Appendix to
Technology adoption under uncertainty

A.1 Conceptual model

This appendix includes the formal proof of Propositions 1 through 4 in the main text. We start by characterizing agents' decisions and types in a more formal way.

A.1.1 Expected Value of Take-Up

The expected net benefit of take-up (which appears in the take-up decision inequality, equation (1)) can be rewritten as

$$\mathbb{E}_{F_1|F_0} \max(R-F_0-F_1, 0) = \Pr(R-F_0-F_1 > 0|F_0) \times \left[R - F_0 - \mathbb{E}_{F_1|F_0}(F_1|R - F_0 - F_1 > 0)\right],$$

where $\Pr(R - F_0 - F_1 > 0|F_0)$ indicates the type-specific (i.e., conditional on F_0) probability of follow-through and $R - F_0 - \mathbb{E}_{F_1|F_0}(F_1|R - F_0 - F_1 > 0)$ is the net benefit, conditional on follow-through.

A.1.2 Adoption types

Under the distributional assumptions stated in the main text:

- $F_0 \perp F_1$
- F_1 takes one of two values: $F_1 = \{f_L, f_H\},$ with $f_L < f_H,$ and $\mathbb{E}_{F_1|F_0}(F_1) = \mathrm{g}_1(f_L)f_L + \mathrm{g}_1(f_H)f_H$, where $\mathrm{g}_1(.)$ is the probability mass function of F_1
- F_0 is continuously distributed across agents with cumulative distribution function $\mathrm{G}_0(.)$,

we can classify agents in three follow-through types: non-adopters, contingent adopters and always adopters.

Non-adopters Non-adopters are characterized by the condition on F_0,

$$R - F_0 < f_L \tag{1}$$

such that even when the realization of F_1 is low (f_L), their net benefit of follow-through is negative. The share of non-adopters is given by $1 - \mathrm{G}_1(R - f_L)$. Their probability of follow-through is always 0 and so is their expected private benefit. Non-adopters take-up only if $c - A > 0$, or if the subsidy exceeds the cost of take-up. Note that even when they take-up (purchase the technology), they never follow-through.

Contingent adopters Contingent adopters are characterized by the condition

$$f_L < R - F_0 < f_H. \tag{2}$$

Contingent adopters follow-through when the realization of F_1 is F_L, but not when the realization is F_H. The share of contingent adopters is given by $\mathrm{G}_0(R - f_L) - \mathrm{G}_0(R - f_H)$, with expected private benefit given by

$$\mathbb{E}_{F_1|F_0}\left[\max(R - F_0 - F_1, 0)|R - F_0 - F_1 > 0\right] = \mathrm{g}_1(f_L)\left(R - F_0 - f_L\right)$$

where $\mathrm{g}_1(F_L)$ is their probability of follow-through. The take-up decision of these agents is characterized by condition $F_0 \leq R - f_L - \frac{c-A}{\delta g_1(f_L)}$.

Always adopters Always adopters are characterized by the condition

$$f_H < R - F_0. \tag{3}$$

Hence, they follow-through whether the draw of F_1 is f_L or f_H: $\Pr(R - F_0 - F_1 > 0|F_0) = 1$. The share of always adopters is given by $\mathrm{G}_0(R - f_H)$, and their private benefit given by

$$\mathbb{E}_{F_1|F_0}\left[\max(R - F_0 - F_1, 0)|R - F_0 - F_1 > 0\right] = R - F_0 - \mathbb{E}(F_1).$$

They take-up only if $F_0 < R - \mathbb{E}(F_1) - \frac{c-A}{\delta}$.

A.1.2.1 Selection and follow-through

Conditions (1), (2), and (3) determine thresholds over the support of F_0 that delimit the shares of always adopters, contingent adopters and never adopters for a given distribution of F_0. Figure 1 illustrates these thresholds on the probability density function of F_0, $\mathrm{g}_0(F_0)$. Note that the bell shaped distribution for F_0 shown in Figure 1 is not a necessary assumption of the model, and is used only to visualize the shares of each agent type as the area under the curve delimited horizontally by the thresholds in grey: $R - f_H$ and $R - f_L$.

The thresholds in black correspond to the take-up decision for each agent type. The take-up threshold for contingent adopters, $R - f_L - \frac{c-A}{\delta g_1(f_L)}$, is always to the right of the threshold for contingent adopters provided that the subsidy, A, is less than or equal to the total cost of the technology, c. Hence, the bigger the subsidy, A, the bigger the share who take-up, but follow-through only if $F_1 = f_L$. The take-up threshold for always adopters, $R - \mathbb{E}(F_1) - \frac{c-A}{\delta}$, may be to the left or to the right of the threshold, $R - f_H$, which defines the group of always adopters. If $\frac{c-A}{\delta} \leq f_H - \mathbb{E}(F_1)$, all always adopters will take-up. However, if $\frac{c-A}{\delta} > f_H - \mathbb{E}(F_1)$, a bigger subsidy may increase take-up among always adopters.

In sum, the subsidy A affects follow-through rates conditional on take-up by determining the shares of always adopters and contingent adopters that take up. When the subsidy is small, such that $A < c - \delta(f_H - \mathbb{E}(F_1))$, not all always adopters take-up. When the subsidy is between $c - \delta(f_H - \mathbb{E}(F_1))$ and c all always adopters take-up, but just a fraction of contingent adopters take-up. For subsidies larger than c, all always adopters, all contingent adopters and some non-adopters take-up.

Proposition 1 *Follow-through conditional on take-up increases as a function of the total (potentially subsidized) take-up cost.*

Conditional adopters are the population of interest for understanding the relationship between uncertainty and technology adoption: they constitute the only group whose follow-through decision is affected by the shock realization. The share of conditional adopters who take-up is given by

$$\frac{\mathrm{g}_1(f_L)\left[\mathrm{G}_0\left(R - f_L - \frac{c-A}{\delta g_1(f_L)}\right) - \mathrm{G}_0(R - f_H)\right] + \mathrm{G}_0\left(R - f_H\right)}{\mathrm{G}_0\left(R - f_L - \frac{c-A}{\delta g_1(f_L)}\right)} \tag{4}$$

if $\frac{c-A}{\delta} < f_H - \mathbb{E}(F_1)$ and is 100 percent if $\frac{c-A}{\delta} \geq f_H - \mathbb{E}(F_1)$. These two expressions show how follow-through depends on A through the take-up decision of the different types of agents: the larger the subsidy, A, the larger the share of contingent adopters that take-up, reducing the overall rate of follow-through among those who take-up.

Proposition 2 *An increase in uncertainty reduces follow-through conditional on take-up.*

Note that in expression (4), an increase in the spread of F_1 (distance between f_H and f_L) results in a bigger increase in the denominator than in the numerator, since part of the numerator is multiplied by $\mathrm{g}_1(f_L)$, which is a number between 0 and 1. Hence, uncertainty worsens follow-through conditional on take-up.

Proposition 3 *An increase in uncertainty weakens the relationship between take-up cost and conditional follow-through shown in Proposition 1.*

Uncertainty increases the share of contingent adopters. This is easy to see since the share of contingent adopters is determined by the probability mass over the support of F_0 between $R - f_H$ and $R - f_L$. The greater the spread of F_1, the bigger the share of contingent adopters, and the less the take-up decision predicts follow-through. In the extreme case of no uncertainty, there are no contingent adopters ($f_H = f_L$) and all we have is either always adopters or never adopters. In this case, A increases take-up among always adopters, but does not lower follow-through conditional on take-up unless adopters are paid to take-up the technology ($A > c$).

A.1.2.2 Option value of the contract

The option value associated with the take-up decision when agents are free to follow-through or not at time 1, i.e. under limited liability, is given by

$$OV(F_0) = \mathbb{E}_{F_1|F_0} \max(R - F_0 - F_1, 0) - \max\left(\mathbb{E}_{F_1|F_0}(R - F_0 - F_1), 0\right) \tag{5}$$

with $\mathbb{E}_{F_1|F_0}(R - F_0 - F_1) = R - F_0 - \mathbb{E}(F_1)$, where $\max\left(\mathbb{E}_{F_1|F_0}(R - F_0 - F_1), 0\right)$ represents the expected profit associated with making the follow-through decision at time 0, or the value of the static contract. Note that for non-adopters, the decision to not follow-through does not change with new information. Hence, $\mathbb{E}_{F_1|F_0} \max(R - F_0 - F_1, 0) = \max\left(\mathbb{E}_{F_1|F_0}(R - F_0 - F_1), 0\right) = 0$. Similarly, always adopters' decision does not change with new information. Hence, $\mathbb{E}_{F_1|F_0} \max(R - F_0 - F_1, 0) = \max\left(\mathbb{E}_{F_1|F_0}(R - F_0 - F_1), 0\right) = R - F_0 - \mathbb{E}(F_1)$. Therefore, the only group with a positive option value is contingent adopters. For them, $\mathbb{E}_{F_1|F_0} \max(R - F_0 - F_1, 0) > \max\left(\mathbb{E}_{F_1|F_0}(R - F_0 - F_1), 0\right)\mathbb{E}_{F_1|F_0} \max(R - F_0 - F_1, 0) > \max\left(\mathbb{E}_{F_1|F_0}(R - F_0 - F_1), 0\right)$, since

$$\mathbb{E}_{F_1|F_0}(R - F_0 - F_1) = R - F_0 - \mathbb{E}(F_1).$$

The share, $G_0(R - \mathbb{E}(F_1)) - G_0(R - F_H)$ would take-up under a contract with commitment (i.e., a static contract where take-up and follow-through decisions are made simultaneously at time 0) since their expected benefit under commitment, $R - F_0 - \mathbb{E}(F_1)$, is greater than zero. The share $G_0(R - F_L) - G_0(R - \mathbb{E}(F_1))$ would only take-up in the contract without commitment, since their expected benefit under commitment, $R - F_0 - \mathbb{E}(F_1$, is less than zero. Hence, for contingent adopters,

$$\max\left(\mathbb{E}_{F_1|F_0}(R - F_0 - F_1), 0\right) = \begin{cases} R - F_0 - F_1 & \text{if} \quad F_0 < R - \mathbb{E}(F_1) \\ 0 & \text{if} \quad F_0 > R - \mathbb{E}(F_1) \end{cases}$$

From the definition in (5), it follows that the option value for contingent adopters with $F_0 < R - \mathbb{E}(F_1)$ is given by

$$\begin{aligned} g_1(f_L)\,(R - F_0 - f_L) - &\quad (R - F_0 - \mathbb{E}(F_1)) \\ &= [g_1(f_L) - 1]\,(R - F_0) + g_1(f_H) f_H \\ &= g_1(f_H)\,(f_H + F_0 - R)\,; \end{aligned}$$

while the option value for contingent adopters with $F_0 > R - \mathbb{E}(F_1)$ is equal to their expected private benefit without commitment: $g_1(f_L)\,(R - F_0 - f_L)$.

In summary, the option value as a function of F_0 is given by

$$OV(F_0) = \begin{cases} 0 & \text{if} \quad F_0 > R - f_L \\ g_1(f_L)\,(R - F_0 - f_L) & \text{if} \quad R - \mathbb{E}(F_1) < F_0 \leq R - f_L \\ g_1(f_H)\,(f_H + F_0 - R) & \text{if} \quad R - f_H < F_0 \leq R - \mathbb{E}(F_1) \\ 0 & \text{if} \quad F_0 \leq R - f_H \end{cases}$$

Proposition 4

The option value associated with take-up is increasing in uncertainty, which results in higher take-up at all take-up cost levels.

For a given agent with $F_0 = f_0$, option value increases with uncertainty. As uncertainty increases (the distance between f_H and f_L), so does the likelihood that $R - f_H < f_0 \leq R - f_L$, which in turn increases the likelihood that the agent becomes a contingent complier. Hence, as uncertainty increases, the share of agents with a positive option value from take-up also increases. As expected, the option value has an asymmetric relationship with the upper and lower bounds of the shock distribution. One can increase the option value indefinitely by lowering f_L (which is equivalent to increasing the realization of the positive shock, since f_L enters as a cost in the profit function). However, lowering f_H leads to an increase in the option value up to the point where $R - \mathbb{E}(F_1) < f_0$; beyond this, the option value remains constant and equal to $g_1(f_L)\,(R - F_0 - f_L)$, which is equal to the expected private benefit of the contract to contingent adopters.

As a function of R, the option value for a given individual with $F_0 = f_0$ is zero up to the point where $R - f_L$ is larger than f_0. Beyond this, the agent becomes a contingent adopter and the option value is increasing with R up to $R = f_0 + \mathbb{E}(F_1)$, where it peaks and then falls up to $R = f_0 + f_H$. After this, the option value becomes 0 again since the value of R is large enough to guarantee follow-through.

A.2 Estimation

The estimation of the model outlined in Section 5 in the main text is done via simulated maximum likelihood.[1] This appendix details the estimation procedure used to recover the structural parameters.

A.2.1 Additional parameters

Our field experiment design included two additional treatment arms in addition to the ones described in Section 5.3: a "surprise reward treatment" group and a monitoring group. In the structural estimation, we modify the profit function to account for the variation in choices that these treatment arms introduce.

Surprise reward treatment Half of the farmers who attended training (52.5 percent) were assigned to a "surprise reward treatment" and did not learn about the threshold reward for follow-through (≥ 35 trees) until after their decision to take-up was made. As explained in Section 3.2, this treatment arm allows us to explore whether liquidity constraints explain the absence of selection effects in the data. In order to keep track of the information differences at the time of take-up in the estimation, we allow for these individuals to have a separate component in the profit from planting any positive amount of trees (a constant "surprise treatment" effect, α_S). If these individuals had identical beliefs about the costs and benefits derived from the trees (which in practice means that random parameters F_0, F_1 and T were drawn from the same distribution as those in the standard treatment, who learned about the reward before choosing to take-up), the surprise treatment effect would be zero. However, we observe reduced form evidence that there was an expectation of a higher profit among those who did not know about the reward before taking up: their take-up rate is higher than the rate among farmers who received a reward of zero. The average take-up among those in the surprise reward treatment was approximately equal to the the take-up rate of farmers in the standard treatment who drew a reward of ZMK 40,000 before they made their take-up decision.[2] Hence, in our estimation, the surprise treatment is left unrestricted and is estimated to be 91.79 (s.e. 8.11) in the main model and 54.42 (s.e. 10.235) in the model with a mean shift in F. Note that this latter coefficient is close to the reduced form effect.

Monitoring group A small share of the program participants, 15.8 percent, were randomly selected to receive regular visits to monitor tree-related activities, which allows us to more closely observe time use. This group experienced higher follow-through rates than farmers who were not assigned to the monitoring group,. Though the treatment was not designed to have an impact and the monitors were explicitly told not to communicate information about tree cultivation to the farmers, monitoring may have influenced farmers in

[1]See Train (2009).

[2]This calculation is performed from the results of a linear regression of take-up on the reward among those who had knowledge of the reward before deciding to take up.

a number of ways. For example, monitoring could have increased the subjective value of the trees by making them seem "more important" or decreased the cost of caring for them by periodically reminding individuals of their location and commitment. Farmers were not aware that they would be monitored when they made their take-up decision. In order to account for the observed effect of monitoring in the estimation, we allow the profit of those in the monitoring group to have a separate component that takes the value of zero if no trees are cultivated and of α_M when any positive number of trees is cultivated. This parameter is estimated to be 238.40 (s.e. 36.844) in the main model and 229.53 (s.e. 37.22) in the model with a mean shift in F.

A.2.2 Objective Function details under Simulated Maximum Likelihood

We use simulation methods to evaluate the objective function, equation (9) (from the main text) , for any given value of the parameters. We use simulated maximum likelihood because there are several quantities in our objective function that do not have a closed form expression.

As is usually done in random parameter models, we integrate away the unobserved random parameters when writing the analytic probabilities for each outcome. These integrals, once more, do not have a closed form solution. Hence, we use numerical integration to write the probability of choosing N trees conditional on parameters $\mu_F, \sigma_{F_0}, \sigma_{F_1}, \mu_T, \sigma_T, \alpha_S,$ and α_M. Before writing the expression for the simulated probabilities of choosing N trees, we note one more aspect of our estimation strategy.

When using simulation methods to estimate discrete choice models with random parameters, numerical integrals are used to approximate theoretical probabilities. This often results in a stepwise as opposed to smooth objective function, since small probabilities are hard to approximate numerically and can be very noisy. In order to smooth the kinks in our objective function, we add an extreme value distributed error term at the end of the profit function. This allows us to compute probabilities between 0 and 1 for each draw of the random parameters, which results in a smoother objective function. Monte Carlo simulations suggest this method will not introduce bias our results provided that we choose a relatively small scale parameter, λ, which we refer to as smoothing factor. In the estimation we use a smoothing factor of 0.5.

Thus, using Train (2009) notation, the simulated probabilities of choosing N trees at $t = 1$ are

$$\check{P}_i(N^* = n|\theta) = \frac{1}{K}\sum_{k=1}^{K} \frac{\exp\left(\frac{1}{\lambda}\Pi(n|F_{0k}, F_{1k}, T_k, R_i\right)}{\sum_{j=0}^{50}\exp\left(\frac{1}{\lambda}\Pi(j|F_{0k}, F_{1k}, T_k, R_i)\right)}$$

where k indexes each draw of the full random parameter vector, (F_{0k}, F_{1k}, T_k), given the vector of parameters $\theta = (\mu_F, \sigma_{F_0}, \sigma_{F_1}, \mu_T, \sigma_T, \alpha_S, \alpha_M)$, and farmer-specific treatments A_i and R_i.

Similarly, the simulated probability of take-up at $t = 0$ is given by

$$\begin{aligned}\check{P}_i(TakeUp|\theta) \quad &= \\ \frac{1}{K}\sum_{k=1}^{K} \quad &\mathbf{1}\left(A_i - c + \delta\check{\mathbb{E}}\left[\max_N \Pi(N|T_k, F_{0k}, F_1, T_i, R_i)\right| F_{0k}, T_k\right] > 0\Big)\end{aligned} \tag{6}$$

where k indexes each draw of the partial random parameter vector, (F_{0k}, T_k), given the vector of parameters $\theta = (\mu_F, \sigma_{F_0}, \sigma_{F_1}, \mu_T, \sigma_T, \alpha_S, \alpha_M)$, and farmer-specific treatments A_i and R_i. Note that the expected profit conditional of random variables F_0 and T, and observed treatments A and R also involves an integral without a closed form solution. We therefore use the simulated version of it in expression (6). More specifically,

$$\begin{aligned}\check{\mathbb{E}}\left[\max_N \Pi(N|T_k, F_{0k}, F_1, T_i, R_i)\right| F_{0k}, T_k] \quad &= \\ \frac{1}{M}\sum_{m=1}^{M} \max_N \quad &\Pi(N|F_{0k}, F_{1m}, T_k, R_i, A_i)\end{aligned} \tag{7}$$

where m indexes each of M draws from a normal distribution with mean F_{0k} and variance given by $\sigma^2_{F_1}$.

For estimation purposes, we use $K = 1500$ and $M = 100$. Each of the k draws are independent across observations. However, the M draws used in (7) are kept constant across observations. This reduces our computing power substantially without affecting the independence assumptions across observations (note that (7) is conditioned on F_{0k} and T_k, which are drawn independently for each farmer).

A.2.3 Maximization algorithm

In order to guarantee that the point estimates correspond to the global maximum of the likelihood function, we first conducted a grid search that would inform our starting values for the numerical maximization. The grid search was conducted over 80 thousand different combinations of the parameters and, to minimize computing time, was conducted with a lower value of K and M (400 and 50 respectively).

In addition, we conducted a three stage recursive maximization (minimization of the negative likelihood) where in each stage we maximized the simulated likelihood along a subset of the parameter vector holding the rest constant. This method worked better than the single step maximization in Monte Carlo simulations. The subsets of parameters in each of the three stages were (μ_T, σ_T, ρ), $(\sigma_{F0}, \sigma_{F1})$, and $(\mu_F, \alpha_M, \alpha_S, \mu_{shift})$ respectively.[3] The three stages were repeated sequentially until a convergence criterion involving changes in the parameter values was reached.[4] In a final stage, we used the resulting parameter estimates as starting values in a single step numerical maximization. This last step yielded small changes in the parameter values (the largest change was less than 6 percent and corresponded to the

[3] μ_{shift} corresponds to the common uniform shock in the mean shifter model discussed in Sections 5 and 6.

[4] The convergence criterion we used was that the square sum of differences between the new parameters and the starting values (the estimated parameters from the last optimization round) was less than 0.0001. The number of iterations was very robust to the critical value chosen and never reached more than four iterations.

monitoring parameter, α_M; the second largest change was of 4 percent and corresponded to the standard deviation of F_0, σ_{F0}). Appendix Table A.2.1 shows the sensitivity of our three-stage results to starting values slightly above, slightly below and at the parameter values that maximized the likelihood in our grid search.

A.2.4 Standard errors

Standard errors were computed using the variance of the numerically approximated scores, which should converge to the negative of the Hessian in the limit provided that the point estimates are the argmax of the log-likelihood function (Train 2009). We chose this method instead of the numerical Hessian because it allowed us to choose the size of the step (h) when calculating the numerical score. Simulated methods often result in "roughness" of the likelihood function, which, in our case, led to a non-positive definite numerical Hessian.[5] In order to verify that we were at a (local) minimum, we plotted the likelihood to verify its curvature along each parameter, one at a time. Appendix Table A.2.2 shows the sensitivity of our standard errors to different values of h. We chose the value of h that led to the smallest gradient.

[5]The default numerical gradient calculation also led to gradient components that far from zero. In contrast, all elements of the numerical gradient we "manually" calculated were very close to zero.

Table A.2.1: Parameter sensitivity to starting values

Panel A. Tuning parameter sets

Estimation Tuning Parameters

Name	Lambda	DiffMinChange	k	m
Set A	0.5	0.1/0.5/2	1500	100
Set B	0.5	0.05/0.05/0.5	1500	100
Set C	2	0.1/0.5/2	1500	100
Set D	0.5	0.1/0.5/2	2500	200

Panel B. Choice set of initial parameters

List of Initial Values of Parameters

μ_T	σ_T	σ_{F0}	σ_{F1}	ϱ	μ_F	α_s	α_m	μ_{Fs}
2	0.5	450	50	0.5	0	0	-100	0
3.2	1.2	500	100	0.7	100			20
3.8	1.8			0.8				50

Panel C. List of runs by initial values and parameters

	Description	Tuning Set	Initial Values of Parameters								
			μ_T	σ_T	σ_{F0}	σ_{F1}	ϱ	μ_F	α_s	α_m	μ_{Fs}
1	Values close to gridsearch	Set A	3.2	1.2	450	100	0.7	100	0	-100	20
2	Values below gridsearch optimum	Set A	2	0.5	450	50	0.5	0	0	-100	0
3	Values above gridsearch optimum	Set A	3.8	1.8	500	100	0.8	100	0	-100	50
4	Below for muT, above for rho	Set A	2	0.5	450	50	0.8	100	0	-100	20
5	Above for muT, below for rho	Set A	3.8	1.8	450	100	0.5	100	0	-100	20
6	All values close to gridsearch except for rho	Set A	3.2	1.2	450	100	0.8	100	0	-100	20
7	All values close to gridsearch except for muT, sdT	Set A	3.8	1.8	450	100	0.7	100	0	-100	20
8	All values close to gridsearch except for muF	Set A	3.2	1.2	450	100	0.7	0	0	-100	20
9	Changing DiffMinChange	Set B	3.2	1.2	450	100	0.7	100	0	-100	20
10	Changing Lambda	Set C	3.2	1.2	450	100	0.7	100	0	-100	20
11	Changing k	Set D	3.2	1.2	450	100	0.7	100	0	-100	20

Panel D. Results without Fshifter

	Description	Log-likelihood	Values of Structural Parameters at Termination								
			μ_T	σ_T	σ_{F0}	σ_{F1}	ϱ	μ_F	α_s	α_m	μ_{Fs}
1C	Values close to gridsearch	11151.17	3.292	1.262	291.12	190.22	0.658	73.54	-79.56	-134.08	-
2C	Values below gridsearch optimum	11160.82	3.132	1.213	188.23	135.91	0.614	60.20	-73.04	-107.22	-
3C	Values above gridsearch optimum	11144.84	3.535	1.399	312.11	202.28	0.803	108.10	-94.64	-224.94	-
4C	Below for muT, above for rho	11168.35	3.235	1.203	136.22	138.87	0.723	84.73	-35.43	-109.05	-
5C	Above for muT, below for rho	11152.62	3.395	1.330	297.10	202.58	0.600	99.11	-56.53	-150.44	-
6C	All values close to gridsearch except for rho	11145.23	3.534	1.364	299.24	200.99	0.775	103.28	-96.51	-221.31	-
7C	All values close to gridsearch except for muT, sdT	11147.79	3.426	1.340	297.28	199.67	0.729	97.16	-93.31	-184.77	-
8C	All values close to gridsearch except for muF	11153.41	3.269	1.225	276.41	191.96	0.684	73.60	-75.01	-147.78	-
9C	Changing DiffMinChange	11145.84	3.416	1.355	288.73	226.57	0.653	107.94	-24.03	-253.77	-
10C	Changing Lambda	# 8295.695	3.117	1.267	300.38	201.87	0.668	75.26	-81.22	-262.02	-
11C	Changing k	# 11147.7136	3.434	1.388	296.62	203.85	0.631	84.61	-57.01	-129.97	-

Panel E. Results with Fshifter

	Description	Log-likelihood	Values of Structural Parameters at Termination								
			μ_T	σ_T	σ_{F0}	σ_{F1}	ρ	μ_F	α_s	α_m	μ_{Fs}
1D	Values close to gridsearch	11153.662	3.469	1.389	279.40	181.33	0.658	51.51	-38.56	-113.99	31.75
2D	Values below gridsearch optimum	11159.819	2.977	1.169	214.61	155.62	0.365	44.00	-36.97	-107.84	37.94
3D	Values above gridsearch optimum	11141.838	3.550	1.389	301.93	195.60	0.823	76.08	-55.03	-217.74	53.61
4D	Below for muT, above for rho	11150.377	3.293	1.268	174.11	134.94	0.753	57.42	-47.36	-158.37	30.79
5D	Above for muT, below for rho	11146.150	3.319	1.256	287.15	190.26	0.592	58.72	-48.54	-222.04	48.25
6D	All values close to gridsearch except for rho	11143.875	3.510	1.365	289.78	198.01	0.772	63.46	-40.43	-189.88	47.32
7D	All values close to gridsearch except for muT, sdT	11145.224	3.400	1.320	288.39	194.84	0.721	71.01	-37.69	-231.67	52.93
8D	All values close to gridsearch except for muF	11145.255	3.381	1.332	225.21	158.53	0.661	63.87	-65.28	-167.64	31.47
9D	Changing DiffMinChange	11150.482	3.255	1.155	267.12	202.41	0.622	85.17	-37.48	-220.85	21.86
10D	Changing Lambda	# 8296.7806	3.052	1.203	293.91	197.92	0.669	55.51	-39.43	-252.52	13.73
11D	Changing k	# 11145.1056	3.444	1.390	289.29	199.59	0.629	68.82	-35.18	-173.34	45.78

Panel F. Final minimization using above results

Description	Log-likelihood	Values of Structural Parameters at Termination								
		μ_T	σ_T	σ_{F0}	σ_{F1}	ρ	μ_F	α_s	α_m	μ_{Fs}
Initial values are 3C results, gradient free algorithm	11142.064	3.539	1.401	307.87	211.42	0.818	107.58	-91.79	-238.40	-
Initial values are 3D results, gradient free algorithm	11138.996	3.579	1.392	290.06	193.05	0.835	74.48	-54.42	-229.53	53.29

Notes: This table list the set of minimization attempts conducted based on grid search results and model exploration. Panels A and B list the choice set of tuning parameters and initial values that are used. Panel C describes the tuning parameters and initial values used for both models. Panels D and E follow the same order as Panel C, listing the results and negative log-likelihood values for each run with and without the F shifter. Panel F takes the best run from Panels D and E and reports the results of an additional minimum search using a gradient free method. # indicates that the log-likelihood cannot be compared to other values.

Table A.2.2: Alternative standard error calculations

Additive h h = k	0.09	0.07	0.05	0.03	0.01	0.009	0.007	0.005	0.003	0.001	0.0005
	Values of k										
SSE(gradient)	0.00020	0.00007	0.00003	0.00030	0.00116	0.00094	0.00047	0.00015	0.00012	0.00123	0.00087
Standard Errors											
μ_T	0.0788	0.0569	0.0536	0.0446	0.0361	0.0363	0.0409	0.0555	0.0742	0.0783	0.0745
σ_T	0.0805	0.0661	0.0720	0.0441	0.0353	0.0358	0.0399	0.0499	0.0585	0.0568	0.0537
ϱ	0.0725	0.0656	0.0578	0.0549	0.0424	0.0407	0.0380	0.0346	0.0293	0.0205	0.0132
σ_{F0}	95.341	93.278	82.250	77.089	58.658	57.753	60.026	57.385	47.803	37.121	27.795
σ_{F1}	50.621	49.953	44.387	42.643	33.717	32.925	33.923	32.853	26.950	21.558	16.626
μ_F	13.083	11.822	10.323	7.398	4.578	4.268	3.875	3.441	3.319	1.129	0.587
α_s	18.244	16.222	13.914	9.604	6.066	5.725	5.054	4.715	4.839	2.945	1.478
α_m	74.432	73.887	69.830	70.173	67.778	67.349	69.312	69.247	64.716	61.108	59.251

Multiplicative h h = x * k	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001	0.0005	0.0002	0.0001
	Values of k										
SSE(gradient)	0.17430	0.00042	0.00099	0.00035	0.00146	0.00449	0.00667	0.00803	0.00840	0.00953	0.01186
Standard Errors											
μ_T	0.1847	0.1451	0.1006	0.0584	0.0438	0.0366	0.0252	0.0219	0.0209	0.0205	0.0203
σ_T	0.0741	0.0420	0.0287	0.0216	0.0170	0.0161	0.0144	0.0137	0.0136	0.0134	0.0129
ϱ	0.1285	0.1066	0.0791	0.0492	0.0374	0.0348	0.0257	0.0213	0.0172	0.0140	0.0111
σ_{F0}	12.250	111.716	114.417	98.676	80.514	56.321	37.237	28.183	22.570	17.511	14.339
σ_{F1}	6.171	61.562	63.773	54.957	45.919	33.423	20.956	16.169	13.436	10.346	7.947
μ_F	27.740	27.335	23.773	19.440	17.004	15.048	11.832	9.402	6.050	3.819	2.419
α_s	40.490	46.246	42.654	38.907	33.646	27.268	20.378	15.132	9.517	8.783	6.360
α_m	78.068	111.782	110.726	95.144	76.055	65.208	57.078	55.481	54.674	54.310	54.300

Additive h h = k	0.09	0.07	0.05	0.03	0.01	0.009	0.007	0.005	0.003	0.001	0.0005
	Values of k										
SSE(gradient)	0.00005	0.00005	0.00014	0.00037	0.00400	0.00425	0.00371	0.00215	0.00054	0.00005	0.00018
Standard Errors											
μ_T	0.0706	0.0682	0.0570	0.0443	0.0371	0.0372	0.0412	0.0486	0.0597	0.0778	0.0730
σ_T	0.0747	0.0799	0.0604	0.0553	0.0352	0.0345	0.0358	0.0379	0.0422	0.0533	0.0499
ϱ	0.0730	0.0645	0.0561	0.0475	0.0303	0.0299	0.0304	0.0301	0.0283	0.0191	0.0156
σ_{F0}	84.622	83.916	81.794	77.381	59.699	56.483	51.711	56.478	51.760	38.025	25.406
σ_{F1}	45.427	45.133	43.960	41.906	34.570	33.396	30.393	32.730	28.430	20.956	13.530
μ_F	15.470	12.581	9.894	6.373	4.062	3.654	2.976	2.176	1.549	0.534	0.296
α_s	20.470	17.634	13.734	9.510	5.561	5.035	4.021	3.800	2.566	1.402	2.400
α_m	74.444	74.339	73.968	75.463	69.887	69.394	68.504	71.408	70.636	65.708	61.092
μ_{Fs}	26.761	25.478	24.328	23.247	22.047	21.828	21.563	21.827	21.527	20.989	20.369

Panel D. F shifter standard errors -- multiplicative h

Multiplicative h h = x * k	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001	0.0005	0.0002	0.0001
	Values of k										
SSE(gradient)	0.23205	0.00143	0.00059	0.00336	0.00084	0.00107	0.00017	0.00009	0.00006	0.00003	0.00017
Standard Errors											
μ_T	0.1857	0.1565	0.1018	0.0732	0.0496	0.0378	0.0550	0.0758	0.0719	0.0618	0.0519
σ_T	0.0839	0.0464	0.0265	0.0169	0.0148	0.0178	0.0348	0.0505	0.0486	0.0416	0.0344
ϱ	0.1275	0.1045	0.0786	0.0572	0.0429	0.0321	0.0307	0.0267	0.0225	0.0173	0.0132
σ_{F0}	10.866	94.981	92.906	87.104	84.390	76.228	62.849	54.639	55.037	41.367	28.522
σ_{F1}	5.403	61.294	60.337	52.771	47.672	43.032	36.870	32.005	30.577	21.769	14.634
μ_F	41.737	63.408	55.454	43.936	33.010	26.111	15.069	12.822	7.683	7.682	5.593
α_s	44.332	42.804	41.230	35.285	32.598	26.815	18.704	13.748	8.751	6.537	4.262
α_m	69.865	97.904	95.421	92.248	83.544	78.012	71.757	68.865	73.371	66.278	61.556
μ_{Fs}	40.510	59.306	53.847	47.573	38.949	32.736	25.922	25.112	22.604	21.766	20.548

Notes: These tables display potential standard error calculations based on different estimates of the numerical derivative. Dotted boxes indicate standard errors reported in the structural parameters table. Panels A and B report the model with no F shifter, while Panels C and D correspond to the model with the F shifter. The central numerical deriviative is calculated as f(x+h)-f(x-h)/2h. In Panels A and C, the change in x given by h is additive across all variables, so h = k. In Panels B and D, the change in x is a multiplicative with respect to x, so h = x * k. These changes in x refer to the transformation of the parameters to an unbounded space -- all standard errors estimates are calculated using these inputs and the delta method. The calculations are performed on the scores of individual log-likelihoods then summed to estimate the gradient. SSE(gradient) gives the sum of squared errors of this gradient estimate, where the true value is assumed to be zero for the minimizing

A.3 Deterministic tree survival assumption

One of the assumptions in the specification of the farmer's optimization problem is that survival of trees is deterministic, conditional on effort. We allow for the cost of tree cultivation to be quadratic in the number of trees, which would capture increasing marginal costs of tree cultivation arising from increasing marginal opportunity cost of time. Our assumption on deterministic survival can be thought of as a two stage optimization process, where the farmer decides on the optimal number of trees to keep alive first, and then allocates the amount of costly effort that guarantees survival to each of those trees. This assumption is less restrictive than one would think.

The two-stage optimization process is roughly consistent with standard optimization under probabilistic survival with a few restrictions: that the probability of tree survival for a single tree as a function of effort, $p(e)$, (a) is independent across trees; (b) attains 1 at some level of effort, $\tilde{e}$, and (c) is a convex function of effort up to $\tilde{e}$; that is $\lim_{e \to \tilde{e}} p'(e) > 0$. In addition we maintain the standard interior solution assumptions of the profit function: (d) increasing and convex cost of effort, $c(e)$ (i.e. $c'(e) > 0$, $c''(e) > 0$), and (e) diminishing marginal returns to the additional tree. We can denote this last assumption as $g_i > g_{i+1}$, where g_i denotes the marginal benefit of the ith tree that survives. Assumption (c) guarantees that the optimal allocation of effort across two or more trees, given an optimal level of total effort $\bar{e}$, is such that the farmer will allocate $\tilde{e}$ to as many trees as possible up to $k\tilde{e} \leq \bar{e}$. If $k\tilde{e} < \bar{e}$, then only the last tree $(k+1)$ will be allocated the remaining effort, $\bar{e} - k\tilde{e}$, making its survival probability less than one. This optimal allocation of effort is thus consistent with deterministic survival of all trees the farmer cultivates, except for possibly the very last tree.[6]

It could be, however, that no amount of effort guarantees the survival of a given tree: i.e., the probability function reaches a maximum of $p\left(\stackrel{\approx}{e}\right) < 1$ at $\stackrel{\approx}{e}$. In order to explore whether such a model fits our data better, we simulate farmer's behavior assuming this is the case. We keep the parameters that govern farmers' heterogeneity and shocks from our estimated mean-shift model, and we add probabilistic survival to the argument of the indicator function for reaching the 35-tree threshold.[7] Table A.3.1 replicates the reduced form comparison exercise in Table 3 of the main text under this alternative assumption. For ease of comparison, Panel A shows the reduced form results using the observed data (i.e. is identical to Panel A of Table 4 in the main text). Panel B shows the reduced form results with simulated data under our baseline deterministic tree survival assumption and the estimated parameters of our mean shifter model (i.e. is identical to Panel C of Table 4). Panels C - E implement the same regressions, with simulated data from a model that keeps our estimated parameters constant (Panel B of Table 4), but models tree survival outcomes as stochastic and governed by either a binomial distribution (Panels C and D) or a beta binomial distribution (Panel E).[8] Panel

[6]The proof behind this optimal distribution of effort across trees consists of showing that there are no interior solutions to the optimization problem where more than one tree is allocated an amount of effort between 0 and $\tilde{e}$. We can prove this by contradiction for the case of two trees. The proof can be easily extended to an unlimited number of trees.

The farmer's maximization problem in the case of two trees is given by

$$\max_{e_1,e_2} \pi(e_1,e_2) = g_1 p(e_1) + g_2 p(e_2) - c(e_1+e_2)$$

where $g_1 > g_2$ (because of assumption (e)), $p(.)$ meets assumptions (a), (b) and (c), and $c(.)$ meets assumption (d).

For a solution to this problem where both trees receive an amount of effort between 0 and $\tilde{e}$ to exist (i.e. $0 < e_1^* < \tilde{e}$ and $0 < e_2^* < \tilde{e}$), the following condition needs to be satisfied

$$g_1 p'(e_1^*) - c'(e_1^*+e_2^*) = g_2 p'(e_2^*) - c'(e_1^*+e_2^*)$$

which can be simplified to

$$g_1 p'(e_1^*) = g_2 p'(e_2^*) l \tag{8}$$

Because $g_1 > g_2$, and $p''(e) > 0$ for $0 < e < \tilde{e}$, condition (8) requires that $e_1^* < e_2^*$. However, it is easy to see that given a constant total amount of effort, e^*, no optimal distribution of this effort, (e_1^*, e_2^*) will be such that $e_1^* + e_2^* = e^*$ and $e_1^* < e_2^*$ as $g_1 p'(e_1) > g_2 p'(e_2)$ for all $e_1 \leq e_2$. I.e., given a constant amount of total effort, the farmer can always do better reallocating some effort to the tree that has the higher return. Thus, no interior solution exists where more than one tree is receiving an amount of effort less than the minimum amount that guarantees survival, $\tilde{e}$.

[7]Recall that the continuous component of the profit function confounds marginal costs and benefits. Thus we cannot introduce probabilistic survival to the benefit portion, without affecting the cost per-tree, which should remain deterministic.

[8]We keep the estimated parameters under the deterministic survival assumption instead of reestimating them under the stochastic survival assumption due mainly to computing time constraints. Thus, the fit of the model may further improve if we let other parameters adjust instead of keeping them constant. However, the little sensitivity of the reduced form responses we see in Table A.3.1 leads us to believe that we would not gain much in terms of fit by reestimating the model under the stochastic survival assumption.

D assumes that the maximum probability of survival, $p(\tilde{\tilde{e}})$, is 0.98, while Panel D assumes that this maximum survival probability is 0.95. We chose relatively high probabilities for the simulation as lower probabilities result eliminate bunching at 35, and thus are inconsistent with what we observe in our data (see Figure A.3.1). The beta binomial distribution in Panel E allows for the maximum probability to vary across farmers according to a beta distribution with parameters 0.57 and 0.37. The purpose of this exercise is to examine whether by relaxing the deterministic survival assumption we can do a better job matching the reduced form results in Panel A than do our main estimates, Panel B.

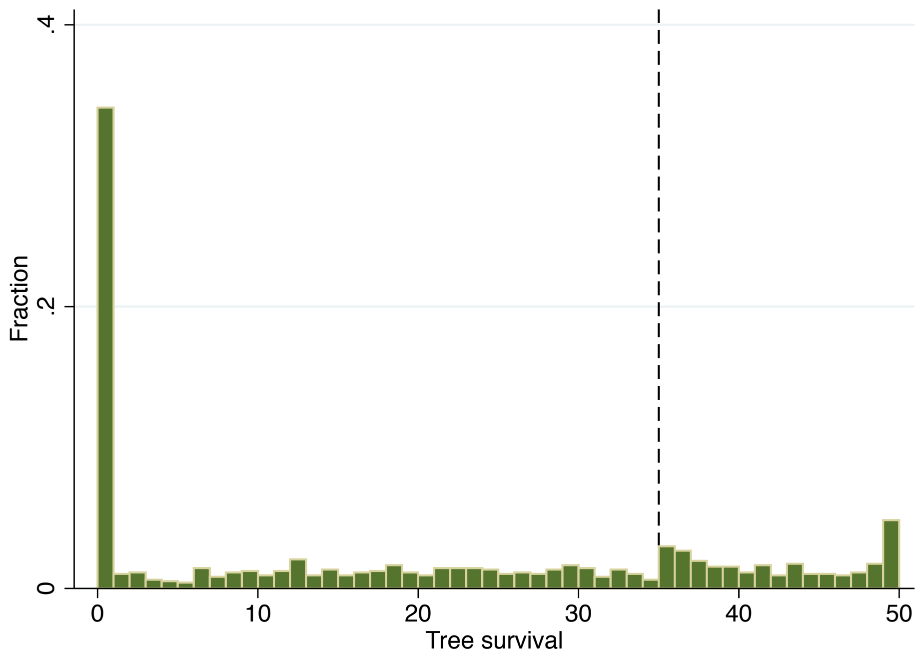
Overall, we see little improvement when stochasticity is introduced into the tree survival outcomes. The main model performs least well on the relationship between the take-up subsidy and the positive number of trees and zero trees (Panel B, columns 3 and 4). Both models overestimate the effect of the reward on the likelihood of reaching the 35-tree threshold and the number of trees for farmers with any surviving trees (Panel B, columns 6 and 7). The model variants in Panel C and D show no improvement on any of these dimensions, and in some cases worsen the fit. Only Panel E improves on the fit compared to our main model (Panel B), and not by much: the coefficients on the reward for the 35-tree threshold attainment and for positive tree survival are closer to the observed data but qualitative differences remain. Importantly, these improvement come at the expense of a poorer match in other responses that are well-fit by our main model, such as the relationship between the reward and take-up and the relationship between the reward and zero-trees cultivated (columns 5 and 8).

Table A.3.1: Stochastic tree survival

	(1) Take-up	(2) 35-tree threshold	(3) # trees | # trees>0	(4) 1.(zero trees)		(5) Take-up	(6) 35-tree threshold	(7) # trees | # trees>0	(8) 1.(zero trees)
Panel A. Observed Data (Repeats Panel A in Reduced Form Table)									
Take-up subsidy	0.022***	-0.004	-0.229	-0.003	Reward	0.001*	0.001***	0.044***	-0.001***
	(0.005)	(0.004)	(0.200)	(0.005)		(0.000)	(0.000)	(0.013)	(0.000)
Observations	1,314	1,092	701	1,092		624	1,092	701	1,092
R-squared	0.071	0.002	0.005	0.001		0.006	0.018	0.022	0.019
Panel B. Mean Shift and No Stochastic Survival (Repeats Panel C in Reduced Form Table)									
Take-up subsidy	0.020***	-0.003	-0.002	0.008**	Reward	0.001*	0.003***	0.094***	-0.001***
	(0.002)	(0.003)	(0.124)	(0.003)		(0.000)	(0.000)	(0.012)	(0.000)
Observations	1,314	1,120	605	1,120		624	1,120	605	1,120
R-squared	0.062	0.001	0.000	0.006		0.006	0.107	0.089	0.013
Panel C. Survival probability = 0.98									
Take-up subsidy	0.020***	-0.002	0.062	0.009**	Reward	0.001*	0.003***	0.105***	-0.001***
	(0.002)	(0.003)	(0.133)	(0.003)		(0.000)	(0.000)	(0.012)	(0.000)
Observations	1,314	1,120	603	1,120		624	1,120	603	1,120
R-squared	0.062	0.000	0.000	0.006		0.006	0.108	0.109	0.012
Panel D. Survival probability = 0.95									
Take-up subsidy	0.020***	-0.002	0.083	0.008**	Reward	0.001*	0.003***	0.101***	-0.001***
	(0.002)	(0.003)	(0.137)	(0.003)		(0.000)	(0.000)	(0.013)	(0.000)
Observations	1,314	1,120	596	1,120		624	1,120	596	1,120
R-squared	0.062	0.000	0.001	0.006		0.006	0.095	0.098	0.012
Panel E. Survival probability distributed beta binomal with mean 0.57 and sd 0.37									
Take-up subsidy	0.022***	-0.001	-0.038	0.006*	Reward	-0.000	0.001***	0.057***	-0.000
	(0.002)	(0.002)	(0.151)	(0.003)		(0.000)	(0.000)	(0.013)	(0.000)
Observations	1,314	1,099	518	1,099		624	1,099	518	1,099
R-squared	0.071	0.000	0.000	0.003		0.000	0.019	0.034	0.002

Notes: This table shows coefficients from regressions of each of four indicator variables (take-up, binary 35-tree threshold, tree survival larger than zero, and no tree survival) on each of our randomized treatments (take-up subsidy and threshold reward) for both non-stochastic and stochastic models. Panel A shows these regression outcomes for the true data. Panel B shows the fit of the strcutural model by simulating all four outocomes using the model estimates and examining the how much the linear relationships between outcomes and treatments resemble those in Panel A. These panels recreate Panels A and C of the reduced form table in the main body of the paper. Panel C here estimates binomial survival assuming that the probabilty any one tree survives is 0.98. Panel D estimates binomial survival assuming that the probabilty any one tree survives is 0.95. Panel E assumes that the probability any one tree survives is distributed beta binomal with mean 0.57 and standard deviation 0.37, corresponding to an alpha parameter of 0.428 and a beta parameter of 0.318.

Figure A.3.1: Observed tree survival outcomes



Notes: Histogram of tree survival outcomes for all farmers assigned a positive reward for reaching the 35-tree survival threshold. The threshold is shown by the dashed vertical line.

A.4 Tables and figures

Table A.4.1: Balance

	A=0 Mean [SD]	A > 0	R=0 Mean [SD]	Reward > 0	Surprise=0 Mean [SD]	Surprise reward	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Respondent is head of household	0.735	0.001	0.694	0.0000	0.702	0.038	1292
	[0.442]	[0.003]	[0.463]	[0.0003]	[0.458]	[0.025]	
Age, respondent	37.872	-0.074	37.439	0.0071	37.25	1.284*	1266
	[13.716]	[0.100]	[12.808]	[0.0073]	[13.684]	[0.709]	
Female headed household	0.135	0.001	0.149	0.0001	0.119	0.009	1292
	[0.343]	[0.003]	[0.358]	[0.0002]	[0.324]	[0.017]	
Years of education, respondent	5.342	-0.039	5.284	0.0033	5.363	-0.009	1292
	[3.276]	[0.028]	[3.133]	[0.0022]	[3.331]	[0.177]	
Household size	5.465	-0.036**	5.328	0.0003	5.246	0.208*	1292
	[2.142]	[0.015]	[2.390]	[0.0013]	[2.217]	[0.114]	
Ordinal discount rate (1 - 5)	2.454	-0.001	2.538	0.0004	2.43	0.106	1262
	[1.627]	[0.013]	[1.714]	[0.0010]	[1.612]	[0.095]	
Non-agricultural assets	9.806	-0.101**	9.343	-0.0013	9.08	0.314	1292
	[5.789]	[0.041]	[5.625]	[0.0034]	[5.506]	[0.287]	
Years working with Dunavant	4.228	-0.033	3.776	-0.0015	3.865	0.069	1292
	[3.748]	[0.035]	[3.404]	[0.0022]	[3.496]	[0.214]	
Total landholdings (hectares)	3.02	-0.022	2.881	0.0000	2.873	0.056	1290
	[2.248]	[0.023]	[2.188]	[0.0014]	[2.253]	[0.115]	
Number of fields	2.874	0.001	2.866	0.000	2.886	-0.054	1292
	[1.065]	[0.009]	[1.194]	[0.0008]	[1.122]	[0.070]	
Average distance from home to plots	20.532	-0.084	19.416	-0.0202	18.397	0.91	1292
	[24.195]	[0.236]	[22.436]	[0.0122]	[20.625]	[1.068]	
Poor soil fertility	0.108	-0.001	0.104	0.0000	0.106	-0.029*	1292
	[0.310]	[0.002]	[0.307]	[0.0002]	[0.308]	[0.017]	
Regular interaction with lead farmer	0.415	0.004	0.448	-0.0006**	0.412	0.007	1290
	[0.493]	[0.004]	[0.499]	[0.0003]	[0.493]	[0.029]	
Affiliated with CFU or COMACO	0.037	0.002	0.037	0.000	0.042	0.003	1292
	[0.189]	[0.002]	[0.190]	[0.0001]	[0.202]	[0.012]	
Prior knowledge of Faidherbia	0.68	-0.002	0.664	0.000	0.64	0.027	1292
	[0.467]	[0.004]	[0.474]	[0.0003]	[0.480]	[0.025]	
Prior planting of Faidherbia	0.111	-0.001	0.09	0.0001	0.088	0.014	1292
	[0.314]	[0.002]	[0.287]	[0.0002]	[0.283]	[0.014]	
Knowledge of risks to tree survival	1.72	-0.005	1.701	0.0000	1.648	-0.013	1292
	[0.905]	[0.006]	[0.785]	[0.0005]	[0.816]	[0.046]	
N	325	967	134	1041	614	678	

Notes: Means are reported for the base group in columns 1, 3 and 5. Coefficients and standard deviations from a regression of the household variable on treatment are reported in other columns. * p < 0.10 ** p < 0.05 *** p < 0.01.

Table A.4.2: Attrition across data collection phases

	Takeup Mean [SD] (1)	Baseline (2)	Endline (3)	Tree monitoring (4)
Take-up subsidy	6.1564	0.0029*	0.0000	0.0000
	[4.5399]	[0.0016]	[0.0020]	[0.0007]
Reward ('000 ZMK)	69.3347	0.0001	0.0000	0.0000
	[48.4713]	[0.0001]	[0.0001]	[0.0000]
Surprise reward treatment	0.5239	0.0097	0.0124	0.0025
	[0.4996]	[0.0088]	[0.0124]	[0.0061]
N, outcome = 1	1317	1292	1232	1083

Notes: Attrition across data collection rounds by treatment. Column 1 shows means and standard deviations for each treatment. Each cell in columns 2 - 4 shows the coefficient from a regression of an indicator being present at the data collection stage regressed on each treatment with standard errors clustered at the farmer group level. Column 4 is conditional ontake-up (N=1092). For observations missing the reward variable (surprise reward treatment, no take up), a missing variable dummy for the reward is added to the regression. Reported coefficients are among non-missing reward values. * p < 0.10 ** p < 0.05 *** p < 0.01.

Table A.4.3: Incentive spillovers within group

Dependent variable is tree survival		
	(1)	(2)
Average reward in group (excl. own)	0.262	0.578*
	(0.240)	(0.317)
Own reward	0.319**	0.642**
	(0.0569)	(0.253)
Group reward x own reward		-0.0230
		(0.0182)
N	1088	1088

Notes: OLS regressions of tree survival on average draw in farmer group, conditional on take-up, and own draw. Standard errors are clustered at the group level. * p<0.10 ** p<0.05 *** p<0.01.

Table A.4.4: Correlation between farmer observables and program outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Take-up		35-tree threshold		Tree survival	
Household head at training	0.0705**	0.0651**	0.0041	0.0010	0.4308	0.2770
	[0.0266]	[0.0229]	[0.0326]	[0.0328]	[1.2126]	[1.1939]
Female household head	0.0298	0.0275	-0.0257	-0.0248	0.2531	0.2885
	[0.0325]	[0.0292]	[0.0379]	[0.0352]	[1.4998]	[1.3798]
Respondent education	0.0009	0.0002	0.0084	0.0079	0.3926*	0.3668*
	[0.0035]	[0.0033]	[0.0044]	[0.0042]	[0.1831]	[0.1744]
Household size	0.0089	0.0104*	0.0082	0.0063	0.2118	0.1240
	[0.0051]	[0.0049]	[0.0058]	[0.0057]	[0.2066]	[0.1994]
Non-agricultural assets	0.0001	0.0017	0.0001	-0.0003	-0.0468	-0.0606
	[0.0020]	[0.0019]	[0.0030]	[0.0029]	[0.1080]	[0.1018]
Years working with Dunavant	0.0050	0.0062	0.0071	0.0077	0.1467	0.1845
	[0.0041]	[0.0034]	[0.0040]	[0.0041]	[0.1623]	[0.1598]
Land size (hectares)	0.0052	0.0052	-0.0037	-0.0043	-0.0041	-0.0261
	[0.0043]	[0.0040]	[0.0064]	[0.0062]	[0.2521]	[0.2424]
Number of fields	0.0108	0.0051	-0.0017	-0.0042	0.7444	0.6132
	[0.0100]	[0.0089]	[0.0127]	[0.0126]	[0.5520]	[0.5467]
Distance from home to plots	-0.0006	-0.0002	0.0003	0.0003	-0.0217	-0.0196
	[0.0007]	[0.0006]	[0.0007]	[0.0007]	[0.0266]	[0.0252]
Poor soil fertility	-0.0310	-0.0200	-0.0171	-0.0237	-1.7101	-1.9494
	[0.0354]	[0.0360]	[0.0517]	[0.0491]	[1.9597]	[1.9100]
Sees YGL often	0.0251	0.0250	-0.0243	-0.0142	0.5124	0.9962
	[0.0194]	[0.0181]	[0.0280]	[0.0270]	[1.1066]	[1.0389]
Affiliated with CFU or COMACO	0.0422	0.0052	0.0793	0.0635	4.9661*	4.0746
	[0.0482]	[0.0430]	[0.0721]	[0.0709]	[2.3327]	[2.3833]
Prior knowledge of Faidherbia	0.0377	0.0352	0.0423	0.0442	0.7249	0.8228
	[0.0288]	[0.0252]	[0.0344]	[0.0321]	[1.3283]	[1.1978]
Prior planting of Faidherbia	-0.0758	-0.0644	0.0643	0.0779	4.3897*	5.0115*
	[0.0461]	[0.0364]	[0.0561]	[0.0573]	[2.1425]	[2.1682]
Knowledge of risks to tree survival	0.0159	0.0157	0.0390**	0.0402**	1.6840**	1.7570**
	[0.0137]	[0.0114]	[0.0127]	[0.0128]	[0.5754]	[0.5794]
Constant	0.6230***	0.3444***	0.0542	-0.0318	8.1421**	2.6194
	[0.0648]	[0.0842]	[0.0620]	[0.0762]	[2.6058]	[3.0433]
R squared	0.0296	0.1898	0.0247	0.0750	0.0314	0.1083
Treatment controls	no	yes	no	yes	no	yes
Obs	1288		1080		1080	
Dep. Var. Mean	0.8385		0.2528		17.6000	

Notes: OLS regressions of outcomes on observables collected as part of the baseline survey, during training. The outcome in columns 3 and 4 is an indicator for reaching the reward threshold (≥ 35 trees). Even columns include controls for the experimental treatments: subsidy level, reward level, reward timing and monitoring. * p<0.10 ** p<0.05 *** p<0.01.

Table A.4.5: Effect of reward timing on tree survival outcomes

	(1) Surprise = 0 Mean/[SD]	(2) Surprise = 1 Mean/[SD]	(3) Reward x Surprise Coef/(SE)
R = 0	11.02	11.32	
	[14.33]	[16.00]	
R = (0,70000)]	14.71	15.87	0.85
	[16.86]	[16.87]	(2.72)
R = (70000,150000]	20.32	21.05	0.43
	[17.99]	[17.48]	(2.66)

Notes: Outcome is tree survival (continuous), conditional on take up. Columns 1 and 2 show means and standard deviations in each reward category, by the reward timing condition. Surprise = 1 indicates that farmers learned about the reward only after the take-up decision. Column 3 reports estimated coefficients and standard errors clustered at the farmer group level for a linear regression of tree survival on reward category interatcted with the surprise reward treatment. We report the coefficient on the interaction term only.

Table A.4.6: Knowledge and experience with the technology

	(1)	(2)	(3)	(4)
Knowledge of risks to tree survival (1-5)	4.5035***			4.4380***
	[0.7828]			[0.7762]
Change in risk knowledge	2.4856***			2.4897***
	[0.4819]			[0.4779]
Prior planting of Faidherbia		5.1819**	4.4554**	4.0145**
		[2.1442]	[1.9009]	[1.9328]
Any prior planting in group			1.5780	1.8883
			[1.7852]	[1.7111]

Notes: OLS regressions of tree survival on proxy measures of knowledge and learning. The sample is restricted to farmers in both the baseline and endline surveys. Knowledge of risks to tree survival counts the number of risks that the farmer was able to recall. Change in risk knowlege measures how that number changed between the endline and the baseline. Prior planting of Faidherbia is an indicator for whether the farmer had adopted the technology prior to the program. Any prior planting in group is an indicator for whether anyone in the farmer group had adopted. All columns control for the remaining variables shown in the balance tables and also for treatment variables. * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table A.4.7: Procrastination

Dependent variable:	Take-up	Survival	Survival	Take-up	Take-up
	(1)	(2)	(3)	(4)	(5)
	Panel A: Self-described procrastinator				
Binary Procrastination Measure	-0.0080	-0.8574	-1.3361	-0.0273	-0.0016
	[0.0208]	[1.1563]	[1.1410]	[0.0440]	[0.0651]
Take-up subsidy	0.0221***	-0.0513	-0.0184	0.0206***	0.0233***
	[0.0044]	[0.2004]	[0.1952]	[0.0047]	[0.0058]
Reward in '000 ZMK			0.0669***		0.0007**
			[0.0114]		[0.0003]
Procrastination x subsidy				0.0031	-0.0028
				[0.0047]	[0.0069]
Constant	0.4507***	9.0534***	4.3205	0.4622***	0.3980***
	[0.0783]	[3.1850]	[3.1716]	[0.0804]	[0.1088]
N	1275	1071	1071	1275	603
	Panel B: Reports procrastination on other activities				
Binary Procrastination Measure	-0.0302	0.0622	0.1404	-0.0938	-0.0521
	[0.0278]	[1.2728]	[1.2767]	[0.0629]	[0.0767]
Take-up subsidy	0.0231***	-0.1045	-0.0745	0.0197***	0.0205***
	[0.0044]	[0.2021]	[0.1958]	[0.0043]	[0.0050]
Reward in '000 ZMK			0.0686***		0.0006*
			[0.0119]		[0.0003]
Procrastination x subsidy				0.0102	0.0076
				[0.0066]	[0.0080]
Constant	0.4589***	8.8880***	3.8534	0.4766***	0.4333***
	[0.0801]	[3.1334]	[3.1569]	[0.0806]	[0.1014]
N	1223	1030	1030	1223	576

Notes: OLS regressions of take up and survival on indicators of procrastination. Standard errors clustered at the group level are in brackets. Columns 2 and 3 condition on take-up. Column 5 conditions on knowing the reward before take-up (excludes the surprise reward treatment). See text for a description of the procrastination measures used in the regressions. * p<0.10 ** p<0.05 *** p<0.01.