household head's age	40.4	48.31	1.54	40.4	36.76	1.937	-0.0006823
Household head's gender	0.9528	0.04502	-4.271	0.965	0.03408	-5.038	-0.0003767
Household's land size owned	1.005	0.004577	14.66	1.013	0.0271	15.03	-0.0068018
Livestock holding	2.381	2.117	4.612	2.382	2.749	4.602	-0.0000898
Price information from SMS/mobile phone	0.06987	0.06507	3.374	0.069 87	0.0651	3.375	0.0004114
Price information from radio/television	0.09164	0.08334	2.831	0.091 63	0.0833	2.831	0.0002979
Observations	566	566	566	684	684	684	

Having established randomized treatment assignment on the basis of the observed covariates, we next proceed to examine the program's effect on households' welfare and investment outcomes. We start with the presentation of the treatment effect estimates of the P4P intervention on welfare outcomes among program-participating households. Table 4 reports estimates of the average welfare effects of P4P from alternative identification strategies. The first column of Table 4 presents estimates of the treatment effect from the conditional DID model.

Table 4: ATT Estimates of Welfare (per capita consumption)

Variables	Cond DID	EB	Semi-parametric DID Model_1	Model_2	Model_3	DRF ct-ols
Treated	147.7	343.3**	353.2**	348.7**	278.4*	
	(317.8)	(155.5)	(137.3)	(141.1)	(144.1)	
Gender*Treated				188.3		
				(602.0)		
Elite*Treated					1,758***	
					(283.0)	
WFP_purchase						151.89
						(214.3)
Constant		1,286***				
		(79.50)				
Observations	1,123	996	1,069	1,069	1,069	996
R-squared		0.005				

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0

The estimate of the P4P impact from this model is statistically insignificant. However, we expect that this estimate is downward biased relative to the true causal effect because identification is not yet fully achieved, partly due to the parallel trends assumption restriction and to the possible inadequacy of the covariate balance in the matching procedure of the conditional DID analysis.

However, a positive and statistically significant treatment effect estimate of ETB 343.3 ($15.60) (27.53%) per year emerges when we estimate the difference in the means of per capita expenditure (the welfare measure) between the treatment group and the reweighted control group. [19] We still expect some biases to remain because we have not yet accounted for the effects of unobservable covariates (CIA restriction) and the parallel trend assumption.

We therefore estimated the semi-parametric DID model to relax both the CIA and parallel trend assumptions. Moreover, we used this model to account for treatment effect heterogeneity across various groups within our sample, i.e., gender and membership in the FO's management team.

Average treatment effects are estimated to be positive and significant in all of the alternative specifications of this model (see Columns 3-5 of Table 4). These effects include increased household per capita consumption. Our estimates of average treatment effects ranged between ETB 278.4 ($12.65) and ETB 353.2 ($16.054) (23.3% and 28.33%) per year, depending

[19] ETB stands for Ethiopian Birr (1USD≅ETB22 in May 2016).

on the group variable that we used to account for treatment effect heterogeneity. These estimates fall within the 95% interval of the EB estimate, suggesting the effects are robust to alternative specifications.

Moreover, our analyses show evidence of treatment effect heterogeneity related to membership in the FO's management team. Specifically, we found that the program appears to earn the average P4P household with membership in the FO management team ETB 1758 ($79.91) higher per capita consumption over and above the average treatment effect of the program. This evidence lends support to Gelo and Dikgang (2019), who found that forest users' cooperatives (FUC) in South-western Ethiopia raised welfare only for households in the middle and upper ranges of income distribution, without having an effect on the bottom income quantiles. In fact, the present effect of ETB 1758 ($79.91) is of the same order of magnitude when compared to the quantile treatment effect of the aforementioned FUC at the 80th decile of income distribution. Acknowledging that the FO management team constitutes part of the FO's elite, we argue that this result points to the existence of elite capture. In fact, it indirectly lends supports to Bernard and Spielman (2009), who confirmed that decision-making within Ethiopian FOs tends to be dominated by management committees and excludes the poorest members of the organization. [20]

The incidence of elite capture within FOs has far-reaching implications for the sustainability of P4P's benefits. As opportunistic behaviour, it erodes members' trust in their cooperative. The waning of such horizontal trust (Biggeri, et.al, 2016) in turn undermines the members' commitment to their cooperatives. For the most part, this takes the form of at least some coalitions of cooperative members opting to sell their products to alternative buyers, commonly private traders, rather than to their cooperatives (Lutz and Tadesse, 2017). Thus, with declining member commitment, gains from aggregation of the P4P commodities decline, as cooperatives are limited in the quantity of these commodities that they can market, which otherwise would afford a greater aggregation benefit.

Moreover, a reduction in the supply of a P4P commodity from cooperatives to WFP, with the required quality, is not consistent with WFP's expectation that the cooperatives would be reliable providers of the food commodities demanded by WFP. The ensuing vertical mistrust between WFP and FOs undermines WFP's attempt to add value that will be fairly apportioned along the value chain of locally purchased grains for food aid.[21]

[20] Lutz and Tadesse (2017) observed that financial control of Ethiopian FOs is weak. The authors confirmed that, although a yearly audit is expected to be done, the lack of auditing officials means that audits take place only every three years. This has afforded an opportunity for the FO officials to manipulate financial reports.
[21] Value creation and its fairer appropriation along the value chain is a critical component of an inclusive market.

Apart from the heterogeneity arising from elite capture, we also tested whether this effect varies with the gender of the P4P participating smallholder. However, we find no evidence of treatment effect heterogeneity associated with the gender of the program participating households. In particular, the results show that the average household with only female representation in the FO does not enjoy more benefits than the average P4P household, suggesting that the program is gender-neutral.

In addition to evaluating the impact of the P4P program on households' welfare measured as per capita consumption, we also investigated the impact of this program on other outcomes such as food expenditure share (a proxy for vulnerability), investment in child education, and asset holding. Contrary to Lenz and Upton (2016), we find evidence of an increased share of food expenditure by 0.897 percentage points among the program participating households relative to control households (see Table 3). This positive effect of the project on food expenditure share is not consistent with our prior expectation of the inverse relationship between household vulnerability (food expenditure share) and the increased per capita income resulting from participation in the program.[22]

Moreover, we find that P4P spurred investment in child education. Specifically, our analysis shows that participation in P4P increased education-related total expenditures by between ETB 220 ($10) and ETB 237.9 ($10.81) per year (see Table 6.a). Likewise, point estimates indicate that P4P participating households spend between ETB 103.9($4.7) and ETB112.3 ($5.10) per child more on child education yearly, compared to non-program households (see Table 6.b). This amounts to a 51.02% to 55.14 % increase in per student education expenditure among program participating households.

[22] The Engel curve literature on food budget share has generally confirmed that food expenditure share falls with an increase in per capita income (expenditure) of a household (Banks et. al 1997). Given that P4P has been observed to increase household per capita income, our finding of a positive effect of P4P participation on food share is not coherent with empirical regularity. However, it is not clear whether this unexpected result is driven by the existence of a threshold effect (a specific non-linearity of the Engel curve) in the context of poor rural households, in which the negative effect of a household's income on food share emerges after a certain income threshold. We call on future research to confirm this hypothesis.

Table 6.a: ATT Estimates of Total Expenditures on Child Education

			Semi-parametric DID	
Variables	Cond DID	EB	Model_I	Model_II
Treated	-231.7	219.6***	220.8***	237.9***
	(148.1)	(65.04)	(65.18)	(74.19)
Treated*gender				536.8
				(410.9)
Constant	-237.2	507.0***		
	(174.0)	(31.47)		
Observations	1,151	1,161	1,091	1,026
R-squared		0.010		

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.

Table 6.b: ATT Estimates of Per Capita Expenditures on Child Education

			Semi-parametric DID	
Variables	Cond_DID	EB	Model_I	Model_II
treated	-46.89	85.58**	112.3***	103.9***
	(84.20)	(33.89)	(34.43)	(34.74)
Treated*gender				167.9
				(262.9)
Constant	37.03	252.8***		
	(72.25)	(16.01)		
Observations	842	814	794	794
R-squared		0.008		

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Apart from child investment, we find evidence of increased asset holding among P4P smallholders compared to control smallholders. Specifically, our EB and semi-parametric DID estimates of P4P effects on measures of asset holding (PCA index) are respectively 0.477 and 0.358 (see Table 7), both of which are statistically significant.

Table 7: ATT Estimates of Asset Holding

Variables	Cond DID	EB	Semi-parametric DID
Treated	-0.382	0.477***	0.358***
	(0.241)	(0.127)	(0.121)
Constant	-2.792***	0.104	
	(0.228)	(0.0847)	
Observations	1,143	1,154	1,084
R-squared		0.013	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The positive P4P effects on child education and asset holding attest to the program's potential to raise long-term living standards apart from its short-term welfare effects. In fact, these results are consistent with the findings of Gertler et al. (2012), who confirmed that conditional cash transfer spurred investment in higher-return (albeit risky) ventures, which helped attain higher long-term living standards among Mexican households. Moreover, the positive effect of the P4P program on child investment corroborates de Janvry et al. (2006), who found that conditional transfers helped protect school enrolment. Overall, these forward-looking behavioural responses point to the program's potential to break poverty traps, beyond directly raising short-term welfare outcomes.

Notwithstanding, a key question remains surrounding the channels through which P4P drives these effects in light of its multi-faceted nature. In other words, we need to confirm whether the P4P effects result from mechanisms related to supply-side interventions spanning increased efficiency (economies of scale and value addition) and reduced transaction costs of connecting to markets, or from mechanisms arising from demand side-interventions, such as increased equilibrium price of P4P commodity markets associated with WFP's purchases in local markets and reduced risk associated with this price (guaranteed sales, partly through forward delivery contracts). In testing these mechanisms, we estimated the effect of P4P intervention on the total transaction cost of selling maize produce as an intermediary outcome and the effect of WFP maize purchase on household welfare/income. If the null of these hypotheses are supported, we attribute the welfare and investment effects of P4P to the efficiency channel alluded to above.

In what follows, we present the treatment effect estimates of the P4P intervention on the total cost of selling maize (a measure of transaction cost) from alternative models. Table 8 reports the statistically insignificant estimate of P4P's treatment effects on the transaction cost of selling maize; this result is robust to alternative specifications (conditional DID with standard errors

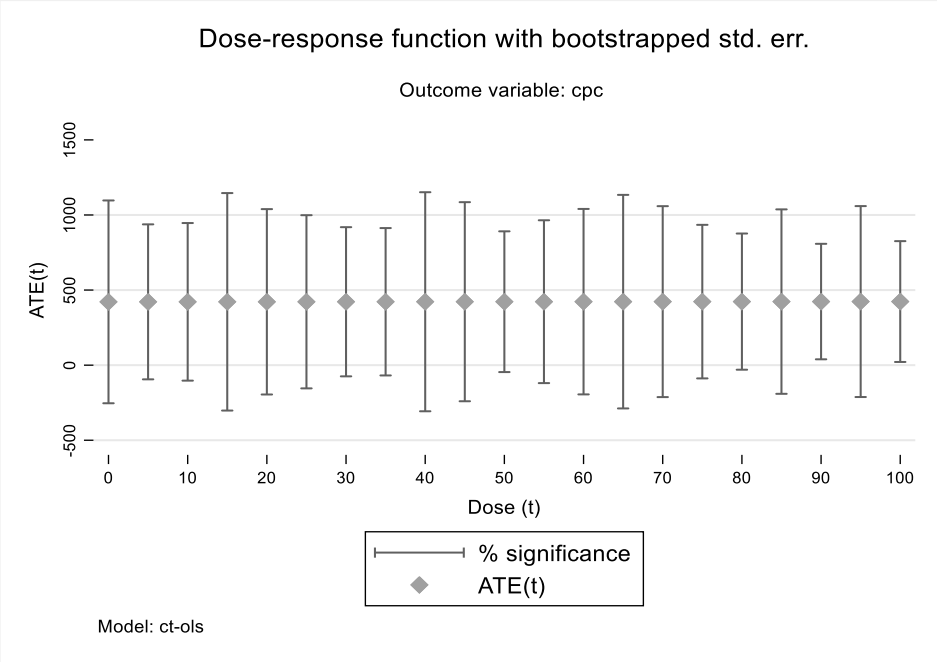
clustered at the FO level, EB model, and semi-parametric DID). Moreover, we estimated the effect of WFP's maize purchase on the welfare outcomes of P4P participating households. Our result from a dose-response function model cannot reject the null hypothesis that WFP maize purchase had no effect on the welfare outcome of P4P households (see column 6 of Table 4 and Figure 2). This is not surprising in light of a possible violation of SUTVA. Violation of SUTVA results from the presence of market spill-over effects, as non-P4P farmers and FOs are not excluded from the beneficial effects of increased equilibrium price of P4P commodities due to WFP purchase (market demand shock). This problem leads to a biased ATT estimate of our analysis, as the outcome of control farmers (untreated unit) cannot represent the counterfactual of treated farmers. In the presence of such bias, it is difficult to ascertain whether this non-significant (zero) estimate represents the lower bound of the demand-side of the P4P program or the true effect of the program. If the former is true, our reported ATT estimates represent the lower bound of P4P's total effects (effects of a combination of supply-side and demand-side interventions).

Table 8. ATT Estimates of Transaction Cost

Variables	Cond DID	EB	Semi-parametric DID
Treated	-5.174	-1.153	-1.701
	(4.490)	(2.673)	(2.800)
Constant	5.672*	9.911***	
	(3.309)	(2.319)	
Observations	1,382	1,388	1,344
R-squared		0.000	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Figure 2: Dose-Response Function Estimate with Bootstrapped Standard Errors



Otherwise, we attribute the positive estimates of smallholders' welfare and investment response to an increased efficiency of treated cooperatives, the latter arising from economies of scale (increased physical and human capital stock) and value addition (premium for meeting WFP's and customer's quality standards).[23]

Overall, our findings do not lend support to previous studies, which generally showed inconclusive evidence of the impact of P4P on a range of smallholder outcomes (welfare, food consumption score, asset holding and income) in Ethiopia, El Salvador and Tanzania. Specifically, a previous Ethiopian study found that the program households had seen income fall compared to non-participants, and insignificant income and welfare effects of the P4P on program households were found in El Salvador and Tanzania (Krieger, 2014a, b, c; Lenz and Upton, 2016).

5. Conclusion

Improving market access for smallholder farmers by reducing transaction costs and information asymmetries has been long recognized to be a priority policy question. One of the key mechanisms through which these could be achieved is by organizing smallholders in the form of FOs or by building the marketing capacity of existing FOs to increase farmers' bargaining power

[23] We thank the editor for suggesting this point and the associated analysis.

and create economies of scale. Such an intervention is expected open up opportunities to increase their incomes.

A recent intervention of the World Food Program's (WFP) Purchase for Progress (P4P) pilot initiative to link smallholders to commodity markets is a case in point. Against this backdrop, we evaluated the impact of WFP's P4P interventions, which involved investment in the capacity of Farmer Organizations (FOs) to aggregate the staple commodity and add value as well as purchasing produce from FOs using guaranteed forward contracts. The P4P procurement and capacity support intervention at the FO level afforded smallholder farmers a value-added and forward market opportunity for their staple produce, the latter of which appears to bolster their income.

Using the Ethiopian panel survey data, we estimated causal welfare and investment effects of the P4P intervention among smallholders. Our analysis shows that P4P interventions have raised per capita consumption of the participating smallholders, suggesting that the program is achieving the intended outcome of increasing smallholders' income.

Estimates of this impact from our preferred models ranged between ETB 278.4 ($12.65) and ETB 353.2 ($16.054) per year, which amounts to a 23.3% to 28.33% increase in per capita expenditure among program households relative to the counterfactual. Apart from average treatment effects on welfare outcomes, we ascertained that there is distributional bias of the welfare benefit of the P4P intervention. Specifically, the analysis showed that smallholders who are members of FOs' management capture a greater welfare benefit compared to the average participant. Acknowledging that the FO management team constitutes part of the local elite, the combination of these findings suggests the existence of elite capture which drives the distributional bias of the P4P welfare effect. On the other hand, we find that the program's welfare impact is gender-neutral.

We also find increased investment in child schooling and increased asset holding among program-participating smallholders relative to comparison farmers. In essence, the investment behavioural responses point to the program's potential to raise long-term living standards of smallholders in addition to raising their short-term welfare outcomes.

However, we also find evidence of an increase in food consumption as a share of income. We generally expect this share to fall with increased income. It's possible that the decrease in share of income does not occur until households reach some baseline of food consumption.

Overall, we find that the returns to the WFP's interventions are positive, although such benefits are not defended on equity grounds. Moreover, in the broader sense, these benefits are attributed to an increased efficiency of treated cooperatives (FOs) arising from economies of scale

(increased physical and human capital stock) and value addition (price premium for meeting WFP's and customer's quality standards) than to demand-side intervention.

The P4P's benefits would be even greater if P4P were complemented by incentives to spur members to invest in FOs. One option would be to help FOs adopt a well-defined property rights bundle, which includes a closed membership policy (restricting the access rights of non-members), establishment of a secondary market for cooperative's shares (transferable and appreciable ownership rights), and enforceable member pre-commitment mechanisms to create greater incentives to invest in the FOs. There is also a clear need for transparent and broad-based participation in FO management decisions by smallholders to assure a more equal distribution of FOs' benefits.

References

Abadie, A. (2005). Semi-parametric difference-in-differences estimators. *Review of Economic Studies, 72,* 1–19.

Abebaw, D., and Haile, M. G. (2013). The impact of cooperatives on agricultural technology adoption: Empirical evidence from Ethiopia. *Food Policy, 38,* 82–91.

Alem. Y., and Broussard. N. (2018). The impact of safety nets on technology adoption: A difference-in-differences analysis. *Agricultural Economics, 49,* 13–24.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28,* 3083–107.

Banerjee, A., Mookherjee, D., Munshi, K., and Ray, D. (2001). Inequality, control rights, and rent seeking: Sugar cooperatives in Maharashtra. *Journal of Political Economy, 109,* 138–190.

Bank. J., Blundell, R., and Lewbel, A. (1977). Quadratic engel curves and consumer demand. *The Review of Economics and Statistics, 79,* 527–539.

Bernard, T., and Taffesse, A. F. (2012). Returns to scope? Smallholders' commercialization through multipurpose cooperatives in Ethiopia. *Journal of African Economies, 21,* 440–464.

Bernard, T., and Spielman, D. (2009). Reaching the rural poor through rural producers organizations? A study of agricultural marketing cooperatives in Ethiopia. *Food Policy, 34,* 60–9.

Bernard, T., Taffesse, A. S., and Gabre-Madhin, E. (2008a) Impact of cooperatives on smallholders commercialization behaviour: Evidence from Ethiopia. *Agricultural Economics, 39,* 147–61.

Bertrand, M., Dufflo, E., and Mullainathan, S. (2004). How much should we trust difference-in-difference estimates? *Quarterly Journal of Economics, 119,* 249–275.

Barrett, C. B. (2008). Smallholder market participation: Concepts and evidence from eastern and southern Africa. *Food Policy, 33,* 299–317.

Bijman, J., and Wollni, M. (2008). Producer organizations and vertical coordination: An economic organization theory perspective. International Conference on Cooperative Studies (ICCS), 7–9 Oct., Köln, Germany.

Biggeri, M., Burchi, B., Ciani, F., and Herrmann, R. (2018). Linking small-scale farmers to the durum wheat value chain in Ethiopia: Assessing the effects on production and wellbeing. *Food Policy, 79,* 77–91.

Biggeri, M., Ciani, F., and Ferrannini, A. (2016). Aid effectiveness and multilevel governance: The case of a value chain development project in rural Ethiopia. *The European Journal of Development Research, 29,* 843–865.

Caeyers, B., Dercon, S., (2012). Political connections and social networks in targeted transfer programs: Evidence from rural Ethiopia. *Economic Development and Cultural Change, 60,* 639–675.

Caliendo, M., and Kopeinig, S. (2005). Some practical guideline for the implementation of propensity score matching. IZA Discussion Paper No. 1588.

Card, D., and Krueger, A. B. (1994), Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84: 772–793.

Cook, M. L., and Iliopoulos, C. (2000). Ill-Defined Property Rights in Collective Action: The case of US agricultural cooperatives. Chapter 22 in C. Menard (ed.) Institutions, Contracts, and Organizations: Perspectives from New Institutional Economics. Edward Elgar, Cheltenham, UK.

Cook, M. L. (1995). The future of U.S. agricultural cooperatives: A neo-institutional approach. *American Journal of Agricultural Economics, 77,* 1153–1159.

Debela. B., Shively. G., and Holden. S. (2015). Does Ethiopia's productive safety net program improve child nutrition? *Food Security, 7,* 1273–1289.

Diamond, A., and Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics, 95,* 932–945.

de. Janvry, A., Fafchamps, M., and Sadoulet, E. (1991) Peasant household behaviour with missing markets: Some paradoxes explained. *Economic Journal, 101,* 1400–17.

de Janvry. A., Finan. F., Sadoulet. E., and Vakis, R. (2006). Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics,* 349–373.

Francesconi1, G., and Heerink, N. (2010). Ethiopian agricultural cooperatives in an era of global commodity exchange: Does organizational form matter? *Journal of African Economies, 20,* 153–177.

Fitzpatrick. J., and Storey. A. (1989). Food aid and agricultural disincentives. *Food Policy, 14,* 241–247.

Gabre-Madhin, E. Z. (2001) Market Institutions, Transaction Costs, and Social Capital in the Ethiopian Grain Market. Research Report 124. Washington, DC: International Food Policy Research Institute.

Gebrehiwot. T., and Castilla. C. (2018). Do safety net transfers improve diets and reduce undernutrition? Evidence from Rural Ethiopia, *The Journal* of *Development Studies, 55:* 1947–1966.

Gelo, D., and Dikgang. J. (2019). Collective action and heterogeneous welfare effects: Evidence from Ethiopian villages. *World Development Perspectives, 16,* 100150.

Gelo, D. (2020). Forest commons, vertical integration and smallholder's saving and investment responses: Evidence from a quasi-experiment. *World Development, 132,* 104962.

Gertler, P., Martinez, S., and Rubio-Codina, M. (2012). Investing cash transfers to raise long-term living standards. *American Economic Journal: Applied Economics, 4,* 164–192.

Gezahegn, T. W., Passel, S., Berhanu, T., D'Haese, M., and Maertens, M. (2019). Big is efficient: Evidence from agricultural cooperatives in Ethiopia. *Agricultural Economics, 50,* 555–566.

Gilligan, D., and Hoddinott, J. (2007). Is there persistence in the impact of emergency food aid? Evidence on consumption, food security, and assets in rural Ethiopia. *American Journal of Agricultural Economics, 89,* 225–242.

Imai, K., and Ratkovic, M, (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society, Series B, 76,* 243–63.

Jayne, T., Strauss, J., Yamano, T., and Molla, J. (2002). Targeting of food aid in rural Ethiopia: Chronic need or inertia? *Journal of Development Economics, 68,* 247–288.

Kirsten, J., and Sartorius, K. (2002). Linking agribusiness and small-scale farmers in developing countries: Is there a new role for contract farming? *Development Southern Africa, 19,* 503–529.

Krieger, D. (2014a). The Impact of P4P on FOs and Smallholder Farmers in El Salvador. WFP:

Purchase for Progress. September.

Krieger, D. (2014b). The Impact of P4P on FOs and Smallholder Farmers in Ethiopia. WFP: Purchase for Progress. March.

Krieger, D. (2014c). The Impact of P4P on SACCOs and Smallholder Farmers in Tanzania. WFP: Purchase for Progress. March.

Hainmueller, J. (2012). Entropy balancing: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis, 20,* 25–46.

Lenz. E., and Upton, J. (2015). P4P Impact Assessments Synthesis of Preliminary Findings, WFP: Purchase for Progress. World Food Programme Report.

Lutz, C., and Tadesse, G. (2017). African farmers' market organizations and global value chains: Competitiveness versus inclusiveness. *Review of Social Economy*, *75*, 318–338.

Oyenubi, A. (2018). Quantifying balance for causal inference: An information theoretic perspective (Doctoral dissertation, University of Cape Town).

Renkow. M., Hallstrom, D. G., and Karanja, D. (2004). Rural infrastructure, transactions costs and market participation in Kenya. *Journal of Development Economics, 73,* 349–367.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Sexton, R.J. (1986). The formation of cooperatives: A game- theoretic approach with implications for cooperative finance, decision making, and stability. *American Journal of Agricultural Economics, 68,* 214–25.

Skoufias, E., and Katayama, R. S. (2011). Sources of welfare disparities between and within regions of Brazil: Evidence from the 2002-2003 household budget survey (POF). *Journal of Economic Geography*, *11* (5), 897–918.

Stiglitz, J. (2009). Moving beyond market fundamentalism to a more balanced economy. *Annals of Public and Co-operative Economics*, *80* (3), 345-360.

Stockbridge, M., Dorward, A., and Kydd, J. (2003) Farmer Organizations for Market Access: Briefing Paper. Crop Post Harvest Research Programme of the Natural Resources Research Programme of the UK Department of International Development.

Sykuta, M. E., and Cook, M. L. (2001). Cooperative and membership commitment: A new institutional economics approach to contracts and cooperatives. *American Journal of Agricultural Economics, 83,* 1273–79.

Tadesse, G., and Shively, G. (2013). Repeated transaction in rural grain markets of Ethiopia. *Journal of Development Studies*, *49*, 1172–1187.

WFP. (2016). Purchase for Progress – P4P Ethiopia. Rome (available at: http://documents.wfp.org/stellent/groups/public/documents/communications/wfp217476.pdf).

Wittenberg. M. (2017). Wages and wage inequality in South Africa 1994–2011: Part I – wage measurement and trends. *South African Journal of Economics*, 85 *(2)*, 279–297.

World Bank. (2003). Reaching the Rural Poor: A Renewed Strategy for Rural Development. Washington, D.C: World Bank.

Zhou, J., Yang, Z., Li, K., and Yu, X. (2019). Direct intervention or indirect support? The effects of cooperative control measures on farmers' implementation of quality and safety standards. *Food Policy*, *86*, 101728.