Signals, Similarity and Seeds: Social Learning in the Presence of Imperfect Information and Heterogeneity

Emilia Tjernström

University of California, Davis

November 3, 2014

1. Introduction
   - Motivation
   - Context
2. Research design
   - Data sources
     - RCT
     - Network info
   - Variable definitions
   - Econometrics
3. Results
   - Data
   - Social network results
   - Heterogeneity

Learning & technology adoption

- Greater use of improved technologies could raise productivity and welfare in developing countries
- Returns are typically unknown and stochastic
- Understanding how individuals learn & decide what technologies to use crucial to boosting prosperity

Learning & technology adoption in agriculture

- Agricultural technologies provide a favorable and important context for the study of learning
- Farmers make production choices in an environment characterized by imperfections, where learning is difficult
  - financial imperfections: credit constraints and imperfect insurance markets
  - incomplete information about the availability and profitability of new technologies
  - complex and heterogeneous information environment
- Social learning plays a role in diffusion and adoption (Foster & Rosenzweig, 1995; Bandiera & Rasul, 2006; Conley & Udry, 2010; Magnan et al., 2013; Cai et al., 2014; Carter et al., 2014; Adhvaryu, 2014)

Agricultural productivity in SSA: low and stagnant

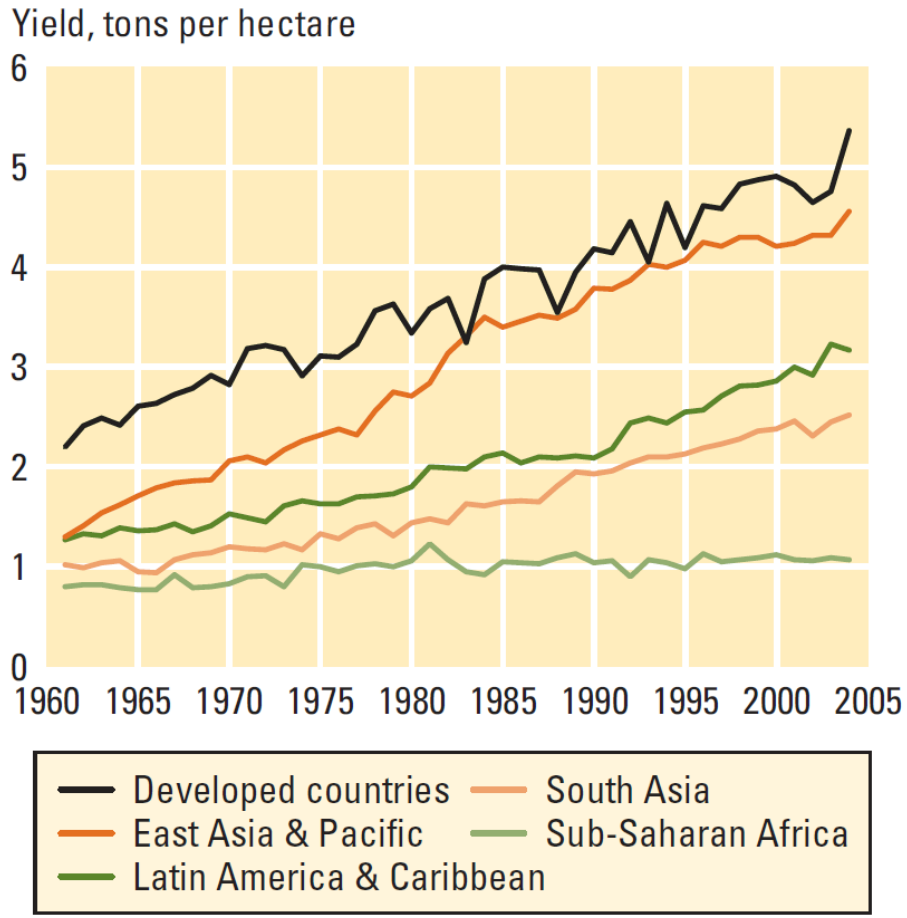


Figure : Cereal yields in SSA & other regions

Hybrids in Kenya

- Hybrid use is higher than many other SSA countries (40-70%)
- Stagnating maize production partly due to slow replacement of old hybrids
  - 2/3 of farmers grow a hybrid developed in 1986, suited for the Kenyan highlands (Tegemeo, 2010)
  - relevant decision is *type* of hybrid & this choice is complex
    - many seeds to choose from
    - soil quality varies widely

Farmers face substantial and growing complexity

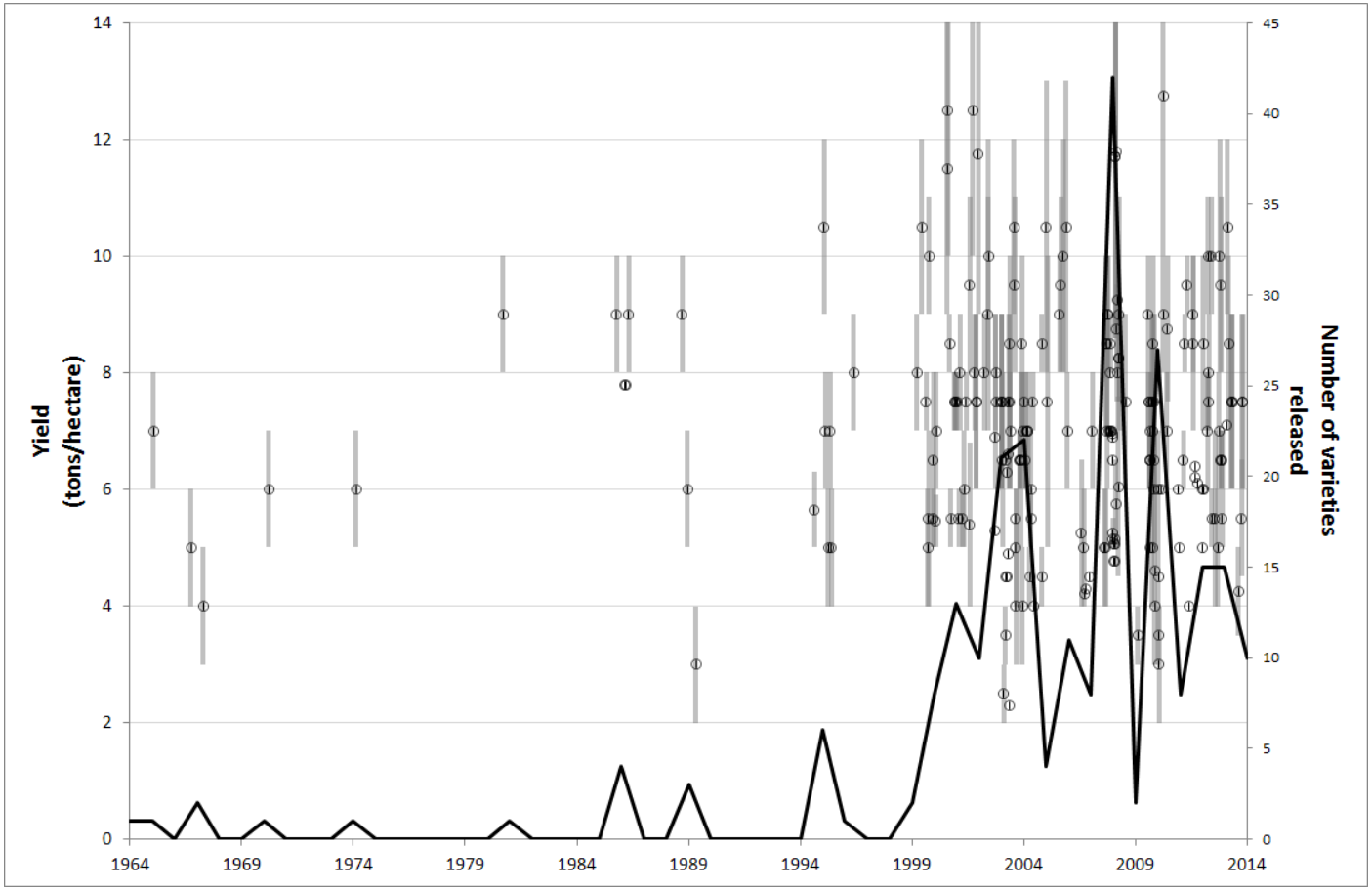


Figure : Number of maize varieties released in Kenya, 1964 - 2014 and their reported yield capacity

Region exhibits significant heterogeneity in soil quality

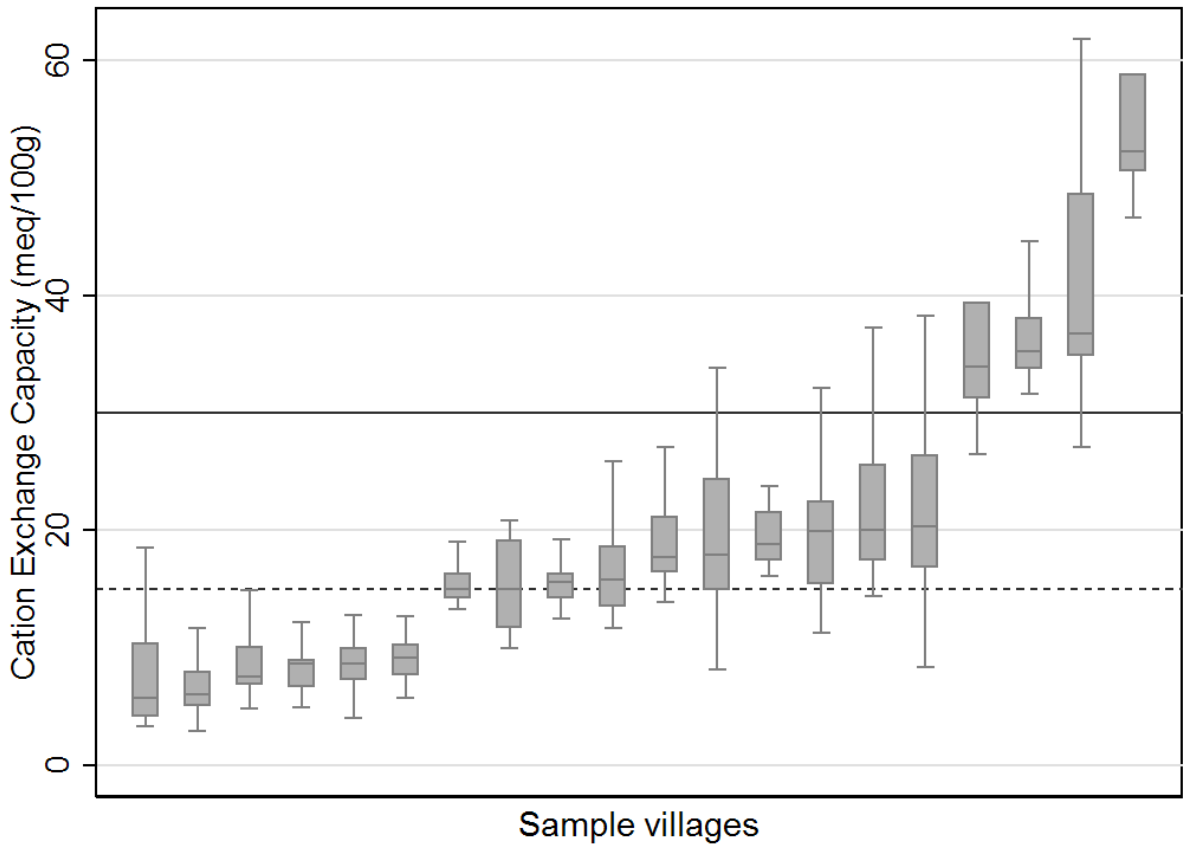


Figure : Box plot of Cation Exchange Capacity across sample villages

What I do & summary of results

- Experimental variation in information available to farmers about new tech
  - construct a measure of the signal in individuals' networks
  - examine how social networks affect familiarity, WTP and adoption of new tech
- Networks matter: they affect
  - familiarity
  - WTP
  - adoption
- Unobserved heterogeneity makes individuals less likely to respond to their peers' experiences

What I do & summary of results

- Experimental variation in information available to farmers about new tech
  - construct a measure of the signal in individuals' networks
  - examine how social networks affect familiarity, WTP and adoption of new tech
- Networks matter: they affect
  - familiarity
  - WTP
  - adoption
- Unobserved heterogeneity makes individuals less likely to respond to their peers' experiences

Impact evaluation

- Large-scale RCT: *"Evaluating the socio-economic impacts of Western Seed's hybrid maize program"*
- Western Seed Company (WSC)
  - high-yielding maize hybrids
  - adapted to mid- & low- altitude areas
- Until recently, limited by capacity-constraints

Impact evaluation

- Study villages are in WSC expansion areas
  - no/little information or marketing
  - no/little access to the seeds
  - may have experience with other hybrids
- Cluster-randomized roll-out
  - information about WSC
  - 250g samples of the seeds
    - could plant small experimental plot
    - $\frac{1}{30}^{th}$ of average farmers land

Impact evaluation

- Villages divided into treatment and control clusters
- Sampled farmers in treatment villages received info & samples
- *Main goal:* induce different adoption levels between treatment and control villages
- *Experiment-within-experiment:* variation within treatment villages in the level of experience with the new technology
    - orthogonal to farmer attributes & social network characteristics

Farmer types

Farmer type	**Village**	**Info + sample**	**Baseline**	**Soil sample**	**Network**
Directly treated	Treatment	Yes	Yes	Yes	Yes
Indirectly treated	Treatment				Yes
Control	Control		Yes	Yes	

Farmer types

Farmer type	**Village**	**Info + sample**	**Baseline**	**Soil sample**	**Network**
Directly treated	Treatment	Yes	Yes	Yes	Yes
Indirectly treated	Treatment				Yes
Control	Control		Yes	Yes	

Impact evaluation - timeline

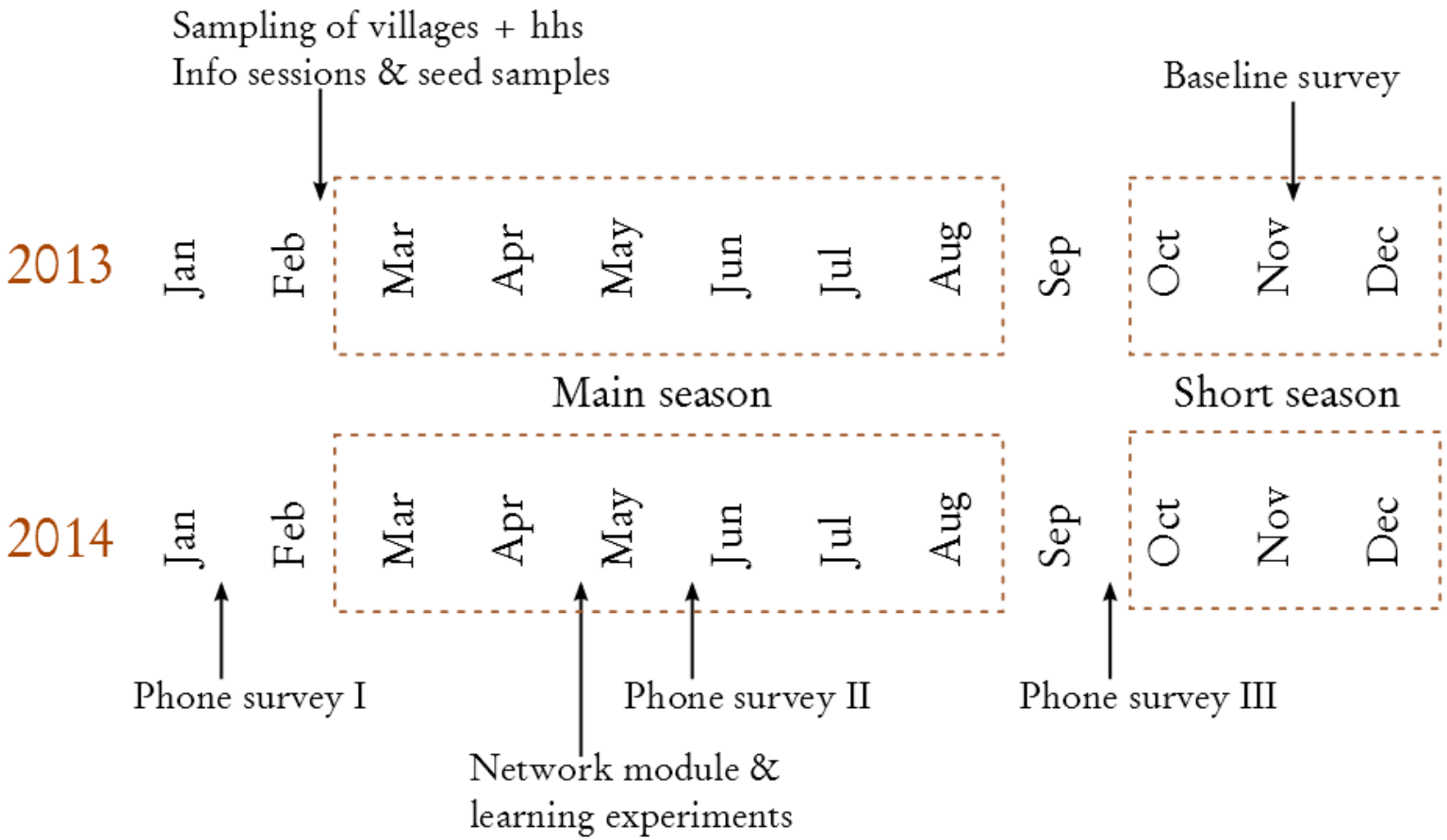


Figure : RCT timeline

Impact evaluation - timeline

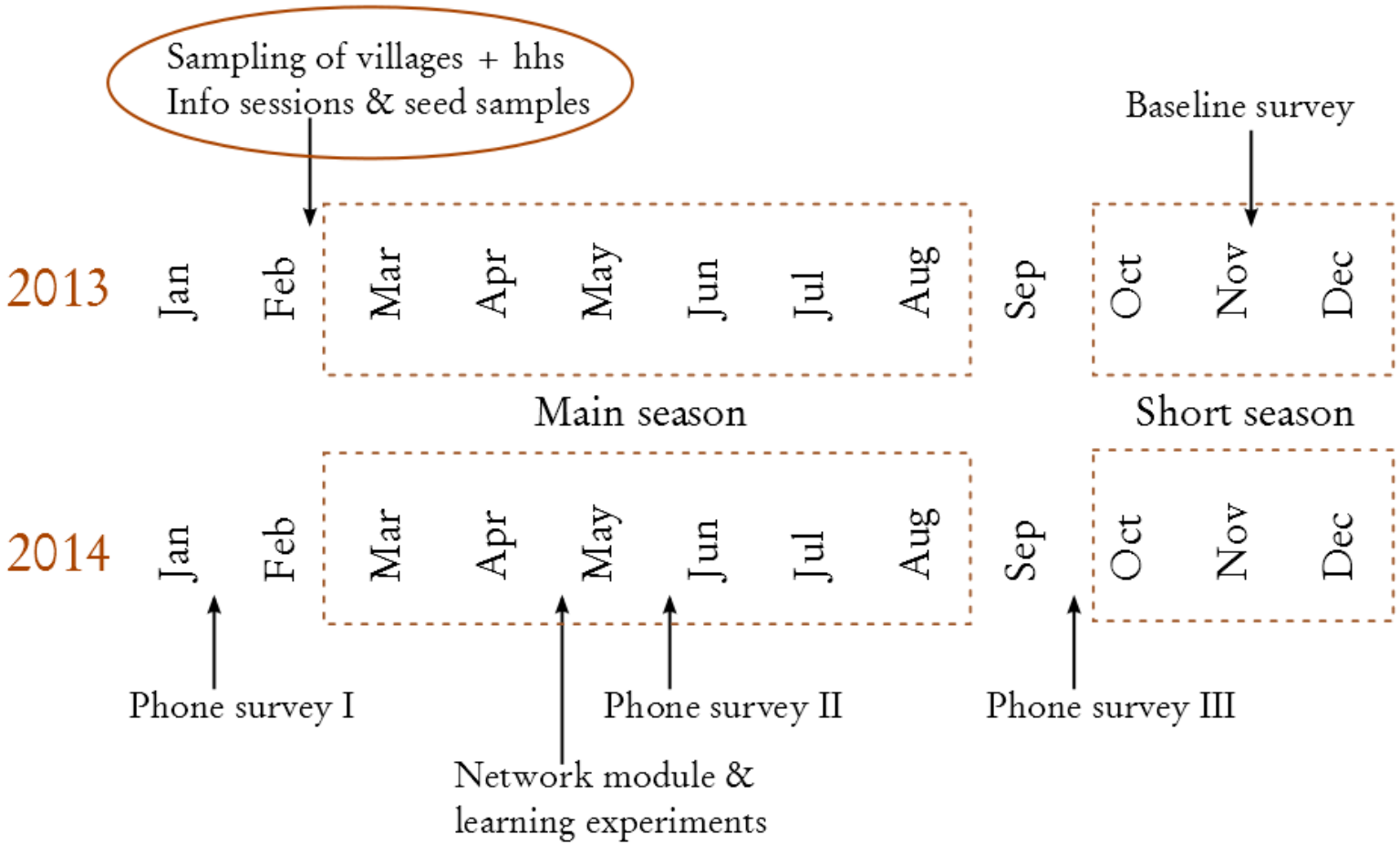


Figure : RCT timeline

Impact evaluation - timeline

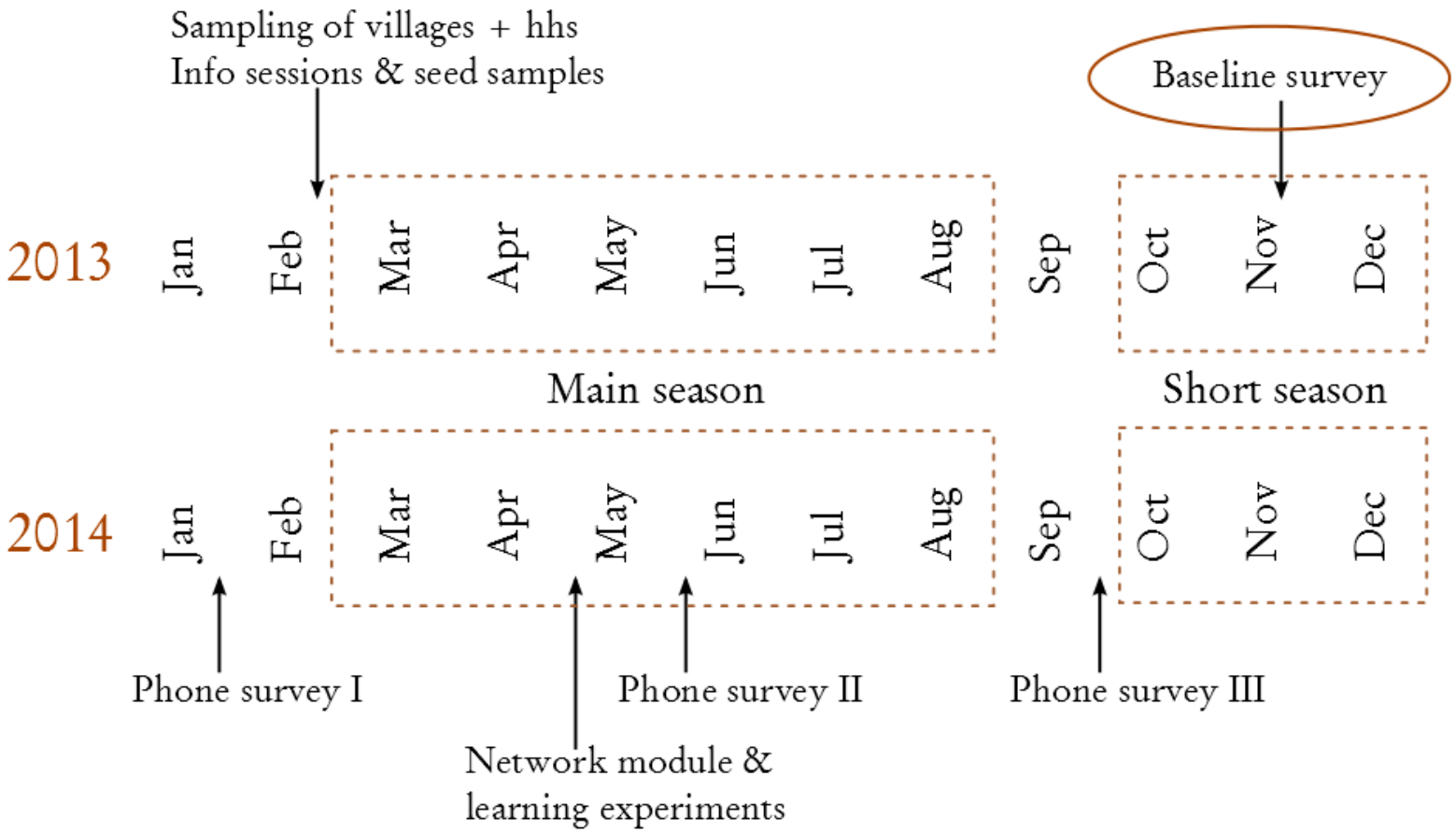


Figure : RCT timeline

Impact evaluation - timeline

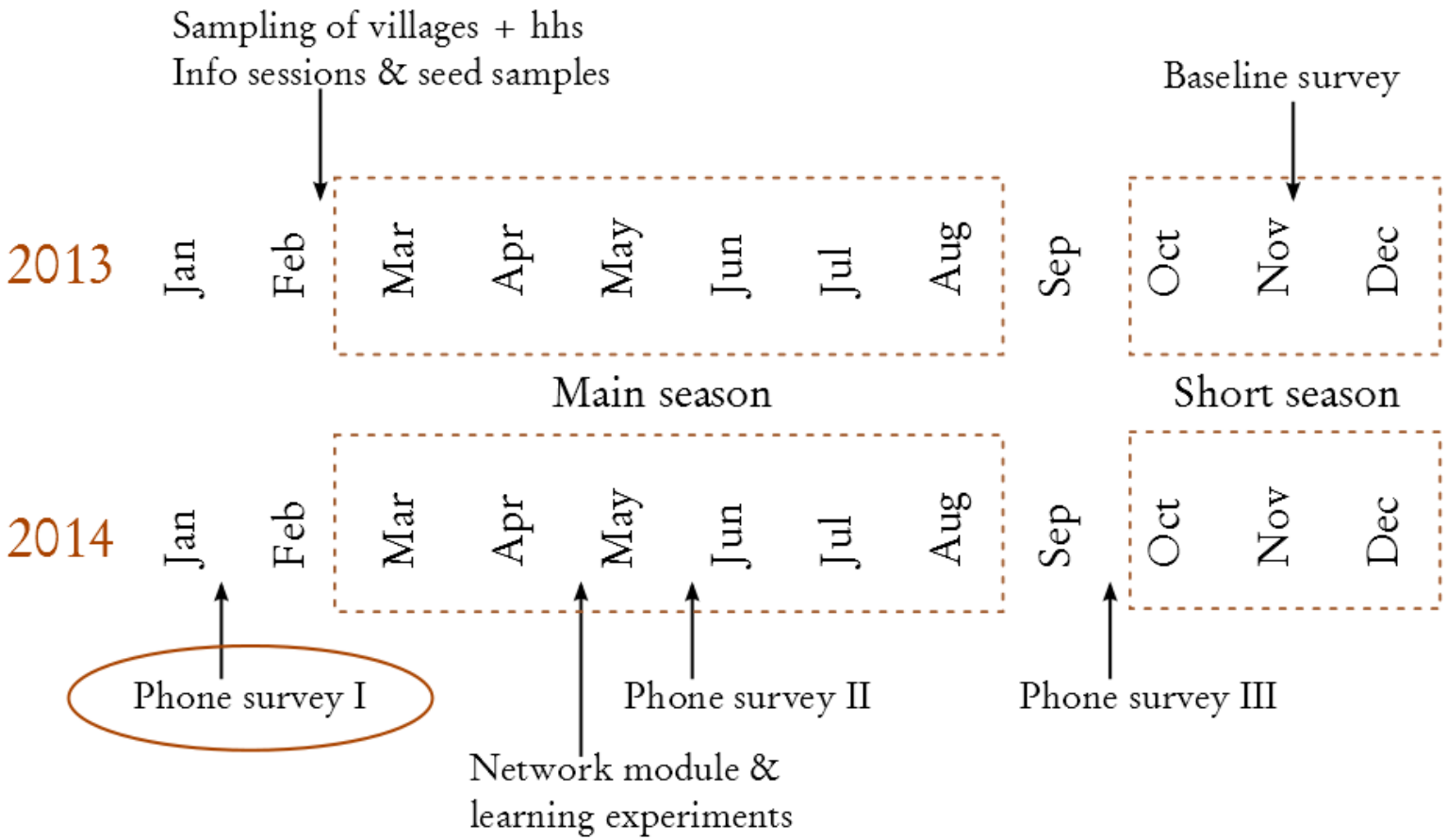


Figure : RCT timeline

Impact evaluation - timeline

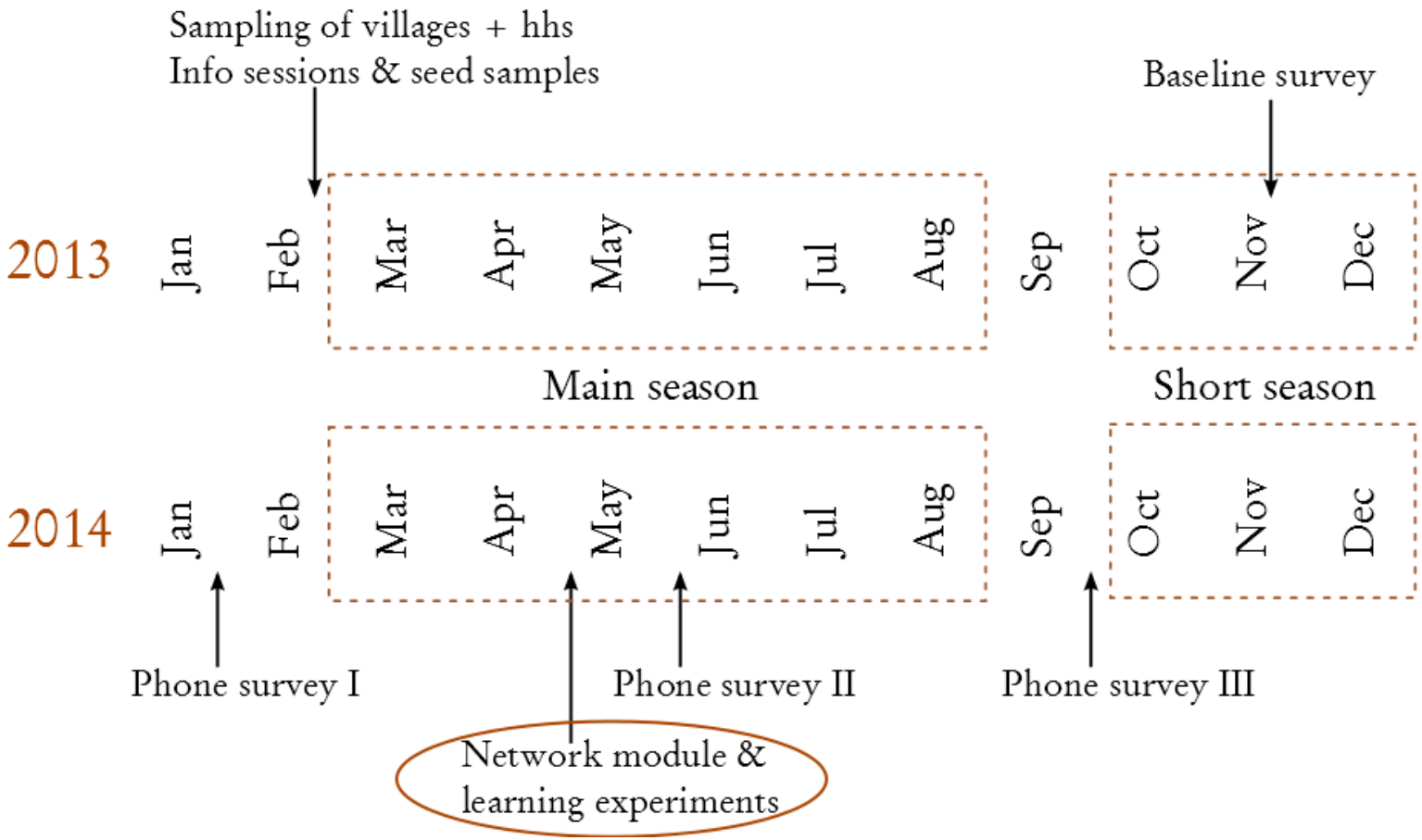


Figure : RCT timeline

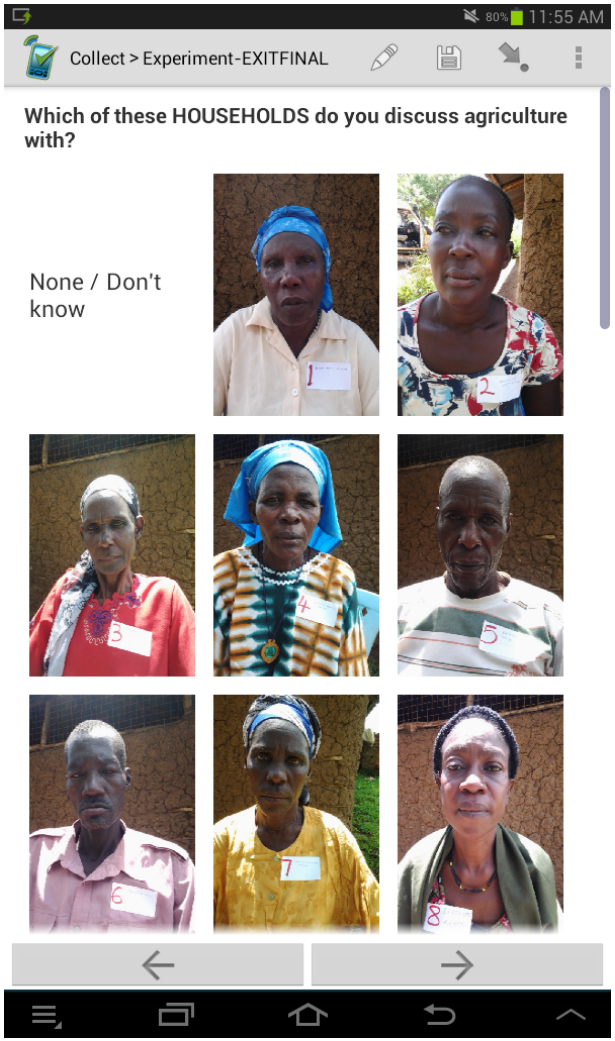
Network information

- Additional survey in 20 treatment villages
  - all directly treated hhs
  - random sample of indirectly treated
- 600 farmers invited; 575 (96%) showed up & participated
- Indirectly treated answered additional survey since not in baseline

Different network types

- Information neighbors
- Talk to (about anything, about ag + at different frequencies)
- Economic (microfinance, women's group, farming group)
- Geographic (walk/bike by, live closest to)
- Information (advice, what seeds they planted/prefer, most similar to you, recommend WSC hybrids)

Tablet network module



Tablet network module

Tablet network module

Tablet network module

Tablet network module

Tablet network module

Tablet network module

Tablet network module

Network definition

- For present analysis, individual *j* is in person *i*'s social network if person *i* listed them in *any* of the network questions
- Many options for defining information networks
  - *reciprocal*: *i* mentions *j* and *j* mentions *i*
  - *corrected*: remove those who spoke about maize for the first time after treatment

Network definition

- For present analysis, individual *j* is in person *i*'s social network if person *i* listed them in *any* of the network questions
- Many options for defining information networks
  - *reciprocal*: *i* mentions *j* and *j* mentions *i*
  - *corrected*: remove those who spoke about maize for the first time after treatment

Information signal

- Several recent papers use experimental variation in networks (Carter et al., 2014; Cai et al., 2014; Magnan et al., 2013; Oster & Thornton, 2012)
- Unlike earlier observational studies that used innovative measures of information, the experimental studies rely on number of treated in network
  - gets around reflection problem (Manski, 1993)
  - implicitly assumes 'social influence' model, rather than social *learning*

Information signal

- Phone survey with treated - elicit their experience with the technology

1. Actual experience (y_i): "How much did you harvest from the sample pack seeds?"
2. Subjective counterfactual ($\tilde{y}_i$): "How much would you have harvested (same weather, input use, etc) if you had planted the seeds you normally grow instead of WSC hybrids?"

- Denote the perceived experimental gains by Δ_i

$$\Delta_i = \frac{y_i - \tilde{y}_i}{\tilde{y}_i}$$

Information signal

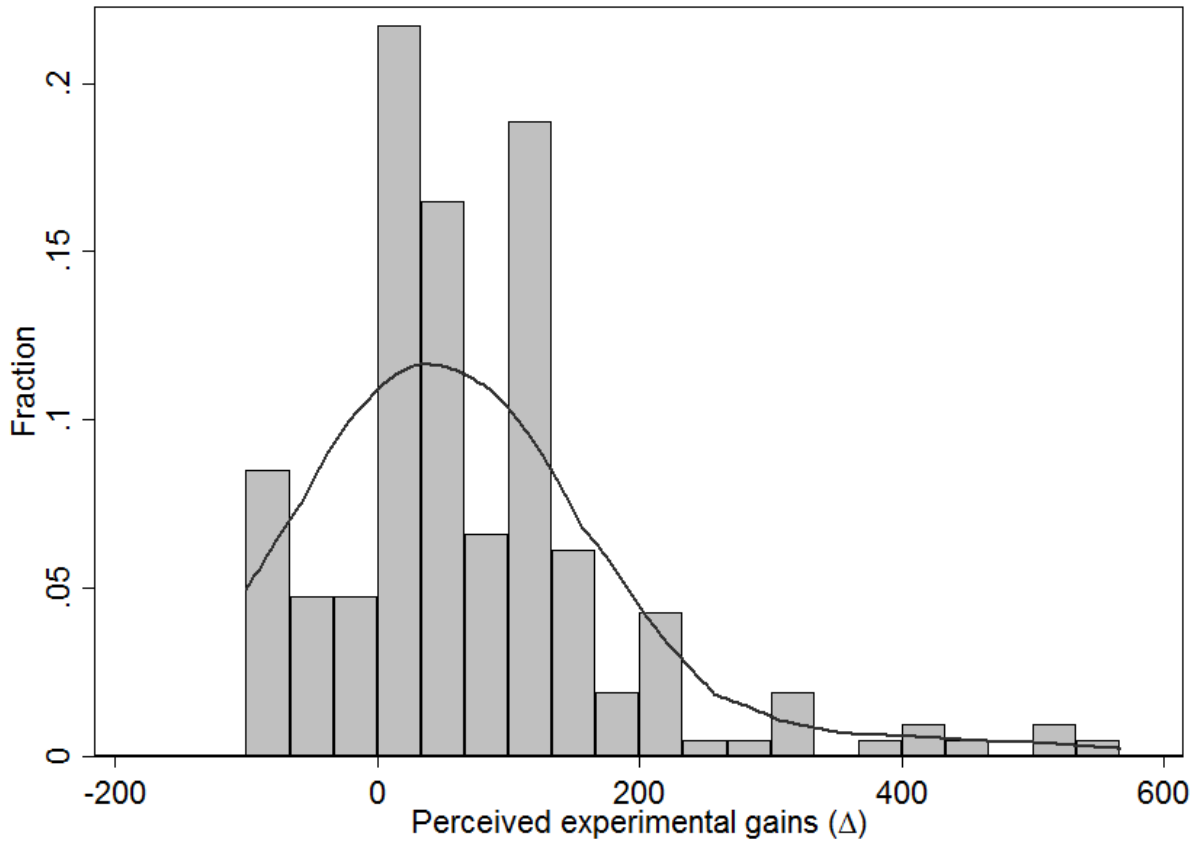


Figure : Distribution of treated farmers' evaluation of the performance of the hybrid seed samples

Information signal

- The experiences of the farmers in person i's network combine to form a distribution of signals from which she can learn
  - compute the mean and variance of the signals in a respondent's network

$$\mu_i = \sum_{j \in N_i} \frac{\Delta_j}{N_i}$$

$$\sigma_i = \sum_{j \in N_i} \frac{(\Delta_j - \mu_i)^2}{N_i}$$

- A higher μ_i should increase likelihood that farmer i adopts
- A higher σ_i, i.e. a noisier signal, should decrease farmer i's response to the signal

Information signal

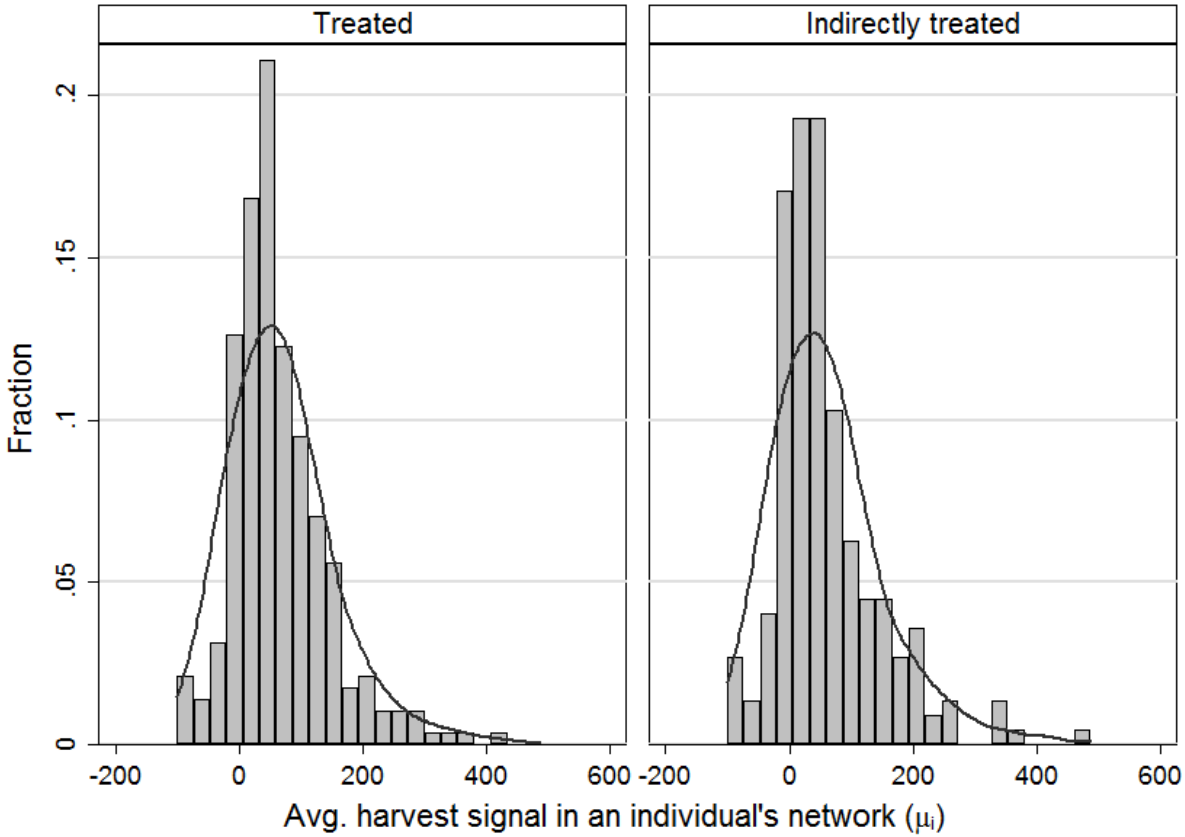


Figure : Distribution of μ_i

Outcome variables

- Familiarity with WSC hybrids
- WTP for WSC hybrids
- Planted a WSC variety
- Planted a non-WSC variety

Familiarity with WSC hybrids

- Indicator variable equal to 1 if respondent is familiar with the technology
- 1^{st} stage of WTP module:
  - respondents shown cards with names of ca. 20 seed varieties
  - asked whether they feel they know enough about the varieties to decide whether or not they would like to plant them
- Measures whether respondent has enough knowledge about WSC hybrid to compare the tech to other seeds?
- Intuitively, have to be familiar with the seed before adopting
  - more restrictive than 'have you heard of WSC hybrids?'

Price-premium based WTP

- 2^{nd} stage of WTP module:
  - rank the seeds with which familiar
- 3^{rd}stage:
  - if ranked a WSC variety over another hybrid, elicited premium
  - add premium to the price of the other hybrid
- Could pick up learning if adoption impacts are limited by liquidity constraints and/or other market imperfections
- Not everyone answers the WTP module

Actual planting behavior

- Planted a WSC variety (0/1)
  - more stringent measure of adoption than other experimental network papers
    - Bandiera & Rasul, 2006; Cai et al, 2014; Oster & Thornton, 2012; Miguel & Kremer, 2004
- Planted a non-WSC hybrid
  - could be 0, positive or negative depending on previous hybrid use and/or spillovers

General specification

$$y_{iv} = f(N_{iv}) + \gamma \mathbf{X}_i + \varepsilon_{iv}$$

- y_{iv} is outcome for household i in village v
- $\mathbf{X}_i$ is vector of baseline control variables
- $f(N_{iv})$ function of information in individual i's network
- s.e.'s clustered at village level

General specification

$$y_{iv} = f(N_{iv}) + \gamma \mathbf{X}_i + \varepsilon_{iv}$$

- N_{iv} represents either
  1. number of treated farmers in farmer i's network
  2. first two moments of distribution of experiences reported by treated individuals in her network
- Recent experimental studies typically only consider 1)

General specification

- These "social influence" models include the number of treated in network in different forms
  - # of treated (Babcock & Hartman, 2010; Oster & Thornton, 2012)
  - share of treated (Cai et al., 2014)
  - indicator vars for having 1,2, 3... treated members (Carter et al., 2014)
  - dummy for having *any* treated network members (Magnan et al., 2013)
- I use dummies for 1 and "2 or more" treated network members

Social networks model

- 'Social influence' model:

$$y_{iv} = \alpha_1 + \beta_k \sum_{k=1}^{K} I_{iv}^k + \gamma_1 \mathbf{X}_i + \varepsilon_{iv}$$

where K in our preferred model is 2+

- Information signal model:

$$y_{iv} = \alpha_2 + \lambda_k \sum_{k=1}^{2} m_{iv}^k + \gamma_2 \mathbf{X}_i + \nu_{iv}$$

m_i^k denotes the k^{th} moment of the distribution of signals in person i's network

Social networks model

- Estimate most models using OLS
- When outcome variable is WTP for technology, use Tobit as it might be censored at 0

Social networks model

- Controls include
  - proxies for prior experience with improved tech:
    - dummy for being in a village where the majority of treated do *not* know where to purchase
    - dummy for having used hybrids & fertilizer
  - household characteristics:
    - size of main maize field
    - risk attitudes
    - understanding score from experiments
    - PPI score
    - microfinance participation
  - network controls:
    - total network size; signal-regressions also dummies for number of treated links

Heterogeneity

- Cation Exchange Capacity (CEC): summary statistic of soil quality
  - often used to gauge soil fertility
  - varies in sample villages & the *extent* of variation also varies between villages
- Compute the coefficient of variation (CV) of CEC: measure of unobserved heterogeneity
- Interact CV_{CEC} with social network variables

Summary statistics

Variable	mean	sd	min	max	mean(T) - mean(I)	*t*-stat
Household characteristics						
Kiswahili spoken at home	0.03	0.18	0	1	-0.001	(-0.06)
Luhya spoken at home	0.19	0.39	0	1	0.045	(1.41)
Luo spoken at home	0.78	0.42	0	1	-0.045	(-1.29)
In womens' or farm group	0.48	0.50	0	1	0.076*	(1.83)
In microfinance group	0.25	0.43	0	1	0.009	(0.25)
General risk taking attitude (0-10)	8.15	2.04	0	10	0.081	(0.47)
Understanding score, exp. games	0.74	0.34	0	1	-0.024	(-0.85)
PPI score (0-100)	44.49	12.41	14	84	1.409	(1.35)

t statistics in parentheses, standard errors clustered at the village level
* p<.1, ** p<.05, *** p<.01

Table : Summary statistics

Summary statistics

Variable	mean	sd	min	max	mean(T) - mean(I)	*t*-stat
Agricultural characteristics						
Size of main maize field (acres)	1.30	1.16	.07	10	0.201**	(2.16)
Nr. of seasons used fertilizer, 4 years	2.57	3.33	0	8	0.479*	(1.71)
Nr. of seasons used hybrids, 4 years	3.32	3.33	0	8	-0.059	(-0.21)
Network characteristics						
Nr. of relatives	2.43	2.23	0	12	0.070	(0.38)
Nr. of treated relatives	1.31	1.39	0	8	0.080	(0.69)
Nr. of links (all)	7.05	3.92	0	29	0.344	(1.08)
Nr. of treated links (all)	4.08	2.51	0	20	0.549***	(2.69)
Nr. of reciprocal links (all)	3.29	2.50	0	22	0.409**	(2.01)
Nr. of treated reciprocal links (all)	1.93	1.71	0	15	0.435***	(3.15)
Nr. of links in corrected network	6.73	3.78	0	29	0.154	(0.50)
Nr. of treated links, corrected network	3.85	2.41	0	19	0.400**	(2.03)

t statistics in parentheses, standard errors clustered at the village level
* p<.1, ** p<.05, *** p<.01

Balance on observables

- Require that treatment induced exogenous variation in number of treated network members in a given individual's network
  - conditional on individual *i*'s total number of links (total network size), the number of *treated* links was randomized
  - test the validity this assumption by regressing baseline characteristics on number of treated links (controlling for total network size)
- Do this separately for treated & indirectly treated
- Test using 3 different network definitions

Balance on observables

Variable	Coeff. on nr. of treated links, controlling for nr. of links			
	Relatives		Corrected	
	T	I	T	I
Household characteristics				
In womens' or farm group	-0.009 (-0.20)	0.008 (0.23)	-0.012 (-0.59)	0.024 (1.21)
In microfinance group	-0.047* (-1.90)	-0.002 (-0.07)	-0.013 (-0.89)	0.040*** (3.57)
General risk taking perception (0-10)	-0.089 (-0.50)	0.018 (0.12)	-0.061 (-1.03)	-0.033 (-0.34)
Understanding score, exp. games	-0.010 (-0.42)	0.035 (1.33)	-0.012 (-1.16)	0.017 (0.88)
Sum of core 10 PPI scores (0-100)	-0.506 (-0.68)	1.248 (1.09)	-0.354 (-0.52)	0.655 (1.02)

t statistics in parentheses, standard errors clustered at the village level

* p<.1, ** p<.05, *** p<.01

Table : Regression of baseline vars on nr. of treated links

Balance on observables

Variable	Coeff. on nr. of treated links, controlling for nr. of links			
	Relatives		Corrected	
	T	I	T	I
Agricultural characteristics				
Size of main maize field (acres)	-0.026 (-0.27)	0.024 (0.35)	-0.029 (-0.55)	-0.038 (-0.69)
Nr. of seasons used fertilizer, 4 years	0.440 (1.37)	0.271 (1.07)	0.303 (1.56)	0.536*** (3.21)
Nr. of seasons used hybrids, 4 years	0.334 (1.26)	0.882*** (2.92)	0.244 (1.32)	0.628*** (3.88)

t statistics in parentheses, standard errors clustered at the village level

* p<.1, ** p<.05, *** p<.01

Table : Regression of baseline vars on nr. of treated links

Familiarity, social influence model

(Dep. variable: Familiar with WSC hybrid?)

Panel A -	**Treated**			**Indirectly treated**	
Nr. of treated links	1	2	3	4	5
1 treated link	0.20 (0.2)	0.097 (0.3)	0.29 (0.3)	0.020 (0.1)	0.53*** (0.2)
2+ treated links	0.31 (0.2)	0.50* (0.3)	0.47* (0.3)	0.082 (0.2)	0.36** (0.2)
Network size	0.0071 (0.006)	0.13 (0.1)	0.0042 (0.007)	0.013 (0.01)	0.19*** (0.06)
(1 treated)*(nw. size)		-0.036 (0.1)			-0.23*** (0.07)
(2+ treated)*(nw. size)		-0.12 (0.1)			-0.18** (0.06)
On-farm trial outcome			0.00067 (0.03)		
(On-farm trial outcome)^2			0.00016 (0.002)		
Additional covars	YES	YES	YES	YES	YES
Observations	319	319	217	255	255
Adjusted R^2	0.078	0.083	0.087	0.229	0.237

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

Table : Social network effects on farmer familiarity with WSC hybrids

Familiarity, information signal model

(Dep. variable: Familiar with WSC hybrid?)

	Treated		**Indirectly treated**
Panel B - Signal in nw	1	2	3
Avg. signal in nw.	0.022	-0.027	0.00024
	(0.03)	(0.04)	(0.01)
Variance of signal in nw.	-0.0000016	0.0022	-0.0046***
	(0.002)	(0.002)	(0.0010)
Network size	0.0066	0.0019	0.014
	(0.006)	(0.007)	(0.01)
On-farm trial outcome		0.0073	
		(0.03)	
(On-farm trial outcome)^2		-0.00017	
		(0.002)	
Additional covars	YES	YES	YES
Observations	294	202	227
Adjusted R^2	0.042	0.006	0.238

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

Table : Social network effects on farmer familiarity with WSC hybrids

WTP, social influence model

(Dep. variable: Willingness to pay for WSC hybrid)

Panel A -	**Treated**		**Indirectly treated**
Nr. of treated links	1	2	3
1 treated link	83.0 (77.0)	84.1 (126.7)	314.9*** (73.9)
2+ treated links	116.8** (51.7)	96.1 (108.9)	263.0*** (66.2)
Network size	2.40 (3.8)	4.13 (4.5)	9.49 (9.3)
On-farm trial outcome		26.6 (18.1)	
(On-farm trial outcome)^2		-1.80 (1.1)	
Additional covars	YES	YES	YES
Observations	224	173	96
Adjusted R^2	0.064	0.087	0.075

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

Table : Social network effects on farmer WTP for WSC hybrids

WTP, information signal model

(Dep. variable: Willingness to pay for WSC hybrid)

Tobit regression	**Treated**		**Indirectly treated**
Panel B - Signal in nw	1	2	3
Avg. signal in nw.	31.0**	25.6	109.0***
	(14.2)	(16.7)	(19.8)
Variance of signal in nw.	-1.55**	-1.03	-17.5***
	(0.8)	(0.9)	(6.1)
Network size	3.92	5.78	14.0
	(4.2)	(5.1)	(8.6)
On-farm trial outcome		30.9	
		(21.8)	
(On-farm trial outcome)^2		-2.04	
		(1.4)	
Additional covars	YES	YES	YES
Observations	215	168	92
σ	227.2***	223.4***	217.5***

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

WSC hybrid adoption, social influence model

(Dep. variable: Planted WSC hybrid?)

Panel A -	**Treated**		**Indirectly treated**
Nr. of treated links	1	2	3
1 treated link	0.35***	0.32***	-0.012
	(0.08)	(0.08)	(0.04)
2+ treated links	0.13**	0.16*	0.029
	(0.06)	(0.08)	(0.03)
Network size	0.0066	0.0051	0.0023
	(0.006)	(0.006)	(0.005)
On-farm trial outcome		0.039	
		(0.02)	
(On-farm trial outcome)^2		-0.0029*	
		(0.001)	
Additional covars	YES	YES	YES
Observations	319	217	255
Adjusted R^2	0.083	0.073	0.045

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

Table : Social network effects on probability of planting a WSC hybrid

WSC hybrid adoption, information signal model

(Dep. variable: Planted WSC hybrid?)

	Treated		**Indirectly treated**
Panel B - Signal in nw	1	2	3
Avg. signal in nw.	-0.023	-0.032	-0.00015
	(0.02)	(0.03)	(0.005)
Variance of signal in nw.	0.0034	0.0044**	0.0012
	(0.002)	(0.002)	(0.002)
Network size	0.0065	0.0048	0.0041
	(0.006)	(0.006)	(0.005)
On-farm trial outcome		0.042	
		(0.03)	
(On-farm trial outcome)^2		-0.0029*	
		(0.001)	
Additional covars	YES	YES	YES
Observations	294	202	227
Adjusted R^2	0.088	0.072	0.035

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

Planted other hybrid, social influence model

(Dep. variable: Planted a non-WSC hybrid?)

Panel A -	**Treated**		**Indirectly treated**
Nr. of treated links	1	2	3
1 treated link	-0.35*	-0.21	0.0079
	(0.2)	(0.2)	(0.2)
2+ treated links	-0.19	-0.14	-0.013
	(0.1)	(0.1)	(0.2)
Network size	0.0080	0.013	-0.0024
	(0.007)	(0.008)	(0.010)
On-farm trial outcome		0.074**	
		(0.03)	
(On-farm trial outcome)^2		-0.0034	
		(0.002)	
Additional covars	YES	YES	YES
Observations	319	217	255
Adjusted R^2	0.166	0.128	0.276

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

Table : Social network effects on probability of planting a non-WSC hybrid

Planted other hybrid, information signal model

(Dep. variable: Planted a non-WSC hybrid?)

	Treated		**Indirectly treated**
Panel B - Signal in nw	1	2	3
Avg. signal in nw.	0.027	0.021	0.0062
	(0.03)	(0.04)	(0.01)
Variance of signal in nw.	-0.0040*	-0.0037*	-0.0089***
	(0.002)	(0.002)	(0.003)
Network size	0.0089	0.012	-0.0016
	(0.008)	(0.008)	(0.01)
On-farm trial outcome		0.077**	
		(0.03)	
(On-farm trial outcome)^2		-0.0035	
		(0.002)	
Additional covars	YES	YES	YES
Observations	294	202	227
Adjusted R^2	0.170	0.110	0.311

In both panels: standard errors in parentheses;
s.e.'s clustered at the village level; * p<.1, ** p<.05, *** p<.01
Network definition used: individual *j* is in person *i*'s network
if person *i* listed them in *any* of the network questions.

Familiarity

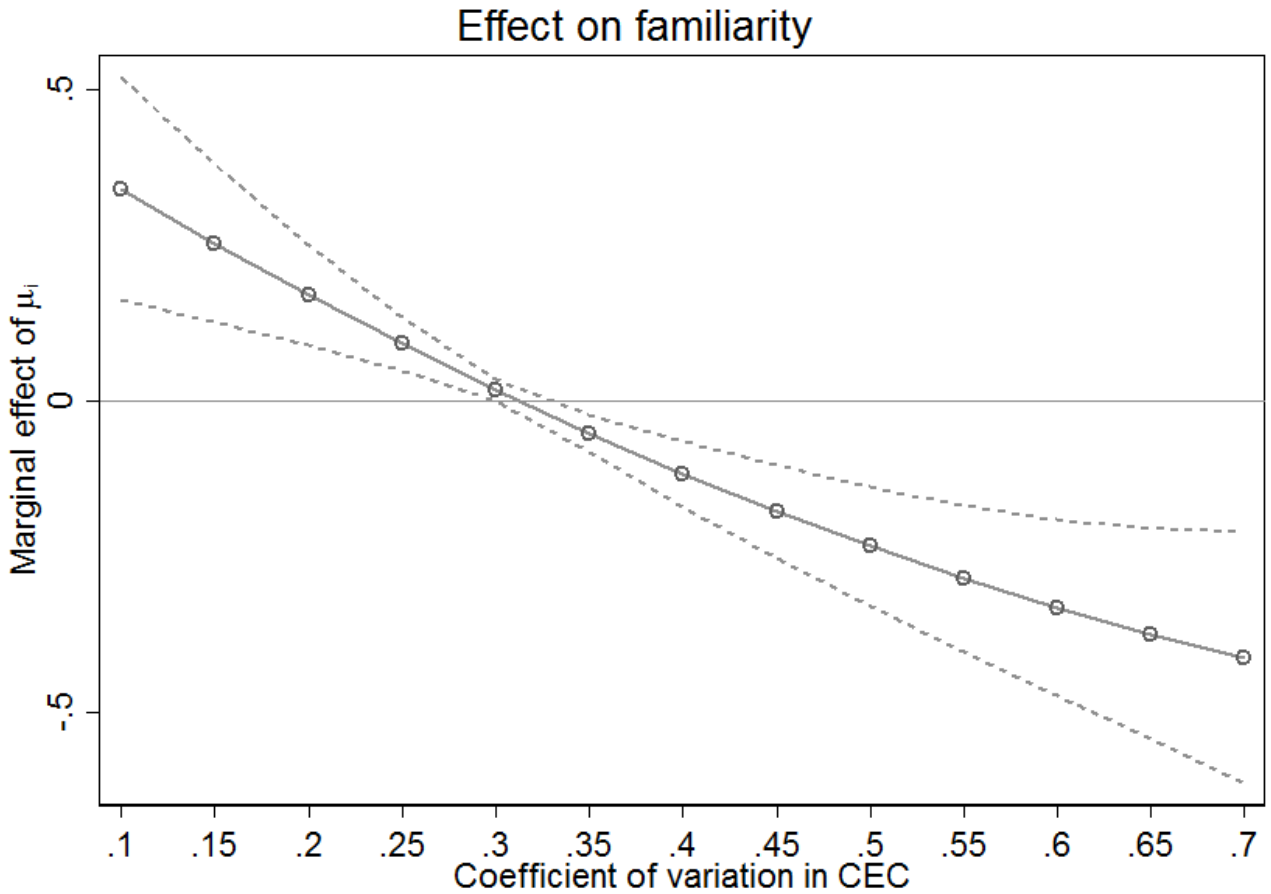


Figure : How impact of avg. signal in nw. varies with heterogeneity

WTP

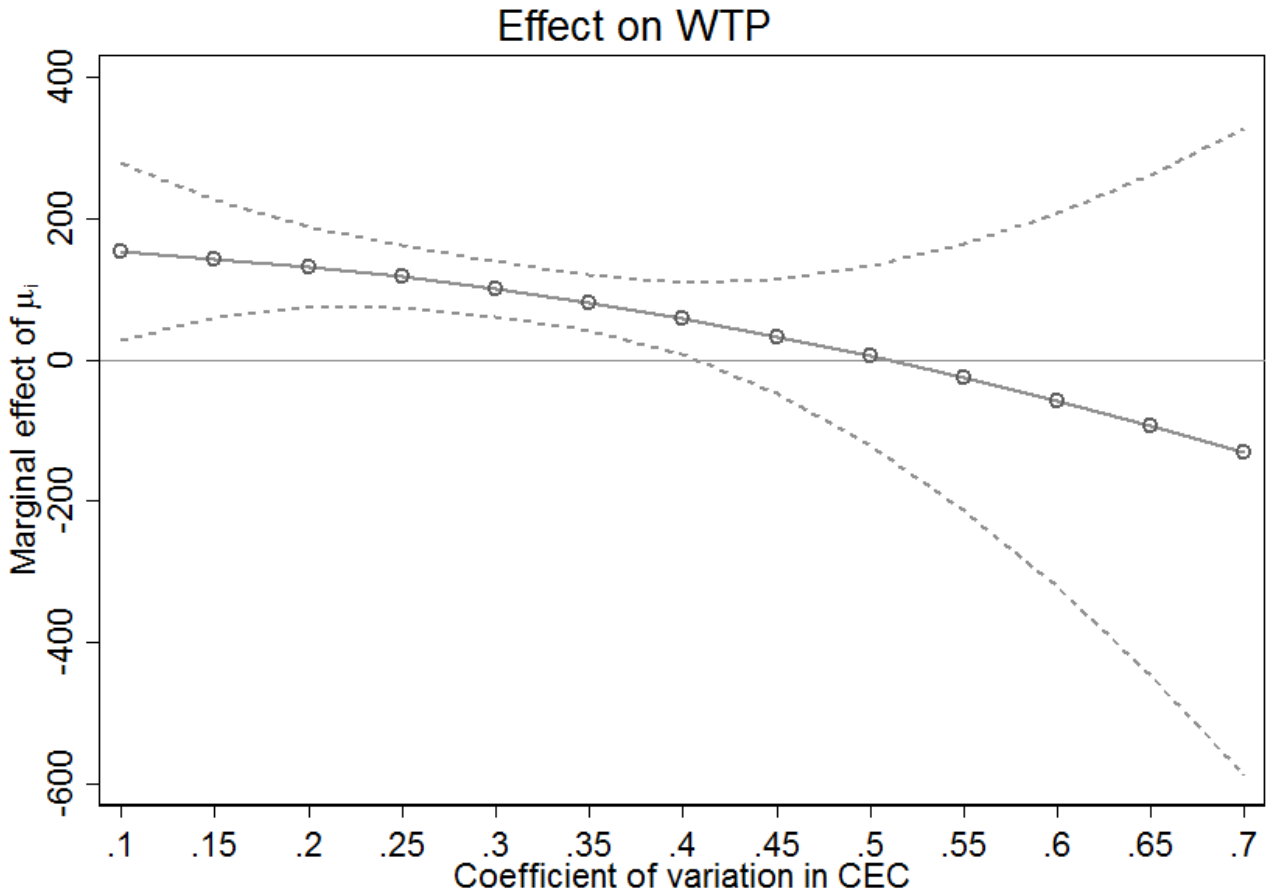


Figure : How impact of avg. signal in nw. varies with heterogeneity

WSC adoption

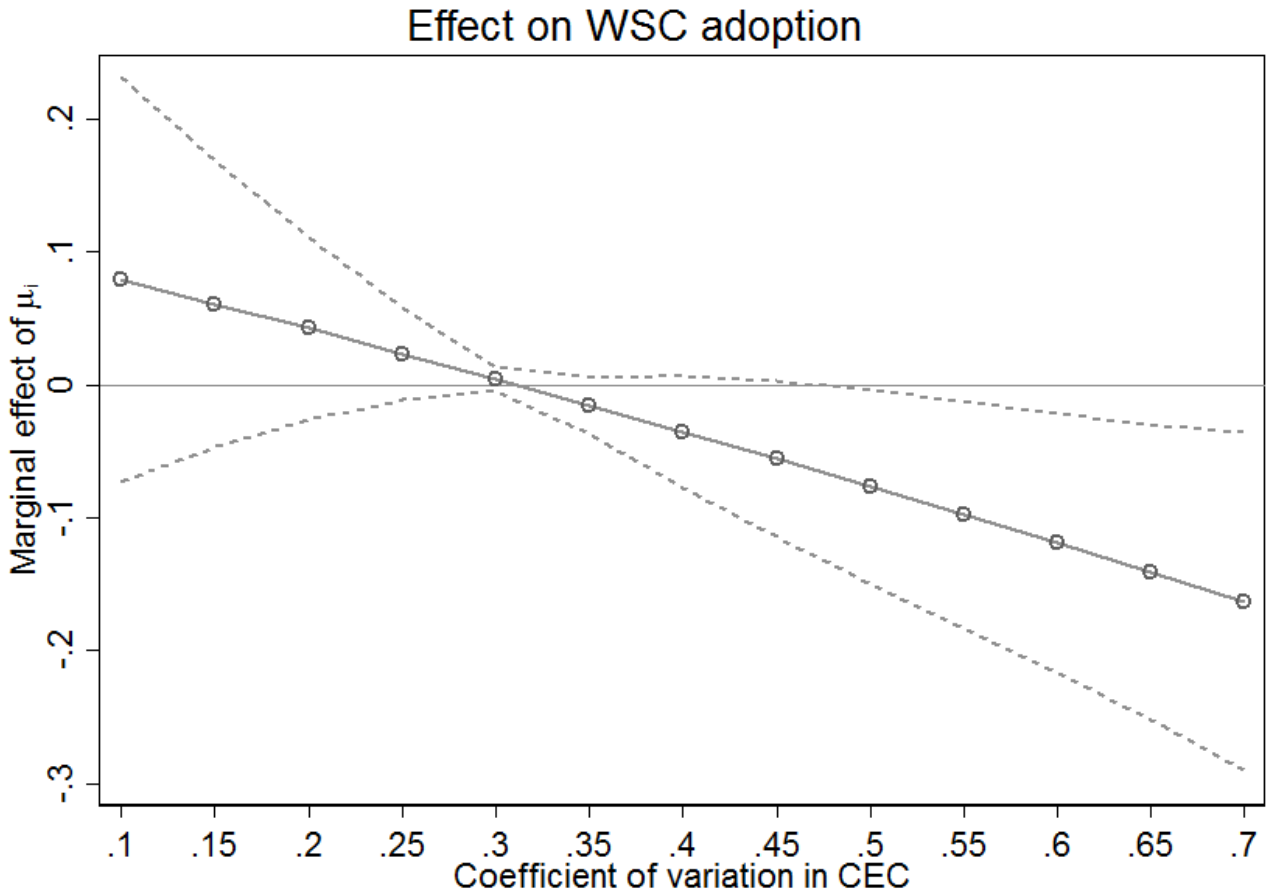


Figure : How impact of avg. signal in nw. varies with heterogeneity

Other hybrid

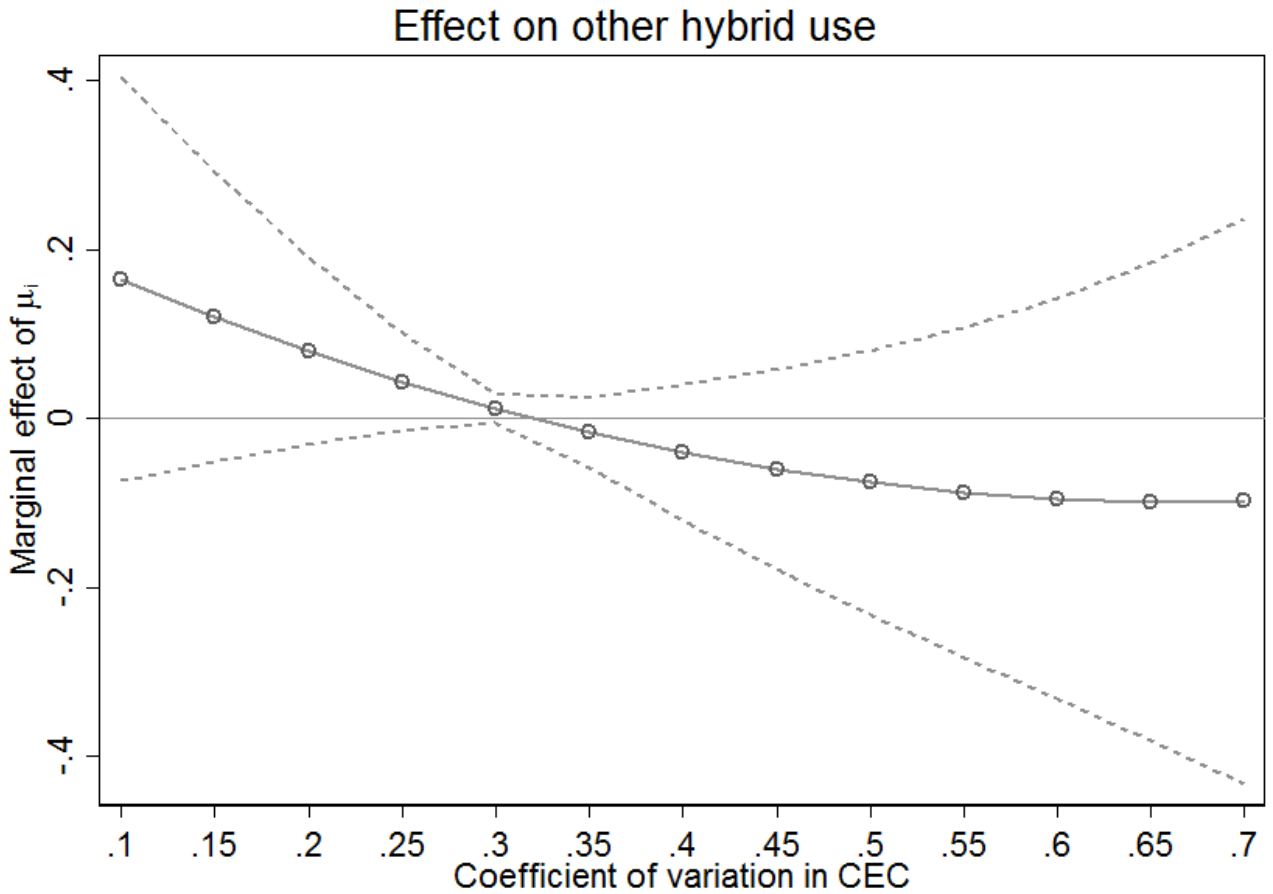


Figure : How impact of avg. signal in nw. varies with heterogeneity

Conclusion

- Use experimental variation in information available through networks to study what farmers learn from their social networks
- Farmers talk and learn from each other BUT heterogeneity that is unobserved to farmers makes them rely less on information from their peers
- Can help us understand why some innovations diffuse slowly
- Can inform policy:
  - when will broad-based extension programs be successful?
  - when do we need to promote individual learning?
- Also useful for thinking about other stochastic tehcnologies