Online Appendix

Learning Through Noticing: Theory and Experimental Evidence in Farming

Rema Hanna
Harvard University, NBER and BREAD

Sendhil Mullainathan
Harvard University and BREAD

Joshua Schwartzstein
Dartmouth College

A Proofs

A.1 Notation and Preliminary Lemmas

Before turning to the proofs of the propositions, we introduce other notation and useful lemmas.

To highlight the role that mixed strategies play in some of the analysis, let $\sigma_{jt} \in \Delta(X_j)$ denote the farmer's input choice along dimension j at time t. Now define

$$v_{jt} = v_j(\sigma_{jt}, a_{jt}; \theta) = \sum_{x'_j} \sigma_{jt}(x'_j) f_j(x'_j|\theta_j) - e \cdot a_{jt}$$

to equal net payoffs along dimension j in period t as a function of the farmer's decisions that period and the (unknown) technology θ, where σ_{jt} necessarily takes on the default uniform distribution when $a_{jt} = 0$. We denote the default distribution by σ_j^d. The farmer's flow payoff in period t then equals $v_t = \sum_j v_{jt}$, so what he is maximizing is the expected value (under his prior) of $v_1 + v_2$.

We denote the maximum attainable payoff given technology θ as $v^*(\theta) = \max_{\sigma,\mathbf{a}} v(\sigma,\mathbf{a};\theta) = \sum_j v_j^*(\theta_j)$, where each $v_j^*(\theta_j)$ is defined in the obvious manner, and arguments that attain this maximum as $(\sigma^*(\theta),\mathbf{a}^*(\theta))$. For a final bit of notation, denote the total loss in profits from period-t farming practices relative to the (static) optimum by $L_t = v^*(\theta) - v_t$ and the loss along dimension j as $L_{jt} = (v_j^*(\theta_j) - v_{jt})$.

An optimal strategy of the farmer entails that in the second (terminal) period, for every input dimension j he selects (σ_j, a_j) to maximize $E[v_j(\sigma_j,a_j;\tilde{\theta}_j)|\hat{h}]$, given any recalled history $\hat{h}$. Let $(\sigma_j(\hat{h}),a_j(\hat{h}))$ denote such an optimal choice of (σ_j,a_j). It will be useful to compare $(\sigma_j(\hat{h}),a_j(\hat{h}))$ to the farmer's optimal terminal period decisions if he (hypothetically) perfectly knew the underlying technology, θ.[1]

Lemma A.1.

1. *If the farmer attended to input j in the first period and followed a first period strategy in which he placed positive probability on each input in X_j on every plot of land, then he learns to optimally set input j given the underlying technology θ: If $a_{j1} = 1$ and $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$, then $(\sigma_j(\hat{h}),a_j(\hat{h})) = (\sigma_j^*(\theta),a_j^*(\theta))$.*

2. *If the farmer did not attend to input j in the first period, then he will not attend to input j in the second period, and consequently may not optimally set input j given the underlying technology θ: If $a_{j1} = 0$ and $e > 0$, then $(\sigma_j(\hat{h}),a_j(\hat{h})) = (\sigma_j^d,0)$.*

Proof. 1. Suppose the farmer attended to input j in the first period and followed a first period strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$. Then, for all $x_j \in X_j$, the farmer can calculate the empirical average gain from using x_j relative to not attending, $\bar{g}_j(x_j)$, which reveals the population expected static gain from using x_j relative to not attending:

$$\bar{g}_j(x_j) = \big(\bar{y}_j(x_j) - e\big) - \frac{1}{|X_j|}\sum_{x'_j}\bar{y}_j(x'_j) = \big(f_j(x_j|\theta_j) - e\big) - \frac{1}{|X_j|}\sum_{x'_j} f_j(x'_j|\theta_j).$$

[1]The notation ignores knife-edge cases where the farmer is indifferent between various options in the second period. To handle these cases, just consider an arbitrary tie-breaking rule.

The farmer's posterior will thus place probability 1 on $\tilde{\theta}$ satisfying $\big(f_j(x_j|\tilde{\theta}_j)-e\big)-\frac{1}{|X_j|}\sum_{x'_j} f_j(x'_j|\tilde{\theta}_j) = \big(f_j(x_j|\theta_j)-e\big)-\frac{1}{|X_j|}\sum_{x'_j} f_j(x'_j|\theta_j)$ for all $x_j \in X_j$.

The farmer then optimally sets

$$\begin{aligned} a_j(\hat{h}) &= \begin{cases} 1 & \text{if } \max_{x'_j \in X_j} \bar{g}_j(x'_j) > 0 \\ 0 & \text{otherwise} \end{cases} \\ &= a^*_j(\theta). \end{aligned}$$

Further, for recalled histories satisfying $a_j(\hat{h}) = 1$, he sets $\sigma_j(\hat{h})[x^*_j(\hat{h})] = 1$, where $x^*_j(\hat{h}) \in \arg\max_{x'_j \in X_j} \bar{g}_j(x'_j)$, which implies that $\sigma_j(\hat{h}) = \sigma^*_j(\theta)$.

2. For any $\hat{h}$ consistent with not attending to input j in the first period, there exists a constant $K \in \mathbb{R}$ such that

$$E[f_j(x'_j|\tilde{\theta}_j)|\hat{h}] = E[\tilde{\theta}_j(x'_j)|\hat{h}] = K, \ \forall x'_j \in X_j. \tag{1}$$

Equation (1) follows from the fact that the farmer's marginal prior distributions over $\tilde{\theta}_j(x'_j)$ are the same across $x'_j \in X_j$, as are the likelihood functions $\Pr(\hat{h}|\tilde{\theta}_j(x'_j))$ given any $\hat{h}$ consistent with not attending to input j.

From (1), the farmer's terminal-period expected payoff function along dimension j satisfies

$$E[v_j(\sigma^d_j, 0)|\hat{h}] = K = E[v_j(x_j, 1)|\hat{h}] + e$$

$\forall\, x_j \in X_j$, given any $\hat{h}$ consistent with the farmer not attending to input j, which implies that $(\sigma_j(\hat{h}), a_j(\hat{h})) = (\sigma^d_j, 0)$ for each $\hat{h}$ when $e > 0$.

■

The first part of Lemma A.1 considers what happens when the farmer attends to input j in the first period and follows a first period strategy in which he places positive probability on each input in X_j on every plot of land, and shows that in this case the farmer learns to optimally set input j. This result leaves open the question of what happens if the farmer attends to input j, but follows a strategy in which he does not experiment with certain inputs in X_j. The next result shows that we need not concern ourselves with such strategies: among strategies where the farmer attends to input j in the first period, it is without loss of generality to confine attention to strategies where he experiments with each input in X_j on a positive measure of plots of land.

Lemma A.2. *If it is optimal for the farmer to attend to input j in the first period, then he optimally follows a strategy under which $\sigma_{j1}(x_j') > 0$ for all $x_j' \in X_j$.*

Proof. The farmer's expected payoff if he attends to input j equals

$$E[v_{-j1} + v_{-j2}] + E[v_j(\sigma_{j1}, 1; \tilde{\theta}_j) + v_j(\sigma_{j2}, a_{j2}; \tilde{\theta}_j)] \leq E[v_{-j1} + v_{-j2}] + E[-e + v_j^*(\tilde{\theta}_j)],$$

where $v_{-jt} \equiv \sum_{i \neq j} v_{it}$, and the upper bound follows from the fact that $E[\tilde{\theta}_j(x_j')] = 0\ \forall x_j'$ and, for all $\tilde{\theta}_j$, $v_j(\sigma_{j2}, a_{j2}; \tilde{\theta}_j) \leq v_j^*(\tilde{\theta}_j)$ by the definition of $v_j^*(\tilde{\theta}_j)$. But, by Lemma A.1.1, this upper bound is attained through a strategy under which $\sigma_{j1}(x_j') > 0$ for all $x_j' \in X_j$ (and clearly cannot be attained through any strategy under which $\sigma_{j1}(x_j'') = 0$ for some $x_j'' \in X_j$, since, in this case, $\sigma_j(\hat{h}) \neq \sigma_j^*(\theta_j)$ with positive probability).

■

In the remaining proofs and extensions, we will make use of the following measure of the degree to which a dimension is relevant.

Definition A.1. Input dimension j is *K-relevant to using the technology θ* if

$$\max_{\tilde{x}_j} f_j(\tilde{x}_j|\theta_j) - \frac{1}{|X_j|} \sum_{x_j'} f_j(x_j'|\theta_j) > K.$$

Otherwise, dimension j is *K-irrelevant to using the technology*.

Input dimension j is K-relevant to using the technology if optimizing the dimension yields an expected payoff of at least K relative to selecting the input along that dimension at random. We say that a dimension is *relevant* when it is 0-relevant.

A.2 Proofs of Propositions

Proof of Proposition 1. Because the payoff function is separable across input dimensions and the prior uncertainty is independent across dimensions, we can separately analyze the farmer's decisions across dimensions. For a given input j, the farmer's first period decision boils down to choosing between:

1. Not attending to input j in the first period.
2. Attending to input j in the first period, and following a strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$.

(By Lemma A.2, we don't need to consider strategies in which the farmer attends to input j in the first period and sets $\sigma_{j1}(x'_j) = 0$ for some x'_j.)

If the farmer does not attend to input j in the first period, then his expected payoff along dimension j equals:

$$
\begin{aligned}
E[v_j(\sigma_j^d, 0; \tilde{\theta}_j) + v_j(\sigma_{j2}, a_{j2}; \tilde{\theta}_j)] &= 2E[v_j(\sigma_j^d, 0; \tilde{\theta}_j)] \\
&= 2E\left\{ \frac{1}{|X_j|} \sum_{x'_j} \tilde{\theta}_j(x'_j) \right\} \\
&= 0, \qquad (2)
\end{aligned}
$$

where the first equality follows from Lemma A.1.2 and the last from the fact that $E[\tilde{\theta}_j(x'_j)] = 0$ for all $x'_j \in X_j$.

Conversely, if the farmer attends to input j in the first period and follows a strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$, then his expected payoff along dimension j equals:

$$
\begin{aligned}
E[v_j(\sigma_{j1},1;\tilde{\theta}_j)+v_j(\sigma_{j2},a_{j2};\tilde{\theta}_j)] &= E[v_j(\sigma_{j1},1;\tilde{\theta}_j)+v_j^*(\tilde{\theta}_j)] \\
&= -e+E[v_j^*(\tilde{\theta}_j)] \\
&= -e+\pi_j\cdot E[v_j^*(\tilde{\theta}_j)|j \text{ relevant}] \\
&= -e+\pi_j\cdot\{\Pr(\ j\ e\text{-relevant }|j \text{ relevant})\cdot E[v_j^*(\tilde{\theta}_j)|\ j\ e\text{-relevant}] \\
&\quad +\Pr(\ j\ e\text{-irrelevant }|j \text{ relevant})\cdot E[v_j^*(\tilde{\theta}_j)|\ j\ e\text{-irrelevant}]\} \\
&= -e+\pi_j\cdot\left\{\Pr(\ j\ e\text{-relevant }|j \text{ relevant})\cdot E[v_j^*(\tilde{\theta}_j)-\bar{v}_j(\tilde{\theta}_j)|\ j\ e\text{-relevant}]\right\},
\end{aligned} \tag{3}
$$

where the first equality follows from Lemma A.1.1, the second from the fact that $E[\tilde{\theta}_j(x'_j)] = 0$ for all $x'_j \in X_j$, the third from the observation that $v_j^*(\theta_j) = 0$ if j is irrelevant, the fourth from expanding the expression using Bayes' rule, and the final from subtracting off $0 = E[v_j(\sigma_j^d,0;\tilde{\theta}_j)] \equiv E[\bar{v}_j(\tilde{\theta}_j)]$ and observing that $E[v_j^*(\tilde{\theta}_j)-\bar{v}_j(\tilde{\theta}_j)|\ j\ e\text{-irrelevant}] = 0$.

By comparing Equations (2) and (3), we see that the farmer chooses to attend to input j if and only if the right hand side of Equation (3) is positive. This is clearly true when $e = 0$ and $\pi_j > 0$ (we have that $E[v_j^*(\tilde{\theta}_j)-\bar{v}_j(\tilde{\theta}_j)|\ j\ e\text{-relevant}] > 0$), so in this case the farmer attends to all input dimensions in the first period. Combined with Lemma A.1.1, this establishes part 1 of the proposition.

For part 2a, recall the assumption that at least one dimension, say j, is worth attending to, i.e., is e-relevant. For π_j sufficiently low, the right-hand-side of Equation (3) is negative and the farmer does not attend to input j in the first period. By Lemma A.1, this implies that the farmer does not optimize along dimension j, where the loss from not optimizing equals

$$L_{j2} = \left(\max_{\tilde{x}_j} f_j(\tilde{x}_j|\theta_j)-e\right)-\frac{1}{|X_j|}\sum_{x'_j} f_j(x'_j|\theta_j) > 0$$

and the inequality follows from the assumption that j is e-relevant.

For part 2b, fix some $K \in \mathbb{R}^+$. Under the assumption that at least one dimension j is worth attending to, there exists a dimension j where $|X_j| \geq 2$. We proceed by establishing that there exists a θ' under which j is e-relevant and $L_{j2} > K$, which implies that the farmer loses at least K from not optimizing when π_j is sufficiently small. Consider some θ', where

$$\theta'_j = (K+e, \underbrace{-K-e-\varepsilon, -K-e-\varepsilon, \ldots, -K-e-\varepsilon}_{|X_j|-1 \text{ times}}),$$

given an $\varepsilon > 0$. Plugging this θ'_j into the above formula for L_{j2} yields

$$\begin{aligned} L_{j2} &= K+e-e-\frac{1}{|X_j|}\big(K+e-(|X_j|-1)\cdot(K+e+\varepsilon)\big) \\ &= K+\frac{1}{|X_j|}\left[\varepsilon+(|X_j|-2)_+\cdot(K+e+\varepsilon)\right] \\ &> K, \end{aligned}$$

where $(\cdot)_+$ denotes an operator where $(Y)_+ = Y$ if $Y \geq 0$ and $(Y)_+ = 0$ otherwise.

For part 3, $a_{j1} = 1$ implies that $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$ (by Lemma A.2), which implies that the farmer learns to optimize dimension j (by Lemma A.1.1).

■

Proof of Proposition 2. First consider dimensions that the farmer attended to in the first period. The farmer learns to optimize such dimensions by Lemmas A.1 and A.2.

Next consider dimensions that the farmer did not attend to in the first period. Because the farmer did not attend to these dimensions, he uses each input in X_j infinitely often in the first period: $a_{j1} = 0$ implies that $\sigma_{j1} = \sigma^d_j$, where $\sigma^d_j(x'_j) > 0$ for all $x'_j \in X_j$. As a result, the empirical average gain from using x_j relative to not attending, $\bar{g}_j(x_j)$, reveals the population expected static gain from using x_j relative to not attending:

$$\bar{g}_j(x_j) = \left(\bar{y}_j(x_j) - e\right) - \frac{1}{|X_j|}\sum_{x'_j}\bar{y}_j(x'_j) = \left(f_j(x_j|\theta_j) - e\right) - \frac{1}{|X_j|}\sum_{x'_j} f_j(x'_j|\theta_j).$$

While the farmer cannot compute $\bar{g}_j(x_j)$ for every x_j, he need only compute $\bar{g}_j(\tilde{x}^*_j)$ to make optimal decisions in the second period, which he can derive from the summary information together with $\hat{h}$: $\bar{g}_j(\tilde{x}^*_j) = \left(\bar{y}_j(\tilde{x}^*_j) - e\right) - \bar{y}$, where $\bar{y} = \int y_{l1} dl = f_j(\sigma^d_j|\theta_j) + \sum_{k\neq j} f_k(\sigma_{k1}|\theta_k)$ equals the empirical average yield across all plots of land. The farmer's posterior will thus place probability one on $\tilde{\theta}$ satisfying $\left(\max_{x_j} f_j(x_j|\tilde{\theta}_j) - e\right) - \frac{1}{|X_j|}\sum_{x'_j} f_j(x'_j|\tilde{\theta}_j) = \bar{g}_j(\tilde{x}^*_j) = \left(\max_{x_j} f_j(x_j|\theta_j) - e\right) - \frac{1}{|X_j|}\sum_{x'_j} f_j(x'_j|\theta_j)$ and $\tilde{x}^*_j \in \arg\max_{x_j} f_j(x_j|\tilde{\theta}_j)$, where $\tilde{x}^*_j \in \arg\max_{x_j} f_j(x_j|\theta_j)$.

The farmer then optimally sets

$$\begin{aligned} a_j &= \begin{cases} 1 & \text{if } \bar{g}_j(\tilde{x}^*_j) > 0 \\ 0 & \text{otherwise} \end{cases} \\ &= a^*_j(\theta) \end{aligned}$$

and, when $a_j = 1$, sets $x_j = \tilde{x}^*_j = x^*_j(\theta)$.

■

B Extensions

This appendix section elaborates on the extensions and results we refer to in Section 5.

B.1 Further Setup and Concepts

Before presenting these extensions, we place a bit more structure on the model and introduce some useful concepts.

To fix ideas, we restrict $\pi_j \in \{\pi^L, \pi^H\}$ for each j, where $0 < \pi^L < \pi^H < 1$. We interpret

$\pi_j = \pi^H$ as a situation where the farmer believes that input j is likely to be relevant, or likely that his input choice along this dimension will impact his payoff in an *a priori* unknown manner, and $\pi_j = \pi^L$ as a situation where he believes it is likely that input j is *not* relevant, or likely that his input choice along this dimension will not impact his payoff.

We will say that a technology is *prior congruent* if the features that matter for using the technology line up well with the features that the farmer thinks should matter, and is *prior incongruent* otherwise. Formally:

Definition B.1. The technology is *prior congruent* if $\pi_j = \pi^H$ for all dimensions j that are e-relevant to using the technology. On the other hand, when $\pi_j = \pi^L$ for some dimension j that is e-relevant to using the technology then the technology is *prior incongruent*.

The role of this definition will be clearer after we state the following simplifying assumption. To limit the number of cases considered, it will be useful to focus on parameter values such that the farmer chooses to attend to an input in the first period when he places sufficiently large prior weight on the importance of that input. This is true, for example, whenever the cost of attending (e) is not too large.

Assumption B.1. Parameter values are such that the solution to the farmer's problem satisfies $a_{j1} = 1$ if $\pi_j = \pi^H$.

Combined with the definition of prior congruence, this assumption implies that a technology is prior congruent if and only if the farmer initially attends to all input dimensions that are worth attending to when optimally set.[2] Whether the technology is prior congruent predicts whether the farmer will learn to optimize the technology.

[2]Thus, a technology is prior congruent even if the farmer initially attends to some variables that are not worth attending to, so long as he attends to all variables that are worth attending to: Our notion of prior congruence ignores biases of commission. This is natural when the focus is primarily on understanding long-run behavior (as is ours), but other notions are relevant to understanding short-run behavior since it is costly for the farmer to attend to relationships that ultimately do not matter for profits.

B.2 Extension 1: Sequential Technologies

To explore how experience with related technologies may influence whether the farmer learns to optimize some target technology, consider an extension of the model where a farmer sequentially uses different technologies $\tau = 1, 2, \ldots$ for two periods each. His behavior over the two periods for a fixed technology is as described by our baseline model, and his prior over the importance of variable j given technology τ, $\pi_j^\tau \in \left\{\pi^L, \pi^H\right\}$, depends on his experience with previous technologies. Specifically, suppose that, when using a technology for the first time, the farmer has a prior belief that input dimension j is important if and only if he attended to that dimension the period before when he used the directly preceding technology:

$$\text{For } \tau > 1,\ \pi_j^\tau \quad = \quad \pi^H \text{ if and only if } a_{j2}^{\tau-1} = 1. \tag{4}$$

To complete the description of the sequential technology extension of the model, we need to specify initial conditions for the farmer's prior; that is, his prior for technology $\tau = 1$. It is intuitive that the farmer's prior is initially symmetric across input dimensions, reflecting the idea that the farmer must learn which dimensions are important. To keep the model interesting, we further assume that he starts off believing that every dimension is likely to be important:[3]

$$\pi_j^1 = \pi^H \text{for all } j. \tag{5}$$

To gain intuition for how previous experience can influence whether a given technology is prior congruent, we first highlight a feature of the baseline model:

Lemma B.1. *For all τ, π_j^τ, and $e > 0$, $a_{j2}^\tau = 1$ only if dimension j is e-relevant for technology τ.*

Proof. Suppose dimension j is *not* e-relevant for technology τ. There are two cases to consider. First, if $a_{j1}^\tau = 1$, then $a_{j2}^\tau = a_j^*(\theta^\tau) = 0$ by Lemmas A.1.1 and A.2. Conversely, if $a_{j1}^\tau = 0$, then

[3] Otherwise, experience with other technologies will not matter: Depending on the level of e, the farmer will either learn to optimize every technology (when e is sufficiently low that he initially attends to everything) or no technologies (when e is sufficiently large that he initially attends to nothing).

$a_{j2}^{\tau} = 0$ by Lemma A.1.2.

■

Lemma B.1 says that, for a fixed technology, the farmer will eventually not attend to any dimension that is not worth attending to. The intuition is straightforward: If the farmer initially attends to a dimension that is not worth attending to, he will learn to stop; if he initially does not, he will continue not to attend. From condition (4), an immediate corollary is that a technology $\tau > 1$ is prior congruent only if all input dimensions j that are e-relevant for technology τ were also e-relevant for technology $\tau - 1$, since the farmer will initially place low weight on the importance of any input dimension that he stopped attending to when using the last technology. We get a stronger result when we assume that e is sufficiently large that the prior influences what a farmer attends to:

Proposition B.1. *Suppose that the farmer sequentially uses technologies* $\tau = 1, 2, \ldots$ *and his prior satisfies conditions* (4) *and* (5)*. For sufficiently large e, technology* $\tau > 1$ *is prior congruent if and only if all input dimensions* j *that are e-relevant for that technology are also e-relevant for every technology* $1, 2 \ldots, \tau - 1$.

Proof. Consider a technology $\tau > 1$. For the "if" direction, suppose that all input dimensions that are e-relevant for τ are also e-relevant for technologies $1, 2, \ldots, \tau - 1$. The goal is to show that $\pi_j^{\tau} = \pi^H$ for any dimension j that is e-relevant for technology τ. Consider any one such dimension, k. Combining the assumption that $\pi_k^1 = \pi^H$ (Equation (5)) with Assumption B.1 implies that $a_{k1}^1 = 1$. Since k is e-relevant it further follows that $a_{k2}^1 = a_k^*(\theta_k^1) = 1$ by Lemmas A.1 and A.2, which implies that $\pi_k^2 = \pi^H$ by Equation (4). Iterating this argument yields the desired conclusion.

For the "only if" direction, suppose that for some technology $\tilde{\tau} < \tau$ there is some dimension j that is not e-relevant for $\tilde{\tau}$, but is e-relevant for τ. Since j is not e-relevant for $\tilde{\tau}$, $a_{j2}^{\tilde{\tau}} = 0$ by Lemma B.1, which implies that $\pi_j^{\tilde{\tau}+1} = \pi^L$ by Equation (4). For e sufficiently large, this then implies that $a_{j1}^{\tilde{\tau}+1} = 0$ from inspecting Equation (3) in the proof of Proposition 1, which in turn implies that $a_{j2}^{\tilde{\tau}+1} = 0$ by Lemma A.1.2, and finally that $\pi_j^{\tilde{\tau}+2} = \pi^L$ by Equation (4). Iterating this argument implies that $\pi_j^{\tau} = \pi^L$, which implies that technology τ is not prior congruent.

■

Proposition B.1 says that a technology is prior congruent if and only if input dimensions that are *e*-relevant to using the current technology were also *e*-relevant to using *every* preceding technology. The farmer starts off attending to everything, but stops attending to a dimension the first time he encounters a technology under which it is not worth attending to.

Proposition B.1 implies path dependence, where the same technology can be prior congruent or incongruent, depending on the technologies the farmer previously encountered. A more unique implication is that experience with related technologies may not be beneficial–and may even be harmful– if the relevant inputs differ from those of the current technology. Proposition B.1 implies the particularly stark result that, all else equal, the same technology is *less* likely to be prior congruent if it occurs later in the sequence $\tau = 1, 2, \ldots$ since this gives a farmer more opportunities to stop attending to a given input dimension.[4]

B.3 Extension 2: Providing Demonstrations

A common intervention for improving productivity is to provide demonstrations. What is the impact of demonstrations when farmers are inattentive?

To incorporate demonstrations in our model, suppose now that there is an additional period 0, where the farmer does not farm, but rather observes a demonstration by an individual with knowledge of the underlying production function. We will call this individual a *best-practice demonstrator*. The best practice demonstrator optimally farms the farmer's plot, i.e., he follows $(\sigma^*(\theta), \mathbf{a}^*(\theta))$, and it is common knowledge that the demonstrator knows how to optimize the technology; the farmer's prior over the demonstrator's strategy is thus derived from his prior over θ. We are sidestepping (potentially interesting) issues that could arise from the farmer being uncertain

[4]While this discussion highlights a potential cost of experience, the model points to other benefits. Experience can teach the farmer to ignore variables that truly are unimportant across many technologies, which reaps benefits the first period the farmer uses a new technology. Further, as emphasized above, experience is always helpful for a *fixed* technology: the farmer will attain a higher payoff the second period he uses a given technology as compared to the first.

about whether the strategy followed by the demonstrator would be optimal for him either because of uncertainty over whether the demonstrator optimizes or over whether his plot is different.

Proposition B.2. *Consider a farmer who learns from a best-practice demonstrator in period* 0 *and then decides how to farm in periods* 1 *and* 2*.*

1. *If there are no costs of attention ($e = 0$), then the farmer learns to optimize the technology from watching the demonstrator: $L_1 = L_2 = 0$.*

2. *If there are costs of attention ($e > 0$), then the farmer may not learn to optimize the technology from watching the demonstrator: For sufficiently large e, $L_1 = L_2 = 0$ only if the technology is prior congruent.*

Proof. To establish the first part of the proposition, consider the case where there are no costs of attention. In this case, the farmer learns $(\sigma^*(\theta), \mathbf{a}^*(\theta))$ by watching the demonstrator. Consequently, he farms according to $(\sigma^*(\theta), \mathbf{a}^*(\theta))$ in periods 1 and 2, which implies that $L_1 = L_2 = 0$.

For the second part, assume that e is sufficiently large that the farmer will choose not to attend to the demonstrator's choice along dimension j, $(\sigma_j^*(\theta_j), a_j^*(\theta_j))$, whenever $\pi_j = \pi^L$.[5] Suppose further that the technology is prior incongruent, so there is at least one dimension, j, that the farmer does not attend to when watching the demonstrator, but is e-relevant. Because the farmer does not attend to what the demonstrator does along dimension j, he does not learn $(\sigma_j^*(\theta_j), a_j^*(\theta_j))$ by watching the demonstration; instead, because his prior is symmetric across $\tilde{\theta}_j(x_j')$, $x_j' \in X_j$, he will remain indifferent between input levels after watching the demonstrator: There exists a $K \in \mathbb{R}$ such that $E[\tilde{\theta}_j(x_j')|\hat{h}_1] = K$ for all $x_j' \in X_j$, where $\hat{h}_1$ denotes the farmer's recalled history at the start of period 1. Consequently, in the first period the farmer either will not attend to input j, in which case $L_1 > 0$, or will attend and set $\sigma_{j1}(x_j') > 0$ for some $x_j' \neq x_j^*(\theta_j)$, in which case $L_1 > 0$ as well.[6]

■

[5] Note that we can find such a range of e consistent with the farmer choosing to attend when $\pi_j = \pi^H$ since the net benefit to attending is lower when $\pi_j = \pi^L$ (as opposed to $\pi_j = \pi^H$), is strictly decreasing in e, and drops below 0 when e is sufficiently large (fixing π_j).

[6] To see why the farmer follows a strategy in which $\sigma_{j1}(x_j') > 0$ for some $x_j' \neq x_j^*(\theta_j)$ if he attends, suppose the farmer instead follows a strategy in which he sets $a_{j1} = 1$ and $\sigma_{j1}(x_j^*(\theta_j)) = 1$. Given his first period beliefs, his expected payoff from following this strategy equals $E[v_{-j1} + v_{-j2} + K - e + v_{j2}|\hat{h}_1] < E[v_{-j1} + v_{-j2} + K - e +$

Proposition B.2 indicates that demonstrations do not ensure that farmers learn to use the production technology when they face costs of paying attention and the technology is prior incongruent. In such cases, farmers will be unable to faithfully replicate practices they saw in the demonstration.[7]

The model may also help understand why demonstration trials can have little long-run impact on technology adoption, as has been observed in some empirical contexts (Kilby 1962, Leibenstein 1966, Duflo et al. 2008b). To briefly incorporate the technology adoption decision into our model, suppose that in periods $t = 1,2$, the farmer faces the additional decision of whether to farm at all, where his outside option is to earn $\bar{v}$ if he chooses not to farm. Assume that farming is profitable when inputs are optimally set, $v^*(\theta) > \bar{v}$, so a farmer who knows best practices will choose to farm. Then, in the standard case where $e = 0$, farmers will necessarily choose to farm in periods 1 and 2 after observing the demonstration: In this case, the demonstration teaches the farmers that farming is profitable relative to the outside option when inputs are optimally set, and further how to optimally set inputs. Matters are more complicated when farmers face costs of paying attention.

To illustrate, consider a simple example where there are two input dimensions and the production technology is such that it is profitable when both x_1 and x_2 are optimally set, but not when the farmer does not attend to one of the inputs: $v^*(\theta) > \bar{v}$, but $v(\sigma, \mathbf{a}; \theta) < \bar{v}$ whenever $a_j = 0$ for some $j = 1,2$. In period 0, the best practice demonstrator will then set $x_1 = x_1^*(\theta)$ and $x_2 = x_2^*(\theta)$ on each plot of land, which will on average yield $f(\mathbf{x}^*(\theta)|\theta) = v^*(\theta) + 2 \cdot e$. If the technology is prior congruent, which in this context means $(\pi_1, \pi_2) = (\pi^H, \pi^H)$ since each input dimension is e-relevant, then the farmer will learn everything necessary to make optimal decisions in periods 1 and 2 from the demonstration (so long as parameter values are such that the farmer attends to the

$v_j^*(\tilde{\theta}_j)|\hat{h}_1]$ (since $E[v_{j2}|\hat{h}_1] < E[v_j^*(\tilde{\theta}_j)|\hat{h}_1]$ if the farmer follows such a strategy). But he could instead get the higher expected payoff by following a strategy in which $a_{j1} = 1$ and $\sigma_{j1}(x_j') > 0$ for all $x_j' \in X_j$ since this alternative strategy guarantees that he would learn to optimize dimension j since he would then experiment with each input in X_j infinitely often.

[7] In principle, the best-practice demonstrator can explain his strategy to the farmer either before or after the demonstration. However, the demonstrator makes many input choices and communicating these choices to the farmer is costly. He may only want to communicate choices along dimensions he believes the farmer is predisposed not to attend to and are relevant for the task at hand. The effectiveness of communication is then increasing in the degree to which the demonstrator has knowledge of the farmer's mental model; i.e., of what he does and does not attend to.

demonstrator's input choices): he will learn $\mathbf{x}^*(\theta)$ and $v^*(\theta)$ with certainty. Consequently, he will choose to farm in periods 1 and 2 and will optimally set inputs 1 and 2.

On the other hand, suppose the technology is prior incongruent since the farmer initially believes it is unlikely that the second dimension is relevant, $(\pi_1, \pi_2) = (\pi^H, \pi^L)$, and the farmer then only attends to the demonstrator's input choice along the first dimension. In this case, the demonstration will not teach the farmer how to optimally farm: while the farmer learns how to optimally set the first input and the expected yield given optimal input choices, $f(\mathbf{x}^*(\theta)|\theta)$, he will remain uncertain about how to set the second input since he did not attend to the demonstrator's choice along that dimension. However, while uncertain, if the farmer initially attached low probability to the second dimension being relevant, $\pi^L \approx 0$, then this uncertainty will be minimal: he is (mistakenly) very confident that it is optimal not to attend to that dimension. He consequently chooses to farm in the second period, expecting to earn a payoff much higher than his outside option with probability close to one, but is disappointed with certainty: he believes that his payoff will equal $f(\mathbf{x}^*(\theta)|\theta) - e = v^*(\theta) + e > \bar{v}$ with high probability if he sets $x_1 = x_1^*(\theta)$ and does not attend to the second input, but he in fact earns a payoff less than $\bar{v}$ since $v(\sigma, \mathbf{a}|\theta) < \bar{v}$ whenever $a_2 = 0$. Since his expected payoff from farming is now at most $\bar{v}$ following this disappointment, the farmer will choose to stop farming in the next, terminal, period.

This example indicates why demonstration trials can have a small impact on adoption decisions: even if people recognize from the demonstration that technologies are profitable when used optimally, they may not come away with knowledge on how to use the technology optimally if it is prior incongruent. Because of this, farmers may not consistently use the technology following the demonstration even though they "saw" how to use it optimally. In fact, farmers may initially try the technology but then give it up when they realize they have not learned to use it profitably.[8]

[8] If the technology is prior incongruent and the farmer does not attend to input dimensions he thinks are unlikely to be relevant both when watching the demonstration in period 0 and when trying the technology for the first time on his own in period 1, then he will not learn to optimize the technology ($L_2 > 0$). However, by comparing his average yield with the average yield of the demonstrator, he will learn that he is not optimizing the technology: He will know he does not know, but will not know what he does not know. In a richer model with more than two periods, this can induce him to experiment, though this desire may be limited if there are many potential inputs that he could attend to. It will be further limited if the demonstration takes place on another plot and the farmer places some weight on that plot differing from his own.

Summarizing:

Proposition B.3. *Consider a farmer who learns from a best-practice demonstrator in period* 0 *and then decides whether and how to farm in periods* 1 *and* 2*, where the technology is profitable when used optimally, but not profitable if the farmer does not attend to a dimension that is e-relevant:* $v^*(\theta) > \bar{v}$*, but* $v(\sigma, \mathbf{a}|\theta) < \bar{v}$ *whenever* $a_j = 0$ *for some dimension* j *that is e-relevant. Suppose the farmer chooses to farm in period* 1*, but does not attend to input dimensions satisfying* $\pi_j = \pi^L$ *both when watching the demonstration and when he farms in period* 1*. Then, if the technology is prior incongruent:*

1. *The farmer will be disappointed with what he earns from farming: he expects to earn over* $2 \cdot \bar{v}$ *combined across periods* 1 *and* 2*, but instead earns less than* $2 \cdot \bar{v}$*.*

2. *The farmer will give up farming: the farmer chooses not to farm in period* 2*.*

Proof. If the farmer decides to farm in period 1, this means that he expects to earn at least $2\bar{v}$ in total across periods 1 and 2, since can earn $2\bar{v}$ by not farming in either period. While such a farmer expects to earn at least $2\bar{v}$, he earns less than $\bar{v}$ in the first period whenever the technology is prior incongruent and he does not attend to dimensions satisfying $\pi_j = \pi^L$ in periods 0 and 1, given the assumption that $v(\sigma, \mathbf{a}; \theta) < \bar{v}$ whenever $a_j = 0$ for some e-relevant dimension j. This means that this farmer in fact earns less than $2\bar{v}$ across periods 1 and 2 so long as he gives up farming in the second period.

To complete the proof we must then verify part (2) of the proposition; that the farmer indeed finds it optimal not to farm in period 2. Since, by assumption, the farmer does not attend to input dimensions satisfying $\pi_j = \pi^L$ both when watching the demonstration and when farming himself in period 1, he does not learn to optimize those dimensions and remains indifferent between different input choices along those dimensions. As a result, he cannot expect to earn more than $v_1 = v(\sigma_1, \mathbf{a}_1; \theta) < \bar{v}$ by farming in the second (terminal) period, and stops.

■

An implication of the analysis is that finding that a demonstration has little impact on long-run beliefs or behavior does *not* necessarily imply that farmers have little left to learn or even that the demonstration provided them with little informative data. Rather, when the technology is prior incongruent, it could reflect that the demonstration did not sufficiently alter farmers' initially mistaken beliefs about which input dimensions matter.

Online Appendix Table I: Randomization Check

	Mean, by Treatment Group			Differences		
	Control	Sort	Weight	Col 1 - Col 2	Col 1 - Col 3	Col 2 - Col 3
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Monthly Per Capita Income)	12.43	12.42	12.67	0.233	-0.016	-0.249
	(1.37)	(0.72)	(0.76)	(0.164)	(0.156)	(0.138)*
HH Head is Literate	0.79	0.84	0.88	0.093	0.050	-0.043
	(0.41)	(0.37)	(0.32)	(0.059)	(0.060)	(0.064)
Number of Assets	8.36	7.95	8.28	-0.079	-0.409	-0.330
	(3.15)	(3.19)	(3.06)	(0.511)	(0.494)	(0.580)
Age of HH Head	43.23	43.95	43.54	0.304	0.718	0.414
	(12.22)	(12.14)	(11.33)	(1.937)	(1.905)	(2.192)
Years Farming	18.00	18.94	18.27	0.275	0.937	0.662
	(6.96)	(7.06)	(7.06)	(1.183)	(1.103)	(1.330)
Parents Farmed Seaweed	0.47	0.55	0.53	0.063	0.081	0.019
	(0.50)	(0.50)	(0.50)	(0.083)	(0.078)	(0.093)
Loans from Someone Sells to	0.31	0.33	0.22	-0.085	0.026	0.111
	(0.46)	(0.48)	(0.42)	(0.088)	(0.085)	(0.097)
Farms Cottonii	0.34	0.33	0.47	0.135	-0.008	-0.144
	(0.47)	(0.47)	(0.50)	(0.082)*	(0.074)	(0.091)
Pod Size at Baseline	109.01	102.83	112.78	3.777	-6.174	-9.951
	(28.79)	(22.57)	(38.00)	(5.850)	(3.884)	(5.930)*
Number of Days of Previous Cycle	36.61	36.00	37.08	0.463	-0.612	-1.075
	(8.24)	(7.31)	(6.54)	(1.178)	(1.190)	(1.281)

Notes: This table provides a check on the randomization. Columns 1 - 3 provide the mean and standard deviation of each baseline characteristic for the control group, sort group, and weight group, respectively. Columns 4 - 6 give the difference in means (and standard errors) between the noted experimental groups. Statistical significance is denoted by: *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Online Appendix Table II: Reasons Why Farmer May Want to Try a New Method

	Mean (1)	Standard Deviation (2)	N (3)
Would Not Want to Make any Changes	0.04	0.18	482
Own Initiative	0.02	0.13	482
Pest or Price	0.02	0.13	482
Advice from Friend	0.10	0.30	482
NGO or Government Recommendation	0.11	0.31	482
Sees Results on Other Farmers' Plots	0.39	0.49	482

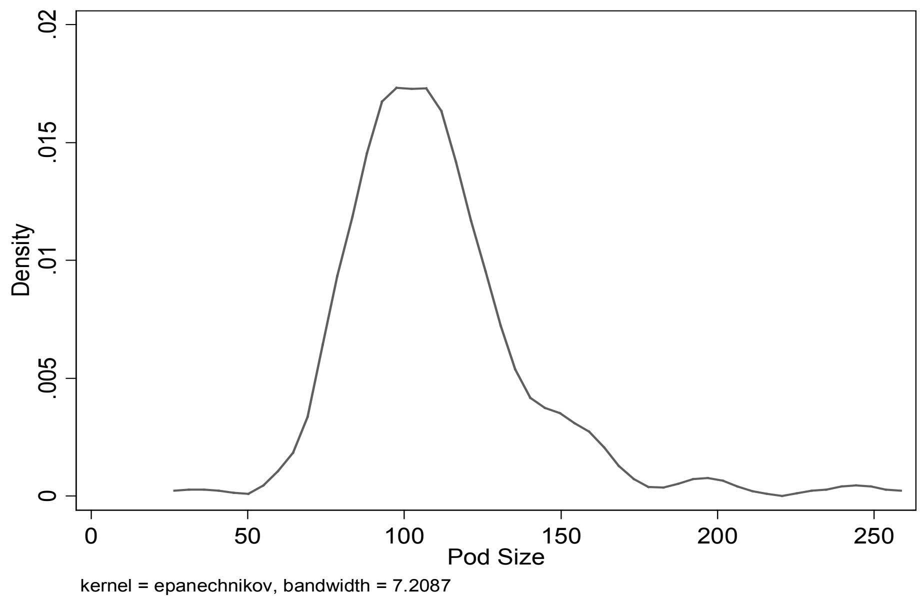
Notes: This table provides the baseline survey responses on the reasons why a farmer may try a new method.

Online Appendix Table III: Effect of Treatment, by Recommendation

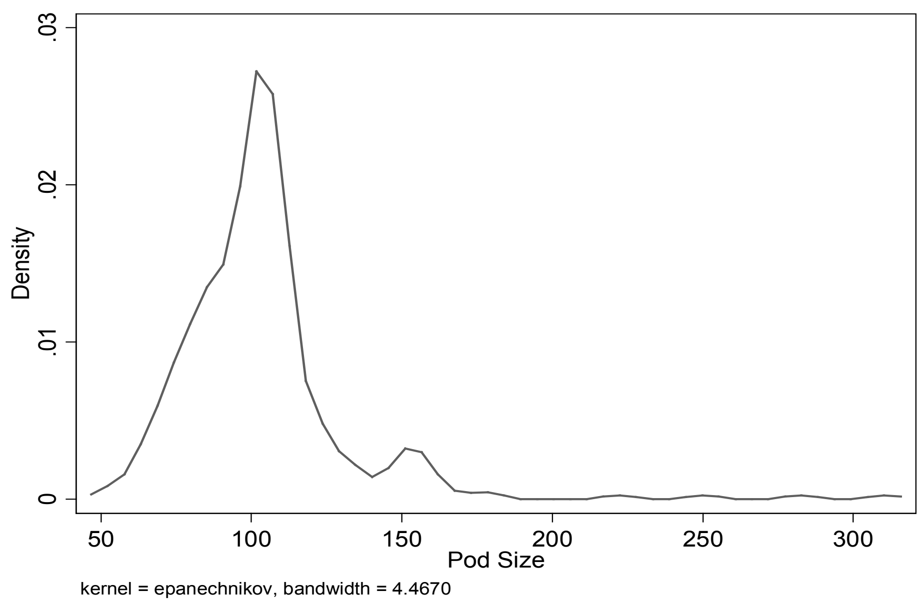
	Pod Size (Grams)	
	(1)	(2)
Increase	-21.091	-50.976
	(4.807)***	(1.851)***
Increase * Trial Participation * After Trial	27.264	27.060
	(6.223)***	(7.627)***
Increase * Trial Participation * After Summary Data	19.502	19.242
	(5.808)***	(7.149)***
Hamlet Fixed Effects	X	
Farmer Fixed Effects		X
Observations	675	675

Notes: This table provides the coefficient estimates of the effect of treatment on farming methods after the trial (follow-up 1) and after observing the summary data (follow-up 2), conditional on baseline farming methods and disaggregated by whether the farmers were recommended to increase or decrease their pod size when observing the summary data. The trial participation dummy indicates that the farmer belongs in either the sort or weight treatment group. Increase is an indicator variable for being told to increase size. All regressions are estimated using OLS and include the main effects as well as the double interactions, and standard errors are clustered at the farmer level. Statistical significance is denoted by: *** p<0.01, ** p<0.05, * p<0.10.

Online Appendix Figure IA: Baseline Pod Sizes for Cottonii Growers

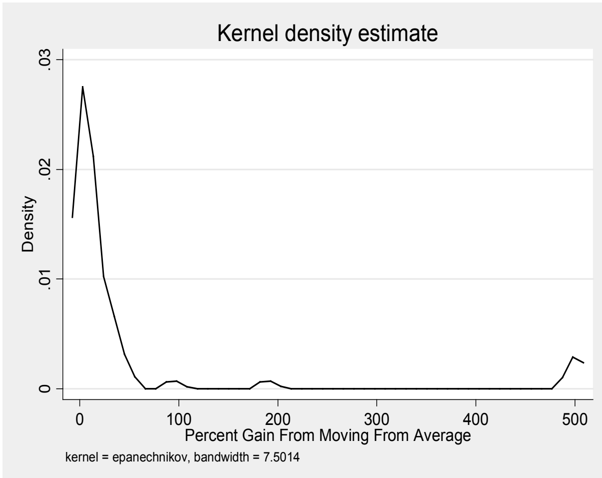


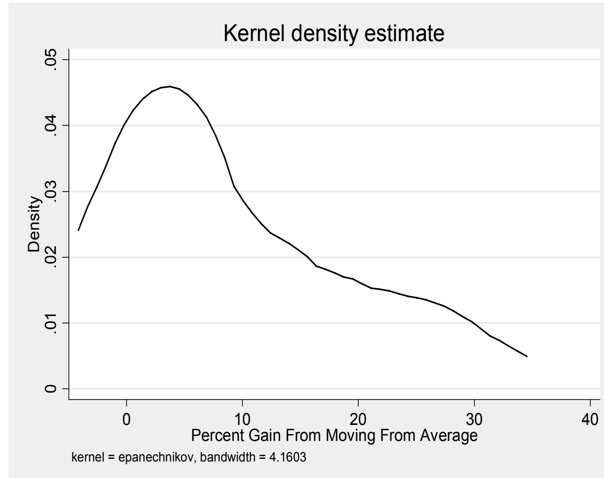
Online Appendix Figure IB: Baseline Pod Sizes for Spinosum Growers



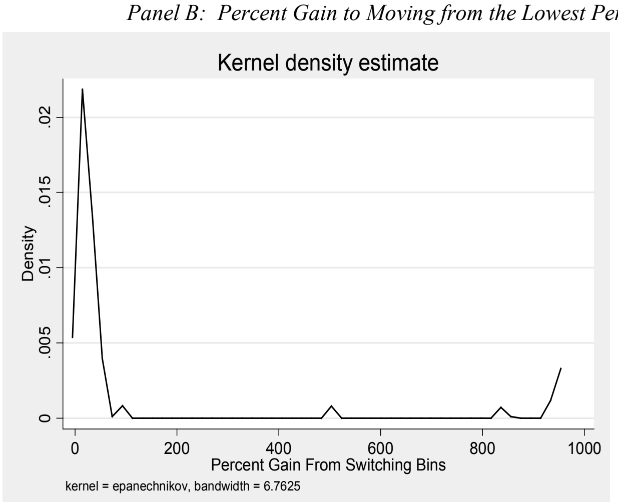
Online Appendix Figure II: Distribution of the Estimated Percent Income Gain from Switching to Trial Recommendations

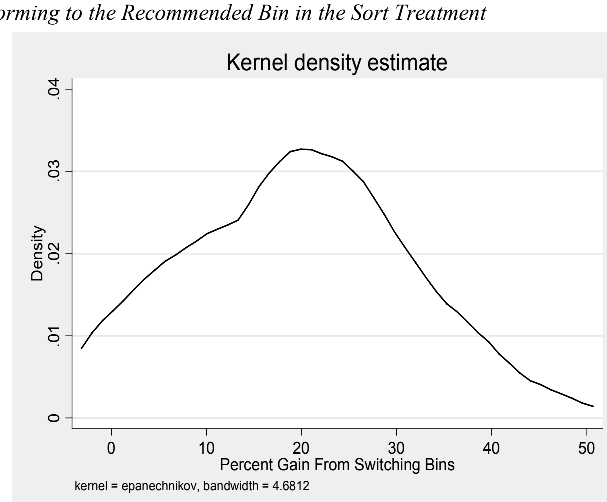
Panel A: Percent Gain to Moving from the Baseline Average to the Recommended Bin in the Sort Treatment



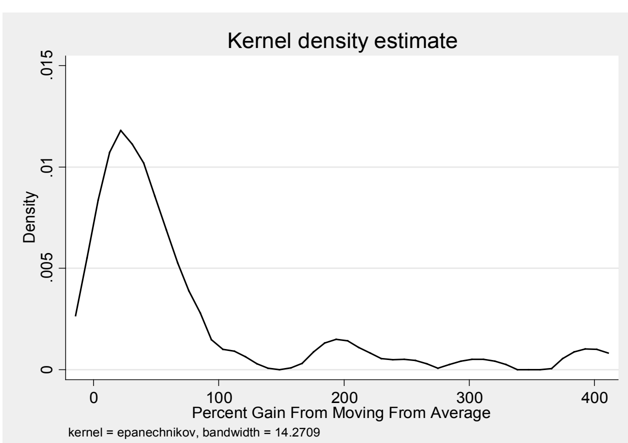


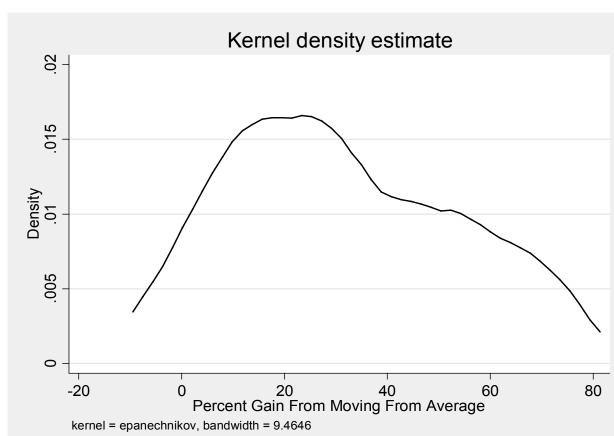
Panel B: Percent Gain to Moving from the Lowest Performing to the Recommended Bin in the Sort Treatment





Panel C: Percent Gain to Moving from the Baseline Average to the Recommended Bin in the Weight Treatment





Notes: The first column of figures provides the full set of returns, while the second column focuses on the bottom 80 percent of the sample.

Online Appendix Figure III: Distribution of Recommended Pod Sizes in the Weight Treatment