<u>ONLINE APPENDIX</u>

Subsidies and the African Green Revolution: Direct Effects and Social Network Spillovers of Randomized Input Subsidies in Mozambique

by Michael Carter, Rachid Laajaj and Dean Yang

Appendix A. Background on Study Context and Locality Groupings

The study that is the subject of this paper is nested within a larger research program on the interaction between input subsidy programs and formal savings programs. Localities in Manica province were selected to be part of the larger-scale research program on the basis of inclusion in the provincial input subsidy program as well as access to a banking program run by Banco Oportunidade de Mocambique (BOM, the implementation partner for the savings component of the research program). To be accessible to the BOM savings program, a village had to be within a reasonable distance to one of BOM's branches (including places visited weekly by a truck-mounted bank branch). These restrictions led to the inclusion of 94 localities in the research program, across the districts of Barue, Manica, and Sussundenga. Each of the selected 94 localities was then randomly assigned to either a “no savings” condition or to one of two savings treatment conditions (“basic savings” and “matched savings”), each with 1/3 probability. As we show in other work, the addition of the savings intervention creates a complex set of interactions with the input subsidy intervention, so we focus here on “no savings” localities to cleanly analyze the impact of subsidies.

The analysis of this paper focuses on the 33 localities randomly selected to be in the “no savings” condition, which did not experience any savings treatment. The mean number of study participants per locality is 15.5, with some variation (standard deviation of 10.1, interquartile range of 12). The mean number of treatment group study participants per locality is 7.4, with standard deviation of 4.9, and interquartile range of 6.

The localities we use were defined by us for the purpose of this project, and do not completely coincide with official administrative areas. We sought to create "natural" groupings of households that had a high level of connection to one another and limited connection to others. In most cases our localities are equivalent to villages, but in some cases we grouped adjacent villages together into one locality, or divided large villages into multiple localities.

Appendix B. Tests for Experimental Balance and Differential Attrition

In this section we document balance on key time-invariant household characteristics with respect to the key randomly-generated independent variables in our analyses: the indicator for assignment to the treatment group ($Treat_{ic}$) and the indicator for having above-median social network contacts in the treatment group ($SocialTreat_{ic}$). The balance test consists of estimating the following regression equation for household characteristic y_{ic} of household *i* in locality *c*:

(A1) $$y_{ic} = \alpha Treat_{ic} + \sigma SocialTreat_{ic} + \boldsymbol{X_{ic}}\gamma + \theta_c + \varepsilon_{ic}$$

This regression is a simplified version of equation (1) in the main text. The dependent variable is a time-invariant household characteristic, so there is only one observation per household; time subscripts *t* are therefore dispensed with. Now $\boldsymbol{X_{ic}}$ are a set of dummies for the number of persons (among study participants) in the social network of household i.

As discussed above, due to uncertainties in the timing of voucher distribution and delays in the creation of the list of study participants, it was not feasible to conduct a baseline survey prior to the subsidy voucher lottery. Instead, we implemented a survey after the distribution of vouchers (in April 2011), which included questions on variables that are not expected to be manipulable in response to treatment. This balance test focuses on four key household characteristics that can plausibly be considered non-manipulable: education, gender, age, and literacy of household head. Education and age are measured in years. Gender is an indicator for the head being male, and literacy is an indicator for being literate.

As in equation (1), $Treat_{ic}$ indicates treatment group households, and $SocialTreat_{ic}$ indicates the household has above-median (two or more) social network contacts who were randomized into the treatment group. The vector of controls $\boldsymbol{X_{ic}}$ includes indicators for having one, two, three, four, or five or more social network contacts who are study participants (omitted category zero). As discussed in the main text, social network size is not exogenously determined and so must be controlled for in the regression for $SocialTreat_{ic}$ to be considered exogenous. θ_c are locality fixed effects (treatment is randomized within locality). ε_{ict} is a mean-zero error term. We report robust (heteroscedasticity-consistent) standard errors.

Results are presented in Table A1. None of the coefficients on either $Treat_{ic}$ or $SocialTreat_{ic}$ are large or statistically significantly different from zero. Conditional on the $\boldsymbol{X_{ic}}$, randomization of treatment status appears to have led to balance on key household characteristics with respect to the key right-hand-side variables of interest, $Treat_{ic}$ and $SocialTreat_{ic}$.

TABLE A1—BALANCE RESULTING FROM TREATMENT ASSIGNMENT.

	Education of household head (years)	Household head male (%)	Household head age (years)	Household head lit (%)
$Treat_{ic}$	-0.059	0.015	0.26	-0.044
	[0.26]	[0.028]	[1.02]	[0.035]
$SocialTreat_{ic}$	-0.14	0.12	0.89	0.020
	[0.58]	[0.072]	[2.54]	[0.085]
Observations	475	504	491	500

Notes: Level of observation is the household. Data are from April 2011 interim survey. Summary statistics of dependent variables: education of household head, mean 4.75 (standard deviation 3.19); 84.9% of household heads male; household head age, mean 46.14 (standard deviation 13.95); 77.8% of household heads literate. Regression is as in equation A1, including locality fixed effects and dummies for one, two, three, four, or "five or more" social network contacts who are study participants. Robust standard errors in brackets.

We also examine whether treatment is related to the social network connections variables we collected in the April 2011 survey. This is important to examine, because the April 2011 survey occurred after treatment, and treatment could have affected reports of social network connections. In Table A2, each cell is a coefficient on an indicator variable that the individual was treated, and the dependent variable is a given number of social network connections (listed on the left-hand-side of the table). In the first column of the table, dependent variables in the regressions are number of social network connections in the study sample. In the second column, dependent variables in the regressions are number of treated social network connections. All coefficients on the treatment indicator are small in magnitude. Out of 15 coefficients reported in the table, only two are statistically significantly different from zero, and those at the 10% level. This is approximately what would be expected to occur by chance. All told, there is no indication that treatment affected reports of social network connections to study participants overall, or to treated study participants. This ameliorates concerns about selection bias resulting from the fact that we only collected the social network connections variables after treatment.

TABLE A2—BALANCE IN THE NUMBER OF SOCIAL NETWORK CONTACTS AND TREATED CONTACTS

X	Number of social network contacts who are study participants is equal to X	Number of social network contacts who are treated is equal to X
0	0.010 [0.048]	0.0088 [0.045]
1	-0.0015 [0.039]	-0.015 [0.029]
2	0.0068 [0.027]	0.00078 [0.030]
3	-0.041* [0.024]	0.019 [0.022]
4	0.017 [0.024]	0.018 [0.025]
5 or more	0.0082 [0.041]	-0.031* [0.018]
2 or more		0.0066 [0.039]
Number of contacts (count)	-0.0011 [0.21]	-0.092 [0.15]
Observations (in every regression)	511	511

Notes: Level of observation is the household. Data are from April 2011 interim survey. Each cell presents coefficient on $Treat_{ic}$ in a separate regression for the given social network outcome variable, including locality (stratification cell) fixed effects. Robust standard errors in brackets.

In addition to testing for baseline balance, it is important to consider attrition from the study sample. We attempted to survey everyone in the initial April 2011 sample at each subsequent survey round (in other words, attrition was not cumulative), so all attrition rates reported are vis-à-vis that initial sample. Across the three rounds, attrition rates range from 7.3% to 9.9%. In Table A3, we examine whether attrition is related to treatment assignment. The regressions are specified as in Equation 1 in the main text, with three observations per household (in 2011, 2012, and 2013). The dependent variable is an indicator for a household having attrited from a given round of the survey. None of the four coefficients are large in magnitude or statistically significantly different from zero at conventional levels. Attrition bias is therefore not a concern in this study.

TABLE A3—TEST FOR ATTRITION RELATED TO TREATMENT

		Dependent variable: Attrition indicator
Direct Impacts	During	-0.015 [0.024]
	After	0.025 [0.021]
Spillover Impacts	During	-0.013 [0.062]
	After	0.013 [0.044]
Observations		1,524

Notes: Level of observation is the household. Data are from 2011, 2012, and 2013 follow-up surveys. Dependent variable is an indicator for a household being missing from the sample in a given round. Attrition rates in the three survey rounds are 8.6%, 9.9% and 7.3% in 2011, 2012, and 2013, respectively. Regression is as in equation (1) in main text. Standard errors clustered by household in brackets.

Appendix C. Imperfect Compliance with Treatment Assignment

We have imperfect compliance with treatment assignment. First, only 40.8% of farmers in the treatment group redeemed and used their vouchers. Most such non-compliance stemmed from an inability to make the input package co-payment (even though claimed ability to pay was a selection criterion). Second, 12.4% of control group farmers reported using subsidy vouchers for the input package. Ground-level agricultural extension agents were instructed by their MinAg superiors to distribute vouchers to study participants in accordance with their randomly-determined treatment status. Control group receipt of vouchers was likely due to a mismatch in incentives between extension agents and their MinAg superiors: extension agents had quotas of vouchers to distribute, and vouchers in study localities that were unused by treatment-group farmers were supposed to have been distributed in other, non-study localities. Some extension agents apparently chose not to bear the travel and effort costs of redistributing unused vouchers in other localities, instead distributing them to some control farmers in study localities. The difference in voucher use rates in the treatment and control groups (28.8%) is statistically significantly different from zero (Table A4). The second column of the table shows that the result is not sensitive to exclusion of locality and network size fixed effects; in this case the coefficient, 0.284, is the simple difference in take-up rates across treated and control farmers. Our experiment therefore constitutes an "encouragement design".

TABLE A4 —COMPLIANCE WITH TREATMENT

	Redeemed voucher (indicator)	Redeemed voucher (indicator)
$Treat_{ic}$	0.288	0.284
	[0.0443]	[0.0374]
Locality dummies	Yes	No
Network size dummies	Yes	No
Observations	511	511
Avg. in Control group	0.124	0.124

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

Notes: Level of observation is the household. Data are from April 2011 interim survey. Dependent variable is an indicator for whether the household received and redeemed the voucher. Regression includes locality (stratification cell) fixed effects and dummies for having one, two, three, four, or "five or more" social network contacts who are study participants. Robust standard errors in brackets.

Appendix D. Definitions of Outcome Variables

The outcome variables of interest are use of fertilizer for maize, use of improved maize seeds, maize yield, expected maize yield with the technology package, and per capita consumption in the household. We describe these variables in detail here, and then turn to the definitions of the social network contacts variables.

We focus on outcomes in log transformation, to deal with extreme values of outcome variables. The log transformation of maize yield, and expected yield with the technology, which contain no zeros, is straightforward. For other variables (fertilizer and seeds) that contain zeros we add one before taking the log. In the robustness checks described in Appendix E below we show that the results are qualitatively similar when using alternative measures such as the variables in levels, or simple dummies for nonzero fertilizer and improved seed use.

Fertilizer use, improved seeds use, and yields are obtained from a section of the survey that first asks the respondent to list all plots where maize is produced, before asking further questions plot by plot. The survey asked what quantity of planting fertilizer was used, and what quantity of top dressing fertilizer was used. The fertilizer used by the household on maize in that season is obtained by summing planting and top dressing fertilizer across all plots. In a similar way, we asked for each plot what quantity of seed was used, and what type of seed was used. Possible types of seeds are local, OPV (open pollinated variety) and hybrid. A list of all the common names of OPV and hybrid seeds were provided to help identify the type of seeds with the respondent. We summed the quantities of OPV and hybrid seeds across all plots to obtain the household's use of improved maize seeds during the season.

For each plot, we asked the respondent about the area cultivated. We also asked about harvested maize production. These two questions allow the use of multiple units, to allow the respondent to use the unit that he/she is most comfortable with. We then used conversion factors to convert all areas into hectares, and all production into kilograms. Then we summed the production across all plots, summed the area across all crops, and divided the total production by the total area to obtain the average maize yield of the household.

Expected yield with the technology package is calculated as follows. We take into account that farmers perceive production to be conditional on weather conditions. We therefore aimed to estimate the distribution of potential yield that farmers have in mind across possible weather realizations. We first have the respondent specify the main parcel cultivated by the household. We then ask what production the farmer would expect if he/she used improved seeds and fertilizer in this parcel in 1) a normal year, 2) a very good year, and 3) a very bad year. We then asked the farmer to say, on average, out of 10 years, how many are very good years, how many are very bad years, and how many are normal years. We then multiply the expected production under each condition by the probability that this condition occurs according to the farmer's perception, to calculate expected production. We then divide the expected production by the area of the plot to obtain the expected yield with the technology package.

The calculation of consumption builds on the following three survey sections:

- Food expenditures over the 7 days prior to the survey (37 items)
- Regular expenditures over the 30 days prior to the survey (13 items)
- Major expenditures over the 365 days prior to the survey (14 items)

Each type of expenditure was divided by the recall period to obtain expenditures per day. The sum of expenditures per day was divided by the adult equivalent size of the household to obtain our measure of consumption per day and adult equivalent household member. Adult equivalents were obtained using weights that are a function of age and gender of each household member.

Social network connections to treatment group members are collected and defined as follows. We have data on social network links prior to treatment, based on elicitation of "information links" (Conley and Udry, 2010). In the April 2011 interim survey, study participants were presented with the full list of other study participants in the same village, and asked one by one whether they talked about agriculture with this person in the season prior to the survey (2009-10), and if so whether they did so "a bit", "moderately", or "a lot". For each study participant, others whom they indicated as having talked to about agriculture "moderately" or "a lot" are considered the participant's social network contacts. We do not require that the social network links be reciprocal. (In other words, Person A can be Person B's social network contact, as reported by Person B, even if Person B is Person A's social network contact, as reported by

Person A.) Because we are interested in understanding spillovers of our randomized treatment within the social network, this elicitation only captures social network links among study participants in the village, not the full set of social network links (which would include study non-participants).

Appendix E. Main Results and Additional Analyses

Table A5 presents the main results, represented graphically in Figure 2 of the main text. In our main regression results (Table A5 and Figure 2 in the main text), all outcome variables are expressed in natural logarithms, so that all coefficients can be interpreted in similar fashion, and to reduce the influence of outliers. In this section we also present results to confirm robustness of our findings to alternate dependent variable specifications.

TABLE A5—DIRECT AND SPILLOVER EFFECTS OF INPUT SUBSIDIES.

		Fertilizer on maize	Improved maize seeds	Maize yield	Daily consumption per capita	Expected yield with technology package
Direct Impacts	During	0.78 [0.16]	0.49 [0.15]	0.21 [0.092]	0.012 [0.049]	0.16 [0.089]
	After	0.30 [0.12]	0.099 [0.13]	0.17 [0.087]	0.091 [0.045]	0.16 [0.091]
Spillover Impacts	During	0.26 [0.32]	0.31 [0.29]	0.25 [0.18]	0.16 [0.093]	0.052 [0.17]
	After	0.73 [0.26]	0.38 [0.25]	0.40 [0.17]	0.13 [0.088]	0.42 [0.18]
Observations		1,428	1,404	1,346	1,393	1,273

Notes: Dependent variable x expressed as $\log(1+x)$ for fertilizer and improved seed outcomes (originally in kilograms, which includes zeros), and $\log(x)$ for other outcomes (data include no zeros). Maize yield originally expressed in kilograms per hectare. Daily consumption per capita originally expressed in Mozambican meticais. Expected yield with the technology is respondent's estimate of maize output (in kilograms per hectare) on household's main farming plot if using the subsidized Green Revolution technology package. Regression specification and control variables are as in equation 1 in text. Standard errors clustered by household in brackets.

In Table A6, we present coefficients from estimation of equation 1 of the main text for the same outcome variables, but with different specifications of the dependent variables. Fertilizer and seed outcomes are expressed as indicators for non-zero use and in kilograms. Maize yield and expected returns are expressed in kilograms per hectare. Daily consumption per capita is expressed in Mozambican meticais (MZN). Because of the existence of large outliers, outcome variables expressed in kilograms or MZN are truncated at the 99th percentile of that variable's distribution in the full sample. Coefficient estimates are qualitatively similar in terms of relative magnitudes and statistical significance levels when compared to the results in Figure 2 in the main text and Table A5.

TABLE A6—REGRESSIONS WITH ALTERNATE SPECIFICATIONS OF DEPENDENT VARIABLES

		Fertilizer on maize		Improved maize seeds		Maize yield	Daily consumption per capita	Expected yield with technology package
		dummy	level (kg)	dummy	level (kg)	level (kg/ha)	level (MZN)	level (kg/ha)
Direct Impacts	During	0.16	17.0	0.16	4.16	177	1.61	181
		[0.037]	[4.88]	[0.041]	[2.81]	[98.1]	[4.56]	[198]
	After	0.066	6.26	0.041	-0.16	210	10.2	308
		[0.027]	[4.03]	[0.037]	[1.96]	[102]	[4.40]	[224]
Spillover Impacts	During	0.047	9.23	0.12	2.52	364	10.6	32.0
		[0.073]	[9.33]	[0.083]	[5.00]	[172]	[8.43]	[357]
	After	0.18	18.9	0.11	5.02	679	14.9	1,285
		[0.057]	[8.78]	[0.071]	[3.93]	[253]	[8.66]	[638]
Nb of observations		1,428	1,428	1,404	1,404	1,346	1,393	1,273
Mean Control		0.22	23.2	0.48	19.3	821	78.7	1654

Notes: Level of observation is the household. Data are from 2011, 2012, and 2013 follow-up surveys. Regressions are as in equation 1 in main text. Dependent variables are as in Table A5 above, but with alternate specifications. Fertilizer and seed outcomes are expressed as indicator variables (dummies) for non-zero use and in kilograms. Maize yield and expected returns are expressed in kilograms per hectare. Consumption is expressed in Mozambican meticais (MZN). Outcome variables expressed in kilograms or MZN are truncated at the 99th percentile of that variable's distribution in the full sample. Standard errors clustered by household in brackets. The last line provides the mean of the variable (in level) among the control group in the first round.

In our main results, we estimate one "after" treatment effect, pooling the two agricultural years after the "during" subsidy year. It may also be of interest to examine direct and spillover effects for the two "after" years separately. We do this below in Table A7, which modifies regression equation (1) (from the main text) by estimating interactions of $Treat_{ic}$ and $SocialTreat_{ic}$ with two different "After" indicators for each of the post-subsidy years. As one would expect due to loss of power, a lower share of coefficients are statistically significant. There is no obvious systematic pattern to the coefficients across the two years, and in no case can we reject the hypothesis that the direct and spillover effects differ across the first and second "after" year.

TABLE A7—DIRECT AND SPILLOVER EFFECTS OF INPUT SUBSIDIES, WITH SPILLOVER EFFECTS ESTIMATED SEPARATELY FOR 1 YEAR AND 2 YEAR AFTER THE SUBSIDY ENDS

		Fertilizer on maize	Improved maize seeds	Maize yield	Daily consumption per capita	Expected yield with technology package
Direct Impacts	During	0.78 [0.16]	0.49 [0.15]	0.21 [0.092]	0.012 [0.049]	0.16 [0.089]
	1 year After	0.30 [0.14]	-0.023 [0.16]	0.22 [0.11]	0.13 [0.052]	0.27 [0.11]
	2 years after	0.30 [0.14]	0.22 [0.16]	0.13 [0.099]	0.056 [0.052]	0.070 [0.11]
Spillover Impacts	During	0.26 [0.32]	0.31 [0.29]	0.25 [0.18]	0.16 [0.093]	0.052 [0.17]
	1 year After	0.92 [0.30]	0.45 [0.31]	0.52 [0.25]	0.067 [0.10]	0.62 [0.28]
	2 years after	0.56 [0.29]	0.32 [0.31]	0.28 [0.18]	0.19 [0.11]	0.24 [0.20]
p-value of test of difference between 1 year and 2 year after	*Direct Impacts*	0.96	0.19	0.43	0.17	0.11
	Spillover Impacts	0.19	0.70	0.36	0.28	0.22
Observations		1,428	1,404	1,346	1,393	1,273

Appendix F. Alternate mechanisms for social network effects

We interpret the social network spillovers as reflecting informational spillovers from treatment group members to their social network contacts. It is important to consider whether other mechanisms may be behind these social network spillovers. The key alternate mechanisms are input sharing (treated farmers sharing fertilizer and improved seeds with their social network contacts) and resource transfers (treated farmers making monetary or goods transfers to their social network contacts). In this section we provide additional evidence and regression results testing whether these alternate channels are likely to be operative.

As explained in the main text, quantitative and qualitative evidence all point to very limited direct sharing of the inputs provided by the package. Among farmers who redeemed their voucher, 88.8%, reported they had already used the inputs for agriculture, 2.8% had not used it at the time of the survey, 1.4% sold the inputs, 1.4% declared that they used the inputs in some other way and 5.6% did not respond to this question. The requirement that vouchers be redeemed only by the original recipients themselves, with checking of names upon redemption, may have contributed to making sharing more difficult.

We can also examine whether a treated farmer's social connectedness with other treatment group members affects whether they used their subsidized inputs. If treated farmers shared their subsidized inputs with social network contacts, then having more social network contacts in the treatment group should reduce one's sharing, raising one's likelihood of reporting one had already used the inputs for the household's agriculture in the 2010-11 season. We run a regression where the dependent variable is an indicator for having "already used" one's inputs for agriculture. The sample is also restricted to the treatment group only (because control group farmers were not asked this question), so the indicator for treatment drops out of the regression. In this regression, the coefficient of interest is the coefficient on $SocialTreat_{ic}$. As shown in table A8, this coefficient is small in magnitude, and not statistically significantly different from zero. There is no indication that social connections to treatment group members affects whether one had already used the inputs for agriculture.

TABLE A8—SPILLOVER EFFECTS ON THE USE OF VOUCHER

VARIABLES	Already used voucher (dummy)
$SocialTreat_{ic}$	0.035
	[0.12]
Observations	245

Notes: Level of observation is the household. Data are from April 2011 interim survey. Only treatment group members included in sample. Dependent variable is an indicator for whether household had already used the voucher in one of its plots at the time of the survey. Regression is as in equation A1 above, but excluding the treatment indicator. Robust standard errors in brackets.

The other possible channel of social network effects is resource sharing from treated farmers to their social network contacts, thus allowing them to alleviate their credit constraint and invest in the inputs. We examine this using survey data on each study participant's sharing of resources with other households. Dependent variables are (1) an indicator for having given any money or other assistance to other households in the last 12 months (sample mean 0.39), and (2) the number of times one has given any money or other assistance to other households in the last 12 months (sample mean 1.65). These questions were asked only in the latter two follow-up surveys (2011-12 and 2012-13, the "after" period), so each household contributes up to two observations to the regression sample. We conduct regression analyses using the specification of equation A1 above, but excluding the "during" variables.

Regression results are presented in columns 1 and 3 of Table A9. If resource sharing from treated households is a channel for the social network effect, we would expect positive coefficients on $Treat_{ic}$ (treatment group members becoming more likely to provide assistance to others), and negative coefficients on $SocialTreat_{ic}$ (there is less need to provide assistance when others in the social network also receive the subsidy). None of these coefficients are large in magnitude or statistically significantly different from zero.

The resource sharing hypothesis is not perfectly tested in equation (A1), because the effect of $SocialTreat_{ic}$ on assistance provided to other households should be heterogeneous depending on one's own treatment status. If resource sharing from treated households is a channel for the social network effect, we would expect negative coefficients on $SocialTreat_{ic}$ only for households in the treatment group. We test this in columns 2 and 4 of Table A9, in which we add two interaction terms with $SocialTreat_{ic}$, one interaction with a dummy for being in the treatment group and another interaction with a dummy for being in the control group. There is no large or statistically significant coefficient between the $SocialTreat_{ic} * Treat_{ic}$ interaction term (nor on the interaction term with the control group indicator). We thus find no indication from these analyses that either input sharing or resource transfers constitute an alternate channel for the social network effects.

TABLE A9—IMPACTS ON ASSISTANCE PROVIDED TO OTHER HOUSEHOLDS

Variables	Indicator: any assistance given to other households		Number of times assisted other households	
Direct beneficiary	0.032 [0.034]	-0.0019 [0.041]	0.11 [0.10]	0.024 [0.12]
Spillover impacts	0.024 [0.064]		-0.064 [0.23]	
Spillover impacts on treatment group		-0.021 [0.073]		-0.18 [0.25]
Spillover impacts on control group		0.069 [0.071]		0.049 [0.26]
Observations	1,022	1,022	1,022	1,022

Notes: Level of observation is the household. Data are from 2012 and 2013 follow-up surveys. Dependent variables are indicator for providing money or other assistance to other households in the last 12 months (columns 1 and 2), and number of times providing such assistance in last 12 months (columns 3 and 4). Regressions in columns 1 and 3 use specification of Appendix Equation A1. Regression in columns 2 and 4 interact $SocialTreat_{ic}$ separately with indicators for households being in the treatment group and in the control group. Standard errors clustered by household in brackets.

Appendix G. Details of the Calculation of the Benefit-Cost Ratio of the Program

Here we describe how we calculate benefit-cost ratio (the ratio of the benefits to the costs) of the subsidy program. We first describe how we calculated the benefits, then how we calculated the costs. Both the benefits and costs are estimated in terms of intent-to-treat effects (ITT).

Benefits

We quantify benefits in terms of maize output, net of the increased costs of the use of the subsidized inputs (fertilizer and improved seeds), including estimates of increased labor requirements.

We break up benefits into the following four categories:

A: the direct effect of the subsidy during the subsidy period

B: the direct effect of the subsidy after the subsidy period

C: the indirect effect of the subsidy during the subsidy period (indirect refers to the effect from having at least two social network contacts in the treatment group)

D: the indirect effect of the subsidy after the subsidy period

We estimate the value of increased maize production in each of the four categories of effects. To calculate these benefits, we take the following steps:

- We estimate impacts on maize yield, fertilizer use, and use of improved seeds by estimating regressions where dependent variables are expressed in levels (Table A6).
- We transform effects on yields into monetary values by multiplying the coefficients by the average area per household (3.35 ha) to obtain the increase in maize production for an average farmer, and then we multiply the result further by the median market price of one kg of maize (5 MZN, obtained from our survey data on maize sales) to obtain the value of the additional maize production.
- We calculate the cost of the additional fertilizer and improved seeds used (retail market prices MZN 24 per kg of fertilizer, and MZN 12 per kg of improved seeds). We also place a monetary cost on the value of additional labor required when using the input package, using estimates from IFDC's agronomic trials in the study area (IFDC 2012). The cost of labor assumption is MZN 1,250 for the full technology package. In categories B, C and D, we then deduct the cost of inputs and labor from the additional value of maize production, to obtain net benefits. To be conservative in our benefit-cost ratios, we take into account any positive effect on input use, even if the coefficient in the regression in Table A6 is not statistically significantly different from zero.
- To calculate net benefits during the subsidy ("during") year for direct beneficiaries (category A), because the additional input use was subsidized, rather than using the market value of additional inputs used, we use the value of the farmer co-pay on the package as the cost to farmers. We take into account that treatment group members were 28.4 percentage points more likely to use the voucher, so to estimate the ITT effect on the cost, we multiply the household co-pay and cost of additional labor by 0.284.

- For spillover effects, we are interested in the ITT effect per subsidy voucher offered. To calculate these effects (categories C and D) we multiply the spillover coefficients of Table A6 by 0.78, which is the proportion of households whose number of social network contact with treatment group members is above the median (0.376), divided by the proportion of households in the treatment group (0.48). We also assume costs of additional labor requirements when using the inputs, in proportion to the kilograms of fertilizer used (MT 12.5 per kg of fertilizer). The cost of labor per kg of fertilizer used is derived from the IFDC (2012) estimate of MT 1,250 additional labor cost when using 100 kg of fertilizer.
- To account for time discounting, we take the present discounted value of benefits occurring in the "after" period, using a conservative (high) 10% annual discount rate.
- Estimated benefits in the various categories (including column and row totals) resulting from these calculations are presented in Table A10.

Costs

The cost of the voucher has two main components: the cost of the input subsidy, and logistical costs of identifying beneficiaries, coordinating the different actors, and distributing the vouchers. The latter includes all types of costs from staff and material to the overhead that pays for the corresponding share of the facilities of the implementing institutions. To estimate intent-to-treat costs, we consider the logistical costs for every household who won the voucher lottery, whether the household used the voucher or not. The subsidy costs themselves are only considered to be disbursed if the household received and used the voucher. We also take into account the subsidy costs arising from the fact that some individuals in the control group managed to obtain and use the voucher.

To calculate costs we take the following steps:

- Logistical costs are calculated from detailed budgets provided by implementing organizations. We first consider costs incurred at the national level by Mozambique's Ministry of Agriculture (MinAg). These costs are for activities such as training and coordination of the program, supervision, and printing of the vouchers. We divide these national costs by 25,000, the number of beneficiaries of the overall national program, to obtain a per-voucher cost. We then consider costs of the provincial authorities in Manica province (including the extension agents who selected farmers and distributed the vouchers), where the research study occurred. These costs are then divided by 5,000, the total number of vouchers in Manica (only a subset of which were for our study participants). We add up national and provincial costs per unit to obtain the logistical cost per voucher intended to be distributed, MZN 694.
- The cost of the input subsidy per voucher used is MZN 2,300 (73% of the total input package value of MZN 3,163).
- To calculate the intent-to-treat costs, we calculate how much a person selected to receive the voucher costs compared to a person not selected to receive the voucher. This cost is equal to the logistics cost (which is spent whether or not the targeted person decides to use the voucher), plus 0.284 times the cost of the input subsidy. The number 0.284 corresponds to the difference in the probability of voucher use in the treatment group (0.408) and control group (0.124).
- Total cost per voucher intended to be distributed (intent-to-treat voucher cost) is MZN 1,347.

Benefit cost ratios

Finally the benefit cost ratios presented in Table 4 in the main text are obtained by dividing the benefits over the costs calculated above. We separately consider the four categories of benefits to illustrate how this ratio varies depending on whether or not post-subsidy period effect is included and whether or not the indirect effect is included.

TABLE A10—NET BENEFITS DUE TO INPUT SUBSIDY PROGRAM.

	During	After	**Total**
Direct Net Benefits	2,363	5,621	7,985
Spillover Net Benefits	4,487	14,226	18,713
Direct and Spillover Net Benefits	6,850	19,847	26,697

Notes: All net benefits are impacts on maize output minus costs of associated increases in fertilizer, improved seeds, and labor requirements. All benefits and costs are denominated in Mozambican meticais (MZN; exchange rate MZN 27 per US$ at time of study). Direct benefits accrue from being randomly assigned to treatment group (being eligible for subsidy voucher oneself). Indirect (spillover) benefits accrue from having above-median (two or more) social network contacts randomly assigned to treatment group. Net benefits are in subsidized "during" period (2010-11 season) or post-subsidy "after" periods (2011-12 and 2012-13 seasons). Net benefits in after periods discounted back to during period using 10% annual discount rate.

Appendix References

International Fertilizer Development Center (IFDC), *Maize Intensification in Mozambique (MIM) Project: 2010-2011 Progress Report*. IFDC, International Fertilizer Association (IFA), International Plant Nutrition Institute (IPNI), and International Potash Institute (IPI). February 2012.