Title

Risky Income or Lumpy Investments? Evidence on Two Theories of Under-Specialization

Author

Shenoy, Ajay

Publication Date

2014-09-13

Risky Income or Lumpy Investments? Evidence on Two Theories of Under-Specialization

Ajay Shenoy*
University of California, Santa Cruz

September 13, 2014
First Version: October 08, 2012

Abstract

Why do the poor have so many economic activities? According to one theory the poor do not specialize because relying on one income source is risky. I test the theory by measuring the response of Thai rice farmers to conditional volatility in the international rice price. Households expecting a harvest take on 1 extra activity when the volatility rises by 21 percent. I confirm the decrease in specialization costs households foregone revenue. I find no evidence to back a second theory in which households under-specialize because they cannot afford lumpy business investments. (JEL Codes: D13, O12)

*Email at azshenoy@ucsc.edu. Special thanks to Raj Arunachalam, David Lam, Jeff Smith, David Weil, Anja Sautmann, Chris Udry, Silvia Prina, Stefanie Stantcheva, seminar participants at Michigan and Brown, and conference attendees at PacDev and MWIEDC in 2013 for helpful comments and advice. I am also grateful to Robert Townsend and the Townsend Thai Project for collecting and distributing the data I use for this project. File formatting based on a stylesheet by Chad Jones.

1 Introduction

> *To take...the trade of the pin-maker...One man draws out the wire, another straights it, a third cuts it...ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day...But if they had all wrought separately and independently...they certainly could not each of them have made twenty, perhaps not one pin in a day...*
>
> -Adam Smith, *Wealth of Nations*

The idea that specialization is efficient is as old as economics itself. The puzzle, then, is to explain why households in poor countries rarely specialize in a single business or a single job (Banerjee and Duflo, 2007). If entering multiple economic activities is costly, why would the world's poorest people fail to specialize?

This paper tests two well-known but unproven theories for why the poor have so many economic activities: the theory of risky income and the theory of lumpy investments. The theory of risky income compares a poor household choosing economic activities to an investor choosing stocks. Like stocks the activities of the poor have risky returns, driving households to diversify their portfolio even though expanding it is costly. Whereas this theory blames under-specialization on a lack of insurance, the theory of lumpy investments blames a lack of credit. The theory posits that households must make a large investment—tailors must buy a sewing machine and bakers must buy an oven—before expanding any business to its optimal scale. Households that cannot borrow enough to create one large business must cobble together income from many small businesses.

From a simple model I derive four tests of the theory of risky income. Each household has a primary activity and pays a fixed cost to enter any side activity. The returns to these activities are random and not perfectly correlated. Therefore the theory's first test is that a rise in the riskiness of the primary activity causes a risk-averse household to self-insure by entering more side activities. But since labor spent on side activities is labor taken from the primary activity, a rise in the average return to the primary activity raises the cost of self-insurance. The theory's second test is that a rise in the return to the primary activity causes the household to exit side activities.

The third test, which uses revenue from side activities rather than total revenue to check whether specialization is efficient, is critical for two reasons. First, my empirical approach rules out any test using total revenue. Second, the household reallocates labor between activities when it enters new activities. For both reasons I cannot identify the fixed cost of entering a new activity, the clearest sign of inefficient under-specialization. I can, however, derive the optimal allocation of labor as a function of the number of activities, and use this allocation to find the change in side revenue caused when the household enters a new activity. I show that if the change is negative then specialization is inefficient, though the converse need not hold.

To run these tests I study how rice farmers in Thailand respond to volatility in the international price of rice. Using a monthly panel I identify the households who expect a rice harvest in the next three months. Higher volatility in the price of rice raises the riskiness of these farmers' income. By comparing their response to the response of farmers who do not expect a harvest I identify the causal effect of riskier income on specialization. By likewise comparing how the two groups respond to changes in the expected price I identify the causal effect of greater returns on specialization. My first two tests confirm the theory of risky income. Greater risk drives households into more activities while higher returns tempt households out of activities. After adjusting for how well international prices predict local prices, my baseline estimates suggest a 21 percent rise in volatility causes a household to enter 1 extra activity.

Since a household that expects a harvest next month sells no rice this month the mean and variance of the rice price change the number of activities without directly affecting current revenue. I use this change to instrument for the number of activities. Households expecting a harvest do not yet have the revenue from their primary activity, ruling out any test of whether additional activities decrease total revenue. But my third test shows that if additional activities decrease revenue from side activities then a failure to specialize is inefficient. Two-stage least squares confirms exactly that. I confirm that these results are not driven by changes in household labor or composition, by negative shocks rather than volatility, or a correlation between the price of rice and the price of other crops.

Finally, I test the theory of lumpy investments. The theory predicts that households with easier access to credit can afford the lumpy investments that let them specialize in one business. I test whether households exit activities after a government program creates random variation in the availability of credit, but find no evidence that credit increases specialization.

Existing work links risk to under-specialization but lacks the exogenous variation needed to show that risk causes under-specialization.[1] Morduch (1990) shows that households more vulnerable to income shocks tend to diversify their crops, and Bandyopadhyay and Skoufias (2012) find that households in areas with riskier weather tend to have spouses with different occupations. But since vulnerability and weather risk are not exogenous, households who endure these problems may endure other problems unrelated to the riskiness of their income. If these other problems also cause under-specialization then estimates of the effect of risk will be biased.

More recent work, on the other hand, uses exogenous variation but does not study the effect of risk on specialization. Karlan et al. (2012b) and Cole et al. (2013) run field experiments to show that weather insurance causes household to shift production towards riskier crops, and Emerick et al. (2014) find similar results in an experiment that distributes drought resistant seeds. But these studies do not examine the number of economic activities. Adhvaryu et al. (2013) study whether households expand their number of activities in response to shocks, but entering activities in response to shocks is not the same as entering activities in anticipation of risk. The first is a way to cope with a bad outcome whereas the second is a way to insure against that outcome.

Since households face shocks on occasion but face risk every day, using plausibly exogenous variation in ex ante risk is the only way to answer my original question: why households in poor countries have so many economic activities. Equally important, I test the theory of lumpy investment alongside the theory of risky income, as this is the only way to assess which theory has more merit.

[1]Many more papers study how imperfect insurance drives households to make other inefficient choices. Those most relevant to this paper whether farmers with riskier profits marry their daughters to men in different occupations (Rosenzweig and Stark, 1989), choose safer but less profitable bundles of investments (Rosenzweig and Binswanger, 1993; Bliss and Stern, 1982), or delay the planting of their crops (Walker and Ryan, 1990, p. 256).

2 Theory: A Model of Risky Income

2.1 Deriving Tests for the Theory

Each household has one primary economic activity and may enter any number of side activities. The household pays a fixed cost for each side activity. For simplicity I model this cost as a literal cash payment, but it might be more realistic to think of it as the opportunity cost of whatever labor is wasted while switching between activities. The household allocates one unit of labor between all activities. Labor produces a constant return, and the household does not know the return to any activity until after it has made its choices.

The household must first choose the number of side activities. Then it chooses the allocations of labor. Then the returns to the side activities are realized. Finally, the household learns the return to its primary activity and consumes.

Suppose for simplicity that the household has constant absolute risk-averse preferences. The household solves

$$\max_{M, L_p, \{L_{s,m}\}} \mathbb{E}[-e^{-\alpha C}]$$

subject to

$$C = Y = w_p L_p + \sum_{m \in M} w_{s,m} L_{s,m} - MF$$

$$L_p + \sum_m L_{s,m} = 1$$

where α is the coefficient of absolute risk aversion, $M \geq 0$ is the number of side activities, and L_p and $\{L_{s,m}\}_{m \in M}$ are the labor allocated to the primary and each side activity. The household consumes its revenue, which is the sum of revenue from primary (p) and side (s) activities minus fixed costs. The primary and side activities yield returns w_p and $\{w_{s,m}\}_{m \in M}$, which are independent normal random variables with $w_p \sim N(\bar{w}_p, \sigma_p^2)$ and $w_{s,m} \sim N(\bar{w}_s, \sigma_s^2)$ for each m.[2]

[2] If the returns to side and primary activities were not independent, the properties of normal random variables let me write the returns to each side activity as $w_{s,m} = \rho^m w_p + \xi^m$ for some correlation coefficient ρ^m. If I then re-label variables accordingly, all the results should go through. I only need the returns to be imperfectly correlated.

Assume the side activities yield weakly lower expected returns: $\bar{w}^p \geq \bar{w}^s$. Also assume the average premium to the primary activity, $\bar{w}_+ = \bar{w}_p - \bar{w}_s$, is not too large: $\bar{w}_+ < \alpha\sigma_p^2$. If this assumption fails the household will specialize despite the risk. If $\bar{w}^p = \bar{w}^s$ the household is no better at the primary activity than any other, but even then it is optimal for the household to specialize.

I make many simplifying assumptions about functional forms, but the important results rest on four crucial assumptions. First, the household is risk-averse. Second, the household cannot perfectly smooth its consumption through insurance or savings. (To sharpen the model's predictions I assume the household has no insurance or savings.) Third, the returns to side activities are not perfectly correlated with returns to the primary activity. Fourth, each activity has (locally) increasing returns. The first two assumptions force the household to insure itself against risk without using financial markets. The third assumption makes under-specialization a form of insurance. The fourth assumption makes under-specialization costly.

To get the intuition of the model, consider the simple case where the household either specializes ($M = 0$) or has one side activity ($M = 1$). The household chooses between two "bundles" of average consumption $\bar{C}$ and variance of consumption V:

	$\boldsymbol{M = 0}$	$\boldsymbol{M = 1}$
$\bar{\boldsymbol{C}}$	$\bar{w}_p$	$\bar{w}_p - \bar{w}_+(1 - L_p) - F$
$\boldsymbol{V}$	σ_p^2	$(L_p)^2\sigma_p^2 + (1 - L_p)^2\sigma_s^2$

Since $L^p < 1, \bar{w}_+ > 0$ and $F > 0$ the household can lower the variance of its consumption by entering a side activity if it accepts a lower expected consumption.

Suppose the household enters a side activity and must now choose how much labor to shift from the primary activity. Since consumption is a normal random variable, expected utility is (the negative of) a log normal random variable. The household now solves

$$\max_{L^p} -e^{-\alpha\bar{C} + \frac{\alpha^2}{2}V}.$$

The first-order condition is

$$
\begin{aligned}
& 0=-e^{-\alpha \bar{C}+\frac{\alpha^2}{2} V} \cdot\left(-\alpha \frac{\partial \bar{C}}{\partial L_p}+\frac{\alpha^2}{2} \frac{\partial V}{\partial L_p}\right) \\
\Rightarrow \quad & 0=-\bar{w}_{+}+\alpha L_p \sigma_p^2-\alpha\left(1-L_p\right) \sigma_s^2 \\
\Rightarrow \quad & L_p=\frac{\alpha \sigma_s^2+\bar{w}_{+}}{\alpha\left(\sigma_p^2+\sigma_s^2\right)}
\end{aligned}
$$

To derive predictions about aggregate statistics, suppose the fixed cost of entering the side activity varies across households because some find it easier to enter activities. For example, two rice farmers might differ only in how closely they live to a construction site where they can find part-time work. For simplicity suppose $F \sim U[0, \mathcal{F}]$ for some upper-bound $\mathcal{F}$.

For any amount of risk there is a household whose fixed cost makes it indifferent between zero and one side activity. Call that household's fixed-cost $\bar{F}_0$. Let $C(M)$ and $V(M)$ be the mean and variance of consumption as functions of the number of side activities. Then $\bar{F}_0$ is defined as the fixed cost that makes this equation hold:

$$
\begin{aligned}
& -e^{-\alpha \bar{C}(0)+\frac{\alpha^2}{2} V(0)}=-e^{-\alpha \bar{C}(1)+\frac{\alpha^2}{2} V(1)} \\
\Rightarrow \quad & -\alpha \bar{C}(0)+\frac{\alpha^2}{2} V(0)=-\alpha \bar{C}(1)+\frac{\alpha^2}{2} V(1) \\
\Rightarrow \quad & \frac{\alpha}{2}[V(0)-V(1)]=\bar{C}(0)-\bar{C}(1)
\end{aligned}
$$

Substitute the expressions from the table above and from the optimal labor allocation:

$$
\bar{F}_0=\frac{\left(\alpha \sigma_p^2-\bar{w}_{+}\right)^2}{2 \alpha\left(\sigma_p^2+\sigma_s^2\right)}
$$

[Figure 1 around here]

Households who pay fixed costs above the threshold $\bar{F}_0$ will specialize while those below enter a side activity. The threshold rises with the variance of the primary activity σ_p^2, and Figure 1 shows the effect on the number of households with a side activity. When their primary activity becomes riskier, households are willing to pay a bigger fixed cost to make their revenue less risky. The threshold $\bar{F}_0$ rises, and the mass of households with fixed costs between the old and new thresholds enter a side activity. The change in the average number of activities in the sample is

$$\frac{\partial \mathbb{E}_F[M]}{\partial \sigma_p^2} = \frac{\sigma_p \left(\alpha\sigma_p^2 - \bar{w}_+\right)\left(\alpha\sigma_s^2\left(\sigma_p^2 + 2\sigma_s^2\right) + \bar{w}_+\sigma_s^2\right)}{\alpha\sigma_s^2\left(\sigma_p^2+\sigma_s^2\right)^2} \cdot \frac{1}{\mathcal{F}} > 0.$$

(Recall that by assumption $\alpha\sigma_p^2 - w_+ > 0$). Similarly we can derive the change in the average number of activities when the average return to the primary activity rises. Since a rise in the expected return makes under-specialization more costly, the threshold will fall and the average number of activities will fall:

$$\frac{\partial \mathbb{E}_F[M]}{\partial \bar{w}_p} = -\frac{\alpha\sigma_p^2 - \bar{w}_+}{\alpha\left(\sigma_p^2+\sigma_s^2\right)} \cdot \frac{1}{\mathcal{F}} < 0$$

The intuition of the case where $M \in \{0,1\}$ holds for any number of activities $M \in \{0,1,2,\ldots\}$, and the simple but tedious proof is left for Appendix A.1.

On average each additional activity will lower total revenue. But the empirical approach of Section 3 studies rice farmers who expect but have not yet collected a harvest. These farmers do not have their total revenue, ruling out any test based on total revenue. I must instead derive the model's predictions about what under-specialization does to revenue from side activities; that is, what happens to the rice farmer's revenue from growing cassava if he starts baking bread.

Consider the revenue of the household just before it gets the output from its primary activity. Its revenue at this stage is simply the revenue from its side activities:

$$y_s = \sum_{m \in M} w_{s,m} L_{s,m} - MF$$

For simplicity treat the number of activities M as continuous.[3] Holding a household's cost of additional activities fixed, a small increase in the number of activities changes side revenue on average by

[3] I could keep the number of activities discrete and compute the average conditional difference, but given that a regression coefficient is meant to capture an average marginal change the simplification seems justified.

$$
\begin{aligned}
\mathbb{E}_F\left[\frac{\partial y_s}{\partial M}\right] &= \mathbb{E}_F\left[\mathbb{E}_{w_p,\{w_{s,m}\}}\left[\frac{\partial y_s}{\partial M} \mid F\right]\right] \\
&= \mathbb{E}_F\left[\frac{\partial}{\partial M}\left[-MF + (1-L_p)\bar{w}_s\right]\right] \\
&= -\mathbb{E}[F] + -\bar{w}_s\frac{\partial L_p}{\partial M}
\end{aligned}
$$

The average change in side revenue, which corresponds to the instrumental variables coefficient estimated in Section 5.1, has two parts: the average fixed cost of a side activity and the effect on side revenue of shifting labor to the side activities. Since an all-else-equal increase in the number of activities makes the portfolio of side activities less risky, the household wants to shift labor away from its primary activity. Then $\frac{\partial L_p}{\partial M} < 0$ and the second term is positive. If large enough it will swamp the cost of under-specialization and make the derivative (and thus the instrumental variables estimate) positive. To see why, suppose the household starts with no side activities and thus no revenue from side activities. If the variance of the primary activity rises sharply and the cost of entering a side activity is small, then the household will want to enter the side activity. Then revenue from side activities will have increased, and though the increase might be small compared to what the household loses from its primary activity, the coefficient I estimate will be positive. Thus a negative estimate is sufficient evidence that an additional activity (and thus under-specialization) is costly, but it is not necessary evidence. This arguement ignores the direct effect that my instruments, the variance and the average returns, have on the labor allocation. But as I show in the proof in Appendix A, the direct effect only strengthens the result.

The model also makes a prediction about the ordinary least squares coefficient, which estimates the average effect of increasing the number of activities without holding their cost fixed. That is, it estimates the average total derivative

$$\begin{aligned}\mathbb{E}\left[\frac{dy^s(M,F)}{dM}\right] &= \mathbb{E}\left[\frac{\partial y^s}{\partial M} + \frac{\partial y^s}{\partial F}\cdot\frac{\partial F}{\partial M}\right]\\ &= \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] + \mathbb{E}\left[\frac{\partial y^s}{\partial F}\cdot\mathbb{E}\left[\frac{\partial F}{\partial M}\mid M\right]\right]\\ &= \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] + \mathbb{E}\left[\frac{\partial y^s}{\partial F}\cdot\frac{\partial}{\partial M}\mathbb{E}\left[F\mid M\right]\right]\end{aligned}$$

The term $\frac{\partial y^s}{\partial F}$ is clearly negative; a higher fixed cost will lower revenue. The term $\frac{\partial}{\partial M}\mathbb{E}\left[F \mid M\right]$ gives the selection bias. It captures the difference in fixed cost paid by households who select into many versus few activities. As Figure 2 illustrates, it is also negative. Since a household takes up a large number of activities if it pays a small fixed cost, the number of activities is informative about their cost.[4] This gives the final test of the model:

$$\begin{aligned}\beta_{OLS} &= \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right] + \mathbb{E}\left[\frac{\partial y^s}{\partial F}\cdot\frac{\partial}{\partial M}\mathbb{E}\left[F\mid M\right]\right]\\ &> \mathbb{E}\left[\frac{\partial y^s}{\partial M}\right]\\ &= \beta_{IV}\end{aligned}$$

[Figure 2 around here.]

To summarize, the model gives four tests for the theory of risky income:

Test 1 (Risk) *Households enter activities when the returns to their primary activity get riskier.*

Test 2 (Return) *Households exit activities when the (expected) returns to their primary activity rise.*

Test 3 (Cost) *The average effect of more activities on revenue is negative only if under-specialization is costly.*

Test 4 (OLS Bias) *Compared to the IV estimate, the OLS estimate of the effect of more activities on side revenue is biased positively.*

[4] Indeed, $\mathbb{E}\left[F \mid M\right]$ is just the demand curve for insurance through under-specialization. Like any demand curve its slope is negative.

2.2 Modeling and Measuring Expectations about Risk

To run these tests I must model farmers' expectations about the returns and volatility of the price of rice. Suppose the household makes its choices at the beginning of period t. It has not yet observed the price w_{pt} and must form its expectation $\bar{w}_{pt}$ using only information from the past. Suppose the monthly price follows the Autoregressive Conditional Heteroskedasticity (ARCH) model of Engle (1982) with one modification: I assume the level of the price follows a random walk. The assumption reduces the number of parameters I must estimate and, as I show below, matches the true series well. Then

$$\begin{aligned} w_{pt} &= w_{p,t-1} + \varepsilon_t \\ \varepsilon_t &= z_t \sqrt{h_t}, \qquad z_t \sim N(0,1) \\ h_t &= \tau_0 + \tau_1 \varepsilon_{t-1}^2. \end{aligned}$$

At the beginning of period t, the household expects a return of $\bar{w}_{pt} = \mathbb{E}[w_{pt}] = w_{p,t-1}$. The variance of the return is $\sigma_{pt}^2 = V(w_{pt}) = V(\varepsilon_t) = h_t = \tau_0 + \tau_1 \varepsilon_{t-1}$. I estimate the model using conditional maximum likelihood.[5] The predicted value $\hat{h}$ is a consistent estimate of the true conditional variance.

In practice I must make several simplifications when I use this measure. I cannot use the actual expected volatility of the price at harvest because the empirical design in Section 3 compares farmers expecting a harvest to non-farmers and farmers who do not expect a harvest. Since I cannot define the volatility at the time of harvest for non-farmers I must use the current volatility. This creates measurement error and may bias my estimates towards zero. I also measure volatility using the conditional standard deviation $\sqrt{h}$ rather than the conditional variance to make the coefficients on the volatility of the price and the expected price comparable.

[Figure 8 around here]

Figure 8 plots the actual price of rice, the predicted mean, and the predicted standard deviation. Simple though it is, the random walk assumption makes very accurate predictions about the mean. A regression of price on its lag gives a coefficient of .995. The estimated equation for volatility is $\hat{h}_t = 53.3 + .39\varepsilon_{t-1}^2$.

[5] The true distribution of z_t need not be normal; the (quasi) maximum likelihood estimator based on a normal distribution is still consistent.

The red lines mark the start and end of the time period covered in the monthly panel data. The sample spans a time when prices are relatively stable, ending well before the massive food price spike of 2008. [6]

3 Empirical Design: Implementing the Tests

3.1 Estimating Risk Response

Changes in the international price of rice—and the responses they evoke in Thai rice farmers—provide the exogenous variation in risk I need to test the model. Between planting and harvest the price can change drastically, and anecdotal evidence suggests farmers follow it closely in newspapers, radio broadcasts, and television reports. Since most of my sample grows at least some of the white rice and jasmine rice that make Thailand the world's biggest rice exporter, the international price matters.[7] In Column 1 of Table 1 I report the correlation between the sample-wide average price farmers receive and the international price. Though not perfect, the correlation is significant and large enough to make following the international price worth a farmer's time. If prices become more volatile the farmers know it and know the value of their harvest has become riskier.

[Table 1 around here]

A response to volatility need not be a response to risky income unless it comes from a specific group of farmers: those who harvest soon. Simply comparing the response of a rice farmer to someone who does not farm rice might just measure how rice farmers differ in their attitude to risk. By contrast, observing a household with rice planted but not yet harvested—a farmer expecting a

[6]The reader may worry if regressions on a regressor generated from a time series model are consistent. Pagan (1984) confirms that the ARCH predicted value (though not the residual) will give consistent estimates, and I have confirmed in monte carlo simulations that panel estimators are consistent as well.

[7]As expected, I find in unreported regressions that farmers harvesting only sticky rice, which is not exported, have a lower response. The size of the difference is too large to be the all-else-equal effect of growing rice that will not be exported. As households who grow only sticky rice are unusual, their response may differ from that of other farmers for reasons beyond the type of rice.

harvest in the next three months—isolates the effect of risky income. Farmers harvest rice roughly four months after planting and cannot hasten or delay the date. Harvesting too soon yields immature grains while harvesting too late risks losses to pests. The International Rice Research Institute states that "the ideal harvest time lies between 130 and 136 days after sowing for late" varieties and gives similarly narrow windows for other varieties (Gummert and Rickman, 2011). Leaving rice on the stalk to wait out low prices is not an option. With a growing period of 3 to 4 months, it would also be difficult for farmers to strategically plant by predicting the volatility at harvest. My estimates from Section 2.2 suggest a 1 dollar rise in volatility at planting predicts only a 2 to 6 cent rise at harvest.

Although in principle a farmer might store rice after harvesting, threshing, and drying, in practice the farmers in my sample sell most of their rice as soon as they harvest it. Colum 2 of Table 1 reports the correlation between how much rice a household sells and how much it harvests conditional on harvesting any during the month. It suggests farmers sell almost every kilogram of rice as soon as it comes from their fields. Households either cannot arbitrage—perhaps because millers and other middlemen only buy at certain times—or they need cash too desperately to wait.

[Figures 7 and 9 around here]

The farmers in my sample are too small to affect the international price and they cannot delay their harvest. After controlling for the responses of non-rice farmers and rice farmers not expecting a harvest, any additional response must be caused by riskier income.[8] Though bad weather in Thailand might affect the price, it would have to hit the entire country. Moreover, though Thailand is a big exporter of rice, it is far from the biggest producer, which is China. Bad weather in China and India is more likely to drive prices than bad weather in Thailand.[9] Since I have a panel I can also control for household fixed-effects to eliminate any fixed source of bias.[10] The regression I run will actually compare the farmer

[8]In principle I can drop the non-rice farmers from my regressions and still get consistent (albeit noisier) estimates. I confirm in Section D that estimating Equation with only rice farmers does not change the results.

[9]Regardless, a big weather shock should change local wages as well the international price of rice, but in Appendix D I find no evidence that wages are affected by my regressor of interest

[10]For example, suppose only rich farmers plant in January to harvest in April. If the rice price always turns volatile in March, I might just estimate the effect of being a rich farmer. If seasonal

to herself at times when prices are volatile but she expects no harvest, and times when she expects a harvest but prices are not volatile. Figure 7 illustrates the specification.

When prices become volatile the farmer must decide whether to shift her efforts away from maximizing the upcoming harvest. Figure 9 graphs the average household labor that rice farmers devote to their fields in the months before and after harvest. Bringing a rice crop to harvest requires ceaseless effort. Working as a laborer or planting cassava on the side detracts from rice farming. Like in the model, side activities detract from the primary activity.

To estimate the response I run the regression

$$
\begin{aligned}
[Activities]_{it} = {} & [FE]_i + \beta_M[Mean]_t + \beta_V[Volatility]_t \\
& + \beta_E[Expecting\ Harvest]_{it} + \beta_H[Had\ Harvest]_{it} \\
& + \beta_{RM}[Rice\ Farmer]_i \times [Mean]_t + \beta_{RV}[Rice\ Farmer]_i \times [Volatility]_t \\
& + \beta_{EM}[Expecting\ Harvest]_{it} \times [Mean]_t + \beta_{EV}[Expecting\ Harvest]_{it} \times [Volatility]_t \\
& + \beta_{HM}[Had\ Harvest]_{it} \times [Mean]_t + \beta_{HV}[Had\ Harvest]_{it} \times [Volatility]_t + \varepsilon_{it}.
\end{aligned} \tag{1}
$$

Aside from the responses of non-farmers and farmers who do not expact a harvest, I must also control for the responses of farmers who just had a harvest. Having had a harvest is negatively correlated with expecting a harvest and cannot be left in the error term. In some regressions I replace the main effects [Mean] and [Volatility] with month dummies. Month dummies eliminate much of the variation in volatility but produce more conservative estimates. Since the volatility is generated, I use a two-stage bootstrap for all inference in the results I report in Section 5.1. The details of the bootstrap are in Online Appendix C.

The coefficient β_{EV} on $[ExpectingHarvest] \times [Volatility]$ measures the average response to volatility of a farmer who expects a harvest, while controlling for the responses of non-farmers and farmers without upcoming harvests. Since the number of activities is my measure of specialization, β_{EV} measures the causal effect of risk on under-specialization. Test 1 predicts it should be

selection and seasonal volatility matter, fixed-effects will deal with them.

positive. The coefficient β_{EM} on $[ExpectingHarvest] \times [Mean]$ measures the response to higher average prices, and Test 2 predicts it should be negative.

3.2 The Costs of Under-Specialization

Risk may drive households into extra side activities, but are they costly? It is hard to imagine why else the household would diversify only when risk increases. If the extra activities were costless the household ought to have as many as possible. Test 3, however, suggests a direct approach: to check whether revenue from side activities falls as the farmer adds more activities.

Rises in volatility will cause farmers expecting a harvest to increase their number of activities, but by construction these farmers have not yet sold their harvest and collected their primary revenue. I cannot run any test on total revenue. Test 3 solves the problem by showing that if revenue from side activities falls when the household adds activities, then under-specialization is costly. Test 3 says that if a rice farmer's revenue from cassava falls when he starts baking bread, and the loss to cassava outweighs the gain from bread, then extra activities are costly. I can confirm extra activities are costly if I have instruments that drive farmers into more activities without directly affecting revenue.

The response of farmers expecting a harvest to the mean and volatility of prices is exactly the instrument I need. Since household revenue before the harvest does not include revenue from rice, movements in the rice price cannot affect revenue directly. They may cause the household to reallocate labor away from rice farming, but Test 3 already accounts for the change in labor (see Appendix A.2). Greater risk might cause a household to invest less in physical and human capital. But the effect of any change in investment will not appear for years to come, whereas my regressions measure changes from month to month.

Since the instruments are exogenous I can run the following first-stage regression:

$$
\begin{aligned}
[Activities]_{it} = {} & [FE]_i + \textstyle\sum_m \beta_{D,m}[Month\ Dummy] \\
& + \beta_E[Expecting\ Harvest]_{it} + \beta_H[Had\ Harvest]_{it} \\
& + \beta_{RM}[Rice\ Farmer]_i \times [Mean]_t + \beta_{RV}[Rice\ Farmer]_i \times [Volatility]_t \\
& + \beta_{EM}[Expecting\ Harvest]_{it} \times [Mean]_t + \beta_{EV}[Expecting\ Harvest]_{it} \times [Volatility]_t \\
& + \beta_{HM}[Had\ Harvest]_{it} \times [Mean]_t + \beta_{HV}[Had\ Harvest]_{it} \times [Volatility]_t + \varepsilon_{it}.
\end{aligned} \tag{2}
$$

The second-stage regression excludes $[ExpectingHarvest] \times [Mean]$ and $[ExpectingHarvest] \times [Volatility]$ like so:

$$
\begin{aligned}
[Revenue]_{it} = {} & [FE]_i + \gamma_A[\widehat{Activities}]_{it} + \textstyle\sum_t \gamma_{D,m}[Month\ Dummy]_t \\
& + \gamma_E[Expecting\ Harvest]_{it} + \gamma_H[Had\ Harvest]_{it} \\
& + \gamma_{RM}[Rice\ Farmer]_i \times [Mean]_t + \gamma_{RV}[Rice\ Farmer]_i \times [Volatility]_t \\
& + \gamma_{HM}[Had\ Harvest]_{it} \times [Mean]_t + \gamma_{HV}[Had\ Harvest]_{it} \times [Volatility]_t + u_{it}.
\end{aligned} \tag{3}
$$

Test 3 predicts the coefficient on $[Activities]$ γ_A should be negative. The final test, Test 4, predicts the coefficient on $[Activities]$ κ_A in the simple OLS regression

$$
[Revenue]_{it} = \kappa_A[Activities]_{it} + \sum_t \kappa_{D,m}[Month\ Dummy]_t + \varepsilon_{it} \tag{4}
$$

should be biased upward relative to the IV regression.[11]

[11]Perhaps simple OLS is not the true empirical version of Test 4, but rather OLS that controls for everything in Equation 3 without instrumenting for the number of activities. Running this alternative regression does not change the outcome of the test.

4 Data

I build my sample using annual and monthly surveys collected by the Townsend Thai Project. In May of 1997 the Project surveyed over two thousand rural households in four provinces. The annual survey followed the households from one-third of the baseline districts up through 2010 (Townsend et al., 1997). The monthly survey followed the baseline households plus several new additions from four of the remaining districts (Townsend, 2012). The monthly survey records changes in household income, crop conditions, and many other features of the household. I combined the survey with the monthly international price of rice from January 1980 to June 2012, taken from the IMF's commodity price dataset.[12]

I use the monthly data to test the theory of risky income. My final sample contains all 743 households that responded to at least two of the seventy-two monthly rounds the project has released. Table 2 summarizes the sample characteristics. I observe the average household for 65 months, but have the full five years of data for over three-quarters of households. I mark a household to be a rice farmer if it harvests rice at any point in the sample. I mark a household as expecting a harvest if it harvests rice in the next three months; I mark it as having had a recent harvest if it harvested rice in the current month or the previous three months.[13] Table 2 shows that households expected a harvest one-fifth of the time. I define the number of economic activities as the sum of the number of "large" businesses, crop-plots cultivated, types of livestock raised, number of jobs held by all members, number of miscellaneous or small businesses, and an indicator for whether the household engages in aquaculture (raising fish or shrimp). I define total revenue as the sum of revenue from each economic activity. I define total consumption as total weekly and monthly household expenditure. Net transfers, which I use to classify households as insured in Section 3.1, are the total incoming transfers minus total outgoing transfers. I deflate

[12] I treat a household-month surveyed in the first half of the month as though observed in the previous month when I merge with time series data and define month dummies. Since the rice price and consumer price index are monthly averages, my convention best matches the survey response period to the horizons of the aggregate prices.

[13] Some fraction of households claim to sell rice during months when they still expect a harvest. In Appendix D I show that dropping these observations does not change the results.

revenue, consumption, and transfers to be in May 2007 Thai baht.[14] Despite its benefits the dataset has some limitations. The Townsend Thai Project has released only part of the monthly survey. I do not observe how much land or wealth a household owns; I do not observe all of its farm expenses; and I cannot link the monthly survey to the baseline survey collected in 1997.[15]

Table 2 shows that the average monthly revenue is 620 U.S. dollars per month at May 2007 exchange rates. This figure is skewed upward because revenue is bounded below by zero but spikes during rice harvests; hence the high standard deviation. Consumption is less seasonal and the mean of 194 dollars is less skewed.

[Table 2 around here]

I test the theory of lumpy investments with the annual panel. In addition to the four provinces and roughly 1000 households followed from baseline, the project added two more provinces and roughly 500 more households several years into the survey (both from the new provinces and from the original villages to counter attrition). My final annual sample for the lumpy investment tests is 1502 households. I construct the number of activities as closely as possible to my monthly measure: the sum of the number of large businesses, cropplots, jobs, herds, an indicator for aquaculture, and a subset of the miscellaneous income sources. [16] The annual average of 4.6 activities is almost identical to the monthly average in Table 2, but it varies less because the annual measure wipes out within-year variation in activities. Though this sample is technically different from the monthly sample, by the design of the survey it is nearly identical in location and characteristics. The main difference is that it contains more households, which if anything means my tests of the theory of lumpy investment should be more likely to yield significant results.

[Figures 3 and 4 around here.]

[14] For more details on how I construct the variables, see Appendix B.

[15] Since I do not observe expenses I cannot compute income, which would also be of interest. But since the short-run costs of under-specialization will likely be the opportunity cost of wasted time, revenue should capture most of the useful variation.

[16] Miscellaneous income sources in the annual survey often include remittances and other sources that do not meet my definition of economic activities (namely, revenue generating activities that require labor). I filter these unwanted sources using regular expressions on the textual descriptions of sources. The 1999 survey unfortunately does not contain textual descriptions, but the year dummies in the annual regressions should account for any 1999-specific measurement error.

The histogram in Figure 3, which shows the distribution of the number of activities in an arbitrary month, confirms that households in Thailand have many economic activities. Rice farmers are particularly under-specialized. Figure 4 graphs the top seven spontaneous responses to "What did your household do in the worst year [for income] of the last five to get by?" The most popular response was to take on an extra occupation, followed by working harder than usual. These responses do not prove households avoid risk through under-specialization, only that they cope with shocks through under-specialization. But if households must smooth their consumption by working harder, then they must have no better option. Borrowing money is only the third most popular response and using savings only the fifth. The fourth most popular response is to consume less, meaning many households lack even second-rate insurance.

[Figure 5 around here.]

Figure 5 shows the correlation between revenue and consumption, which is direct evidence of imperfect insurance. I compute the correlation between monthly revenue and consumption expenditure for each household over however many months I observe it (72 months for the majority). If a risk-averse household has perfect insurance its consumption should be independent of its current revenue; indeed, consumption should be constant. A household without perfect insurance cuts consumption when revenue falls, making the correlation positive. A higher correlation is evidence of less insurance. The figure plots the density of the correlation among rice farmers and non-rice farmers. Since zero is modal it appears many households do have near-perfect insurance, but many more do not. The distribution is heavily skewed towards less insurance with rice farmers particularly uninsured.[17] Some households have a negative correlation because of sampling error: the true correlation might be zero, but my estimate fluctuates around the truth and lands below zero for some households.

[17]The result may seem at odds with the high degree of insurance Townsend (1994) finds, but recall his result is that household consumption moves only with village-level and not household-level income. Figure 5 does not control for village-level shocks because a household cares only about having stable consumption, not where instability comes from. The shock I use for identification in Section 3 is a village-level shock: the international price of rice. It is precisely the village's inability to hedge against the price that drives households to under-specialize.

5 Risk and Under-Specialization Results

5.1 Main Results

Table 3 reports the results of the four tests derived in Section 2 and implemented as described in Section 3. Column 1 and 2 estimate (1) first without and then with month dummies. Column 3 estimates (4), and Column 4 estimates (3). Aside from the ordinary least squares regression reported in Column 3, all regressions use the generated measure of volatility. I calculate the p-values and confidence intervals of these regressions using a two-stage bootstrap.[18] The bootstrap, which I describe in detail in the online appendix, corrects for the generated volatility measure and within-household correlation in the error term across time.[19]

[Table 3 around here.]

The model's first test—Test 1—states that greater risk causes entry into more activities. The effect of risk on activities is the coefficient on $[ExpectingHarvest] \times [Volatility]$ in Column 1 of Table 3, and as predicted it is positive and significant. The model also predicts in Test 2 that higher expected returns to the primary activity (rice farming) should cause a decrease in activities. The coefficient on $[ExpectingHarvest] \times [Mean]$ confirms that higher returns have a negative and significant effect on the number of activities. Column 2 verifies that both results hold when I include month fixed-effects, though the estimate of the effect of risk on activities becomes smaller.

Test 3 states that if the extra activities cause (side) revenue to fall, then the failure to specialize is costly. I implement the test by running using the regression in Column 2 as a first-stage regression for (3) using the response of farmers expecting a harvest to expected price and volatility as instruments for the number of activities. Column 4 of Table 3 reports that the two-stage least squares coefficient on $[Activities]$ is negative and significant, confirming that under-specialization is costly. Column 3 reports the results of the simple ordi-

[18]It is not clear how to bootstrap the F-statistic on the excluded instruments or the Hansen's J Statistic. However, I can simply replace the generated volatility with the $|P_2 P_1|$ in the first stage. Since this is perfectly collinear with the generated measure it produces algebraically identical coefficients, but since it is not generated the standard F and J statistics are valid.

[19]I cluster by household rather than village because there is variation in harvest times within villages.

nary least squares regression with month dummies of revenue on number of activities (Equation 4). Test 4 states that the ordinary least squares coefficient on $[Activities]$ should be biased positively relative to the two-stage least squares coefficient because the farmers who pay lower costs for additional activities are exactly those who select into more activities. The coefficient is biased so strongly the sign flips, making under-specialization appear efficient.

What do the sizes of these coefficients mean? Since the average price volatility for all available months is 8.8, the regression in Column 1 implies a 10 percent rise in volatility causes the farmer to enter $.18/(1/8.8) * 10/100 = .16$ additional activities. A similar calculation shows that the more conservative estimate in Column 2 implies a 10 percent rise in volatility causes the farmer to enter .04 activities. Recall from Table 1, however, that the international price of rice is not perfectly correlated with the actual price the farmer receives. This may be because government price supports give the farmer some insurance. Regardless of the cause, since the international price has a correlation coefficient of roughly 1/3, a one unit rise in the volatility of the international rice price predicts a 1/3 unit rise in the volatility of the price the farmer receives. We can adjust the earlier numbers by dividing by 1/3, yielding estimates of .48 and .13 for the baseline and conservative estimates. The baseline estimate suggests the household enters an additional activity when local prices become 21 percent more volatile.

The two-stage least squares estimate implies a household will forego over 13 thousand baht, or over 60 percent of its average monthly revenue, in any year. According to the model in Section 2 this estimate is actually biased upward, suggesting the true cost is even higher than implied. But recall that the average household has a little over four activities at once, making an additional activity a very large increase. Further, if the cost of an activity varies across households the estimate is not the average cost. If there is an upper bound on the number of activities a household can juggle, then the households with fewer activities are those most likely to respond to the instruments. These are also the households for whom an extra activity is most costly. Then the estimate, which is the continuous equivalent of the local average treatment effect, might be higher than the average cost of a side activity.

The responses of households who had a recent harvest bear some explanation. First, the coefficient on $[RecentHarvest] \times [Mean]$ is negative. Since the expected price after the harvest is correlated with the price received at harvest, the negative coefficient confirms Figure 4 and the results of Adhvaryu et al. (2013), both of which say that households increase their number of activities in response to bad shocks. Finally, the positive and significant coefficient on $[RecentHarvest] \times [Volatility]$ seems puzzling, as risk should not matter after a household has had its harvest. There are two explanations for this. First, since the current volatility is correlated with past volatility, this may just reflect that the household faced risk before the harvest and took on extra activities. Since the household cannot drop the extra activities immediately after harvest—temporary jobs must be finished and small businesses must be wound down—the household may still have more activities than usual after harvest. The second possibility is that a high current volatility implies the price has moved drastically in the recent past. Since the current expected price ([Mean]) does not perfectly capture the price at harvest, a high volatility means it is more likely the household had a low price at harvest. Since households take on activities to recover from low prices, the coefficient on post-harvest volatility may be picking up the response to negative income shocks.

5.2 Robustness

Table 4 reports several robustness checks. The theory in Section 2 assumes the total labor supplied by the household is fixed, but in truth the household may work less when the returns to its labor grow riskier. Alternatively, the household might send some members to work abroad or in Bangkok. Columns 1 and 2 show that the effects of higher volatility and higher returns on the number of activities remain unchanged when I control for the household's total labor and the number of household members. Likewise, Column 6 shows that the effect of additional activities on revenue remains unchanged.

[Table 4 around here]

Columns 3 and 7 of Table 4 both answer a simple concern: should we believe Thai rice farmers use a model of autoregressive conditional heteroskedasticity to decide how to spend their time? The model only formalizes a simple intuition: when prices fluctuate they are risky. Columns 3 and 7 confirm that simpler measures of the mean and volatility—the current price and the absolute value of the change in the price since last month—do not much change the results.

If the volatility of the price is just a proxy for unexpected decreases in the price, then what I assume is a response to risk may in truth be a response to changes in the household's permanent income. If this story is true, then the household should respond more strongly to simple changes in the price than to my measure of volatility, which is proportional to the absolute value of the change. Column 4 of Table 4 runs a regression that replaces my measure of volatility with the simple change in price. Households expecting a harvest do not respond to simple changes in the price.

The reader may also worry whether the expected price and the volatility are valid instruments for side revenue. If the price of rice and the price of corn, say, are correlated then the expected price is no longer a valid instrument for rice farmers who also grow corn. Column 5 of Table 4 verifies that the second stage results hold when I use a measure of revenue that excludes earnings from crops. In Appendix D I show that median wages in each village are not affected by average movements in the regressor of interest.

[Table 5 around here]

I measure "specialization" with the number of economic activities, but this measure may seem arbitrary. Table 5 tests whether two alternate measures of specialization respond to risk. The first measure is an indicator for whether anyone in the household holds an unsteady job, which I define as a job that lasts for five months or fewer. If volatile prices drive a household into casual labor it is another sign that risk causes inefficient under-specialization. The second measure is the number of non-crop economic activities. Since these households farm rice it is a sign of under-specialization if they expand their activities beyond the fields.

Columns 1 and 3 show that farmers who expect a harvest are more likely to get unsteady jobs and will take on more non-crop activities when rice prices grow more volatile. Columns 2 and 4 show that the effects grow smaller and insignificant when I include month fixed-effects. This may be because month fixed-effects absorb much of the variation in $[Expecting\ Harvest] \times [Volatility]$. Since I must adjust the standard errors for generated regressors the reduced variation makes it hard to find effects. But the regressions provide some suggestive evidence to support the main results.

[Table 6 around here]

Is what I measure really a response to risk? To answer this question I examine whether households with better insurance respond less to changes in the volatility. In poor countries a household often relies on family and friends for support in hard times. [20] Figure 10 shows that the rice farmers in my sample are no different. When the international price is low rice farmers tend to receive more transfers. I calculate for each household the monthly correlation between its net incoming transfers and its revenue, and call a household "insured" if that correlation is negative.

Columns 1 and 2 of Table 6 report the separate responses of the uninsured and insured sample. As expected, the response of the insured is smaller and insignificant. Since my measure of insurance is not exogenous I cannot rule out that households with insurance differ from uninsured households in ways that change how they might respond to volatility. Still, though not a perfect test of the model the result is consistent with the model.

6 The Alternative Theory: Lumpy Investments

If "the poor cannot raise the capital they would need to run a business that would occupy them fully" (Banerjee and Duflo, 2007) then poor households cannot specialize. Suppose a man can learn to sew or bake but cannot can sew more than a few shirts or bake more than a few loaves unless he buys a sewing machine or an oven. Since he cannot afford either investment he cannot grow either business. To support his family he must sell both shirts and bread. This is the theory of lumpy investments.[21]

[20] Rosenzweig (1988) found that households structure themselves to ease income sharing. Townsend (1994) and more recently Munshi and Rosenzweig (2009) find village and caste networks provide insurance in India. Yang and Choi (2007) show that rural Filipino households who suffer bad rainfall shocks receive more remittances from overseas family.

[21] The inability to invest may create another source of under-specialization: the need to take on extra jobs because one may only work so long at any single task. Suppose labor and capital are complements, and make it simple with an extreme example: perfect complementarity. Suppose an activity m produces revenue with production function $y^m = A^m \min[L, K]$, with $m = T, B$ for tailoring or baking. Suppose $A^T > A^B$ for some household. If the house-

To test the theory I exploit a government program that produced quasi-experimental variation in the supply of credit. The theory predicts that households that get more credit should be better able to make the lumpy investments that let them specialize. The Million Baht Program gave one million baht to a fund for public lending set up in every village in my sample. Kaboski and Townsend (2009, 2011), who are the first to exploit the program, argue that the boundaries of villages in Thailand are set by bureaucratic fiat rather than economic logic. The sizes of villages are effectively random. Since every village got the same amount of credit the per-household increase in credit is also random, with smaller villages exogenously given more credit. Kaboski and Townsend (2011) confirm in their first table that small and large villages have parallel trends.

Since I do not know when in 2001 the program reached each village, I use the annual data and treat 2001 as the year of implementation. Given that the annual data follow a nearly identical population, it is reasonable to compare the results between samples.[22] The effect of the program is measured by the interaction of the year of implementation interacted with some measure of village size. In one specification I use an indicator for whether the village is in the bottom size quartile; in the other I use the actual per-household injection (1 million/number of households).[23] The theory predicts that the signs of the coefficients should be negative and significant.

[Table 7 around here]

According to Table 7 the coefficients are insignificant and have the wrong sign. The positive coefficients on $2001 \times Small$ and the other interactions are not consistent with the lumpy investment theory, but might be consistent with the model from Section 2. If risk is really what drives under-specialization and some households want more activities but cannot afford to pay the fixed-cost, giving them credit might let them enter more activities. But this story, which lacks

hold's labor endowment is $\bar{L}$, it will specialize in tailoring with $K^* = \bar{L}$. But suppose increasing capital beyond $\tilde{K} < K^*$ requires a lumpy investment the household cannot afford. If the household specialized, it would be left with $\bar{L} - \tilde{K}$ units of unused labor. In other words, it would be idle. The alternative is to spend its remaining time baking, so its total revenue is $A^T\tilde{K} + A^B(\bar{L} - \tilde{K}) < A^T\bar{L}$.

[22]Since the annual data cover more households than the monthly data, if anything I stack the odds towards finding statistically significant evidence for the theory of lumpy investment versus the theory of risky income.

[23]The results do not change when I use the log of the injection as in Shenoy (2014).

direct evidence and rests on coefficients that are not signficant, remains only a story. Only the coefficient on $2002 \times Credit/HH$ in Column 2 is (marginally) significant, and even that significance vanishes when I restrict the estimation in Column 4 to a balanced panel. All told the data do not support the theory.

My results do not mean credit constraints have no effect on specialization. As always, a lack of evidence is not a rejection. But given that I did find statistically significant evidence for the theory of risky income in a nearly identical but smaller sample, the lack of evidence is suggestive. Likewise, though a lack of evidence for the theory in Thailand does not mean I would find no evidence in other countries, my results are supported by other studies. Bianchi and Bobba (2012), for example, find that the Progresa conditional cash transfer helped households start businesses by mitigating risk rather than alleviating a lack of funds. Likewise, Karlan et al. (2012a) find that giving small firms extra capital does not cause them to grow.

The bigger concern is that I test only a limited form of the theory. The smallest villages received per-household credit injections of half the median income. If households need sewing machines the credit injection could pay for them. Since most micro credit charities believe small entrepreurs need small loans, finding no effect from a small rise in the supply of credit is not meaningless. But if a few households want to build factories that would provide a single salaried job to everyone else, the Million Baht Program is too small to finance the transformation.

7 Summary

I show that Thai rice farmers expecting a harvest increase their number of economic activities when confronted with more volatile prices. My estimates suggest a 21 percent rise in volatility causes a household to enter 1 extra activity. I use this exogenous change in the number of activities to verify that underspecialization reduces revenue. Finally, I test an alternative theory of underspecialization—that the poor run many small businesses because they cannot afford the lumpy investments needed to grow any one—and find no supporting evidence.

The pin-maker wastes time when he switches from straightening wires to cutting them, and I find evidence of this waste in rural Thailand. My results do not measure the talent wasted when the poor forego expertise in a single trade. This kind of under-specialization, which changes the structure of an economy, is a long-run cost that requires a long-run study. Future research must test whether long-run risk causes long-run under-specialization.

References

ADHVARYU, A., N. KALA, AND A. NYSHADHAM (2013): "Booms, Busts, and Household Enterprise: Evidence from Coffee Farmers in Tanzania," .

BANDYOPADHYAY, S. AND E. SKOUFIAS (2012): "Rainfall Variability, Occupational Choice, and Welfare in Rural Bangladesh," *World Bank Policy Research Working Paper*.

BANERJEE, A. AND E. DUFLO (2007): "The Economic Lives of the Poor," *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 21, 141.

BIANCHI, M. AND M. BOBBA (2012): "Liquidity, Risk, and Occupational Choices," *The Review of Economic Studies*, rds031.

BLISS, C. J. AND N. H. STERN (1982): *Palanpur: The Economy of an Indian Village*, Oxford University Press.

COLE, S., X. GINÉ, AND J. VICKERY (2013): "How Does Risk Management Influence Production Decisions? Evidence from a Field Experiment," .

EMERICK, K., A. DE JANVRY, E. SADOULET, AND M. H. DAR (2014): "Risk and the Modernization of Agriculture," .

ENGLE, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica: Journal of the Econometric Society*, 987–1007.

GUMMERT, M. AND J. RICKMAN (2011): "When to Harvest," Tech. rep., International Rice Research Institute.

KABOSKI, J. AND R. TOWNSEND (2009): "The Impacts of Credit on Village Economies," *MIT Department of Economics Working Paper No. 09-13*.

——— (2011): "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative," *Econometrica*, 79, 1357–1406.

KARLAN, D., R. KNIGHT, AND C. UDRY (2012a): "Hoping to Win, Expected to Lose: Theory and Lessons on Micro Enterprise Development," .

KARLAN, D., R. D. OSEI, I. OSEI-AKOTO, AND C. UDRY (2012b): "Agricultural Decisions after Relaxing Credit and Risk Constraints," .

MORDUCH, J. (1990): "Risk, Production and Saving: Theory and Evidence from Indian Households," *Harvard University, Manuscript*.

MUNSHI, K. AND M. ROSENZWEIG (2009): "Why Is Mobility in India So Low? Social Insurance, Inequality, and Growth," Tech. rep., National Bureau of Economic Research.

MURPHY, K. AND R. TOPEL (2002): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business & Economic Statistics*, 20, 88–97.

PAGAN, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247.

ROSENZWEIG, M. AND H. BINSWANGER (1993): "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," *The Economic Journal*, 103, 56–78.

ROSENZWEIG, M. R. (1988): "Risk, Implicit Contracts and the Family in Rural Areas of Low-Income Countries," *The Economic Journal*, 98, 1148–1170.

ROSENZWEIG, M. R. AND O. STARK (1989): "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India," *The Journal of Political Economy*, 905–926.

SHENOY, A. (2014): "Market Failures and Misallocation: Separating the Costs of Factor and Financial Market Failures," .

TOWNSEND, R. (1994): "Risk and Insurance in Village India," *Econometrica: Journal of the Econometric Society*, 539–591.

TOWNSEND, R., A. PAULSON, S. SAKUNTASATHIEN, T. J. LEE, AND M. BINFORD (1997): "Questionnaire Design and Data Collection for NICHD Grant *Risk, Insurance and the Family* and NSF Grants," .

TOWNSEND, R. M. (2012): "Townsend Thai Project Monthly Survey (1-60) Initial Release," Murray Research Archive [Distributor] V1 [Version].

WALKER, T. S. AND J. G. RYAN (1990): *Village and Household Economics in India's Semi-Arid Tropics*, Johns Hopkins University Press.

YANG, D. AND H. CHOI (2007): "Are Remittances Insurance? Evidence from Rainfall Shocks in the Philippines," *The World Bank Economic Review*, 21, 219–248.

A Proofs

A.1 Generalizing the Risk and Return Predictions

Letting $M \in \{1, 2, \ldots\}$, the optimal labor allocation is

$$L_p = \frac{\bar{w}_+ M + \alpha\sigma_s^2}{\alpha\left(M\sigma_p^2 + \sigma_s^2\right)} \tag{5}$$

Consider the threshold fixed cost that separates households who choose M activities from those who choose $M + 1$ activities:

$$-e^{-\alpha\bar{C}(M)+\frac{\alpha^2}{2}V(M)} = -e^{-\alpha\bar{C}(M+1)+\frac{\alpha^2}{2}V(M+1)}$$

The threshold is

$$\bar{F}_M = \frac{\sigma_s^2\left(\alpha\sigma_p^2 - \bar{w}_+\right)^2}{2\alpha\left(M\sigma_p^2 + \sigma_s^2\right)\left((M+1)\sigma_p^2 + \sigma_s^2\right)}.$$

The derivatives with respect to σ_p^2 and $\bar{w}_p$ are

$$\frac{\partial \bar{F}_M}{\partial \sigma_p^2} = \frac{\sigma_p \sigma_s^2 \left(\alpha \sigma_p^2 - \bar{w}_+\right) \left(\alpha \sigma_s^2 \left((2M+1)\sigma_p^2 + 2\sigma_s^2\right) + \bar{w}_+ \left(2M \left((M+1)\sigma_p^2 + \sigma_s^2\right) + \sigma_s^2\right)\right)}{\alpha \left(M\sigma_p^2 + \sigma_s^2\right)^2 \left((M+1)\sigma_p^2 + \sigma_s^2\right)^2}$$

$$\frac{\partial \bar{F}_M}{\partial \bar{w}_p} = -\frac{\sigma_s^2 \left(\alpha \sigma_p^2 - \bar{w}_+\right)}{\alpha \left(M\sigma_p^2 + \sigma_s^2\right) \left((M+1)\sigma_p^2 + \sigma_s^2\right)}$$

Since by assumption $\left(\alpha \sigma_p^2 - \bar{w}_+\right) > 0$, $\frac{\partial \bar{F}_M}{\partial \sigma_p^2} > 0$ and $\frac{\partial \bar{F}_M}{\partial \bar{w}_p} < 0$ for all M. Then a rise in the riskiness of the primary activity will cause all the thresholds to rise, meaning households will be willing to pay more for any number of activities. This will cause the average number of activities in the sample to rise. A parralel argument shows a rise in the average return decreases all thresholds and decreases the average number of activities.

QED

A.2 Verifying the Cost Prediction

We can rewrite expected side income as

$$\begin{aligned}
\mathbb{E}[y_s] &\approx \bar{w}_s - F\left(\frac{\partial M}{\partial \sigma_p^2}\sigma_p^2 + \frac{\partial M}{\partial \bar{w}_p}\bar{w}_p\right) + \bar{w}_s\left(-\frac{\partial L_p}{\partial M}\left[\frac{\partial M}{\partial \sigma_p^2}\sigma_p^2 + \frac{\partial M}{\partial \bar{w}_p}\bar{w}_p\right] - \frac{\partial L_p}{\partial \sigma_p^2}\sigma_p^2 - \frac{\partial L_p}{\partial \bar{w}_p}\bar{w}_p\right) \\
&= \bar{w}_s + \left(-F - \bar{w}_s\frac{\partial L_p}{\partial M}\right)\left[\frac{\partial M}{\partial \sigma_p^2}\sigma_p^2 + \frac{\partial M}{\partial \bar{w}_p}\bar{w}_p\right] + \bar{w}_s\left[-\frac{\partial L_p}{\partial \sigma_p^2}\sigma_p^2 - \frac{\partial L_p}{\partial \bar{w}_p}\bar{w}_p\right] \\
&= \bar{w}_s + \left(-F - \bar{w}_s\frac{\partial L_p}{\partial M}\right)\hat{M} + \bar{w}_s\left[-\frac{\partial L_p}{\partial \sigma_p^2}\sigma_p^2 - \frac{\partial L_p}{\partial \bar{w}_p}\bar{w}_p\right] \\
\rightarrow y_s &= \bar{w}_s + \left(-F - \bar{w}_s\frac{\partial L_p}{\partial M}\right)\hat{M} + \eta + \varepsilon
\end{aligned}$$

where $\hat{M}$ is the predicted number of activities from the first-stage regression, η is the direct effect of labor reallocation from changes in the volatility and average returns to the primary activity, and ε is an independent error term.

The instrumental variables estimate is consistent for the value

$$\gamma_A = -F - \bar{w}_s\frac{\partial L_p}{\partial M} + \mathbb{E}[\hat{M}\gamma]$$

If $-\bar{w}_s \frac{\partial L_p}{\partial M} + \mathbb{E}[\hat{M}\gamma] > 0$, then $\gamma_A > -F$, which implies that if $\gamma_A < 0$ then $-F < 0$ and thus under-specialization is costly.

First I show that $\frac{\partial L_p}{\partial M} < 0$. From the expression for L_p found in (5) in Appendix A.1 we have that a 1 unit increase in M will cause a rise in the numerator of $\bar{w}_+$ and a rise in the denominator of $\alpha\sigma_p^2$. Since by assumption $\alpha\sigma_p^2 > \bar{w}_+$, the denominator rises by more than the numerator and the total effect is negative. Thus, $\frac{\partial L_p}{\partial M} < 0$.

Now I show that $\mathbb{E}[\hat{M}\gamma] > 0$. The expectation equals

$$
\begin{aligned}
\mathbb{E}[\hat{M}\gamma] &= +\bar{w}_s \left[-\frac{\partial L_p}{\partial \sigma_p^2}\mathbb{E}[\hat{M}\sigma_p^2] - \frac{\partial L_p}{\partial \bar{w}_p}\mathbb{E}[\hat{M}\bar{w}_p] \right] \\
&= -\bar{w}_s [\underbrace{\frac{\partial L_p}{\partial \sigma_p^2}}_{-} \underbrace{\frac{\partial M}{\sigma_p^2}}_{+} + \underbrace{\frac{\partial L_p}{\partial \bar{w}_p}}_{+} \underbrace{\frac{\partial M}{\bar{w}_p}}_{-}]
\end{aligned}
$$

where the final equality applies the definition of $\hat{M}$; applies the predictions of the effects of risk and returns to the number of activities to get the signs of $\frac{\partial M}{\sigma_p^2}$ and $\frac{\partial M}{\bar{w}_p}$; and takes the deriviatives of L_p found in (5) in Appendix A.1 with respect to σ_p^2 and $\bar{w}_p$. This proves that $\mathbb{E}[\hat{M}\gamma] > 0$.

QED

Table 1
Rice Prices and Sales

	(1) Avg. Transaction Price	(2) Rice Sold
Int. Rice Price	0.333**	
	(0.14)	
Rice Harvested		0.856***
		(0.01)
Constant	1.500	-2043.744***
	(1.53)	(70.44)
N	62	2126

Note: **Column 1 —** The dependent variable is the sample-wide average price of a kilogram of rice based on actual transactions, and the independent variable is the international price of rice in baht per kilogram. Not all survey rounds include any sales of rice—hence the number of observations is smaller than the number of survey rounds. Standard errors are robust to heteroskedasticity. **Column 2 —** The unit of observation is the household-month conditional on positive rice harvest.

Table 2
Descriptives of the Monthly Sample

Household-Month Mean and Standard Deviation				Fraction of Households or Household-Months	
Number of activities:	4.6	Revenue:	21352.8	Rice Farmers:	0.48
	(3.3)		(79854.7)		
				Of Whom	
Household size:	5.3	Consumption:	6692.2	Fraction of time expecting harvest:	0.23
	(2.4)		(24449.2)		
Total Labor:	80.0	Net Transfers In:	667.0	Fraction of time just had harvest:	0.31
	(75.6)		(35274.8)		
Households:	743	Avg. Obs/HH:	65.0	Observations:	48329

Table 3
Main Results

	(1) Activities	(2) Activities	(3) Revenue	(4) Revenue
Activities			1851.26***	-13883.30**
			[0.000]	[0.035]
Mean	-0.00*			
	[0.096]			
Volatility	-0.08***			
	[0.000]			
Rice Farmer				
- × Mean	0.01***	0.00		-128.97
	[0.010]	[0.359]		[0.135]
- × Volatility	-0.20***	-0.10***		-272.90
	[0.002]	[0.009]		[0.618]
Expecting Harvest				
- Main	1.82***	1.89***		4993.52
	[0.006]	[0.000]		[0.310]
- × Mean	-0.02***	-0.02***		(Exc. Inst.)
	[0.000]	[0.000]		
- × Volatility	0.18***	0.05*		(Exc. Inst.)
	[0.001]	[0.089]		
Recent Harvest				
- Main	-0.76	-0.57		-34753.04**
	[0.475]	[0.303]		[0.034]
- × Mean	-0.03***	-0.01***		300.00
	[0.000]	[0.002]		[0.129]
- × Volatility	0.41***	0.17***		234.48
	[0.002]	[0.004]		[0.819]
Household Fixed-Effects	Yes	Yes	No	Yes
Month Fixed-Effects	No	Yes	Yes	Yes
F-Stat Exc. Inst.				13.604
Hansen's J Stat.				0.125
Households	743	743	743	743
Observations	48329	48329	48329	48329

Note: These regressions run the four tests of the theory of risky income (see Section 2): Test 1 (Risk): risk increases the number of activities; Test 2 (Returns): higher returns decrease the number of activities; Test 3 (Cost): more activities may cause side revneue to fall; Test 4 (OLS Bias): OLS is biased upwards. Column 1 estimates Equation 1, Column 2 estimates Equation 2, Column 3 estimates Equation 4, and Column 4 estimates Equation 3. The bracketed values are p-values. I compute the p-values in Columns 1,2 and 4 using a two-stage bootstrap that corrects for generated regressors and clusters by household (see Appendix C). I compute the p-values in Column 3 using asymptotic standard errors that cluster by household. The value of the F-statistic on the excluded instruments from the first stage meets common standards for strength. The value of the J-statistic for overidentification is much too small to reject the null of exogenous instruments.

Table 4
Robustness

	(1) Activities (Labor)	(2) Activities (HH Size)	(3) Activities (Non-Gen)	(4) Activities (Price Change)	(5) Revenue (Exc. Crop Rev.)	(6) Revenue (Labor & Size)	(7) Revenue (Non-Gen)
Activities					-11232.10**	-17060.61***	-9305.78*
					[0.031]	[0.009]	[0.064]
Rice Farmer							
- × Mean	0.00	0.00	0.01***	0.00	-112.59	-91.88	-21.25
	[0.249]	[0.441]	[0.001]	[0.854]	[0.178]	[0.278]	[0.795]
- × Volatility	-0.10***	-0.10**	-0.05***	0.02***	-493.25	-791.17	-397.12
	[0.006]	[0.013]	[0.000]	[0.000]	[0.378]	[0.265]	[0.120]
Expecting Harvest							
- Main	1.78***	1.86***	2.58***	2.09***	7295.65	4897.58	1537.60
	[0.000]	[0.000]	[0.000]	[0.000]	[0.118]	[0.345]	[0.725]
- × Mean	-0.01***	-0.01***	-0.02***	-0.01***	(Exc. Inst.)	(Exc. Inst.)	(Exc. Inst.)
	[0.000]	[0.000]	[0.000]	[0.000]			
- × Volatility	0.05*	0.05*	0.02***	-0.00	(Exc. Inst.)	(Exc. Inst.)	(Exc. Inst.)
	[0.094]	[0.088]	[0.003]	[0.881]			
Recent Harvest							
- Main	-0.54	-0.61	1.30***	0.24	-28550.62*	-34442.87**	148.02
	[0.306]	[0.281]	[0.000]	[0.423]	[0.052]	[0.030]	[0.992]
- × Mean	-0.01***	-0.01***	-0.02***	-0.01*	167.55	288.46	-44.62
	[0.002]	[0.005]	[0.000]	[0.062]	[0.381]	[0.129]	[0.758]
- × Volatility	0.16***	0.17***	0.05***	-0.04***	1018.61	161.63	965.18**
	[0.003]	[0.002]	[0.000]	[0.000]	[0.390]	[0.878]	[0.012]
Household Labor	0.01***					275.42***	
	[0.000]					[0.000]	
Household Size		0.13***				950.03	
		[0.000]				[0.312]	
Household Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Stat Exc. Inst.					13.604	13.154	20.660
Hansen's J Stat.					0.018	0.273	0.369
Households	743	743	743	743	743	743	743
Observations	48160	48164	48329	48329	48329	48160	48329

Note: Columns 1 and 2 verify that controlling for household labor and household size do not change the effect that the mean and volatility of rice prices have on the number of activities. Column 6 confirms that controlling for both does not change the effect of activities on revenue. Columns 3 and 7 verify that a simpler measure of volatility–the absolute change in rice prices–does not qualitatively change the results for either number of activities or revenue. Column 4 verifies that the effect is caused by volatility rather than the (non-absolute) change in prices, as would be so if the effect were driven by a change in permanent income rather than volatility. Column 5 confirms that even excluding income from crop-plots the revenue results holds. The bracketed values are p-values. I compute the p-values in Columns 3, 4, and 7 using asymptotic standard errors clustered at the household level. I compute all other p-values using a two-stage bootstrap that corrects for generated regressors and clusters by household (see Appendix C).

Table 5
Check: Other Measures

	Have Unsteady Job		Non-Crop Activities	
	(1)	(2)	(3)	(4)
Mean	0.00***		-0.00*	
	[0.000]		[0.076]	
Volatility	-0.01***		-0.01	
	[0.007]		[0.302]	
Rice Farmer				
- $\times$ Mean	-0.00*	-0.00***	-0.00	-0.00
	[0.058]	[0.000]	[0.336]	[0.113]
- $\times$ Volatility	0.01	0.01**	-0.02	0.01
	[0.244]	[0.040]	[0.173]	[0.726]
Expecting Harvest				
- Main	-0.07	-0.03	0.11	0.29
	[0.369]	[0.661]	[0.630]	[0.113]
- $\times$ Mean	-0.00**	-0.00	-0.01***	-0.00***
	[0.020]	[0.150]	[0.000]	[0.001]
- $\times$ Volatility	0.02***	0.01	0.06***	0.02
	[0.009]	[0.111]	[0.010]	[0.278]
Recent Harvest				
- Main	0.22***	0.10	0.80***	0.64***
	[0.001]	[0.196]	[0.000]	[0.002]
- $\times$ Mean	-0.00***	-0.00	-0.01***	-0.01**
	[0.000]	[0.834]	[0.000]	[0.011]
- $\times$ Volatility	0.01*	-0.01	0.03*	-0.01
	[0.060]	[0.134]	[0.072]	[0.479]
Household Fixed-Effects	Yes	Yes	Yes	Yes
Month Fixed-Effects	No	Yes	No	Yes
Households	743	743	743	743
Observations	48329	48329	48329	48329

Note: I define an "unsteady job" as one held for less than five months. Columns 1 and 2 look for effects of volatility on an indicator for whether anyone in the household had an unsteady job. I define "non-crop activities" as the total number of activities minus the number of crop-plots farmed. Columns 3 and 4 look for effects on the number of non-crop activities. I present all coefficients with p-values calculated using the two-stage bootstrap.

Table 6
Check: Insurance

	(1) Activities (Uninsured)	(2) Activities (Insured)
Rice Farmer		
- × Mean	0.01**	-0.00
	[0.027]	[0.922]
- × Volatility	-0.09*	-0.10**
	[0.069]	[0.013]
Expecting Harvest		
- Main	1.65***	1.99***
	[0.001]	[0.000]
- × Mean	-0.02***	-0.02***
	[0.000]	[0.000]
- × Volatility	0.09*	0.04
	[0.067]	[0.277]
Recent Harvest		
- Main	-0.77	-0.39
	[0.186]	[0.557]
- × Mean	-0.01	-0.01***
	[0.221]	[0.006]
- × Volatility	0.14**	0.19***
	[0.025]	[0.003]
Household Fixed-Effects	Yes	Yes
Month Fixed-Effects	Yes	Yes
Households	270	473
Observations	16933	31396

Note: I split the sample into households who receive transfers of income when their consumption is low ("insured") and those that do not ("uninsured"). I confirm that volatility has a larger effect on the number of activities among households that are uninsured.

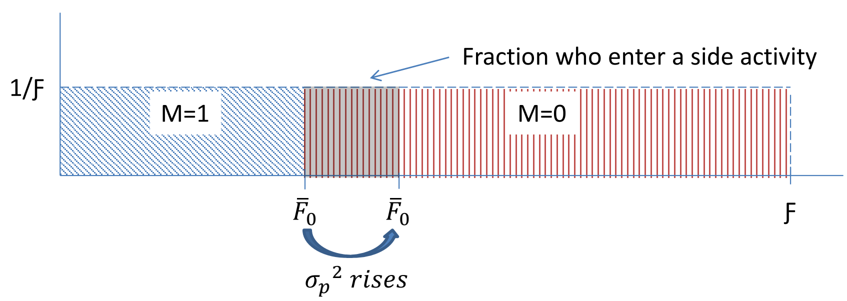
Table 7
Testing the Theory of Lumpy Investments

	(1) Activities b/se	(2) Activities b/se	(3) Activities b/se	(4) Activities b/se
Small Village	-0.010		0.102	
	(0.11)		(0.11)	
2001 X Small	0.132		0.175	
	(0.16)		(0.16)	
2002 X Small	0.213		0.144	
	(0.14)		(0.15)	
Credit/HH		3.010		9.977
		(13.47)		(13.28)
2001 X Credit/HH		11.540		11.569
		(8.63)		(8.20)
2002 X Credit/HH		21.857*		16.619
		(11.55)		(11.86)
Household Fixed-Effects	Yes	Yes	Yes	Yes
Month Fixed-Effects	Yes	Yes	Yes	Yes
Villages	80	80	64	64
Households	1502	1502	706	706
Observations	15340	15340	9884	9884

All standard errors clustered by village

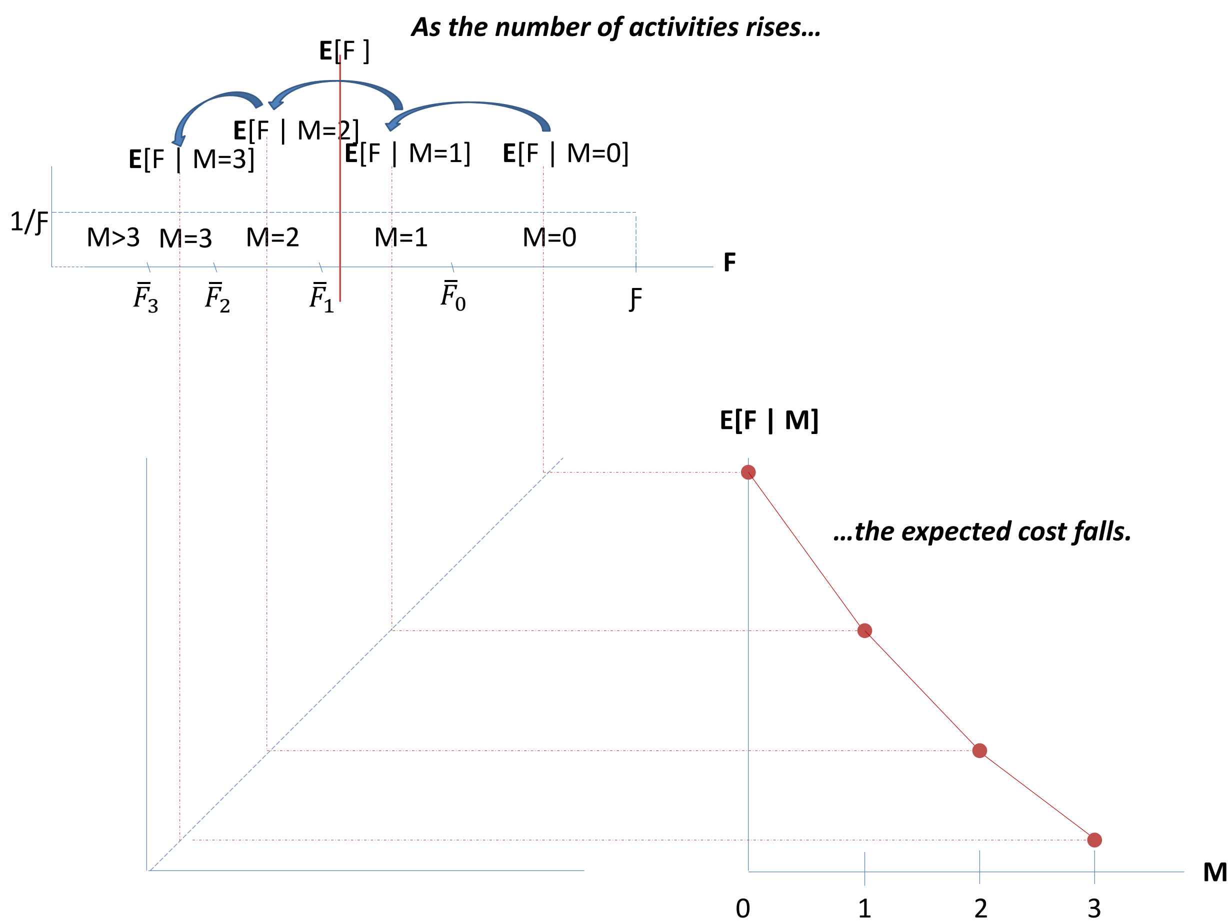
Note: The regressions test the lumpy investment theory using the Million Baht Program. The coefficient on the interaction of village size with the year of implementation (2001) estimates the effect of relaxed credit on number of activities. The measure of number of activities is similar as possible to that in the risk regressions. A village in the bottom quartile of number of households is "small". The alternative specification uses the average per-household credit injection (one million divided by number of households). The first two columns use the largest possible sample of households while the last two use a balanced panel. The lumpy investment theory predicts the program's impact should be negative and significant, which it is not. All inference uses asymptotic standard errors clustered at the village.

Figure 1
Intuition of the Simplified Case



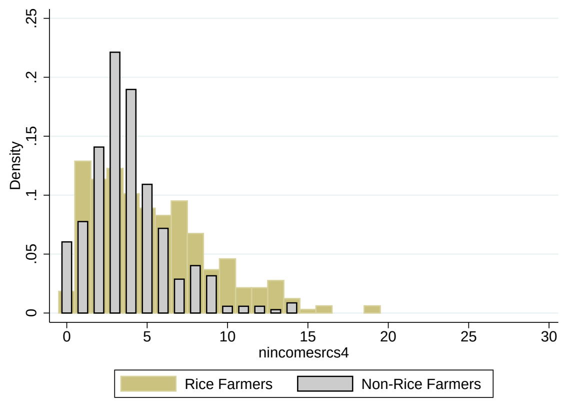
Note: M is the number of side activities; $\bar{F}_0$ the threshold fixed cost for moving from zero to one side activity; σ_p^2 is the variance of the primary economic activity. A rise in the variance raises the threshold fixed cost, which represents the amount households are willing to pay for insurance. In response the highlighted mass of households switches from specialization to having a side activity.

Figure 2
Why is OLS Upward-Biased?



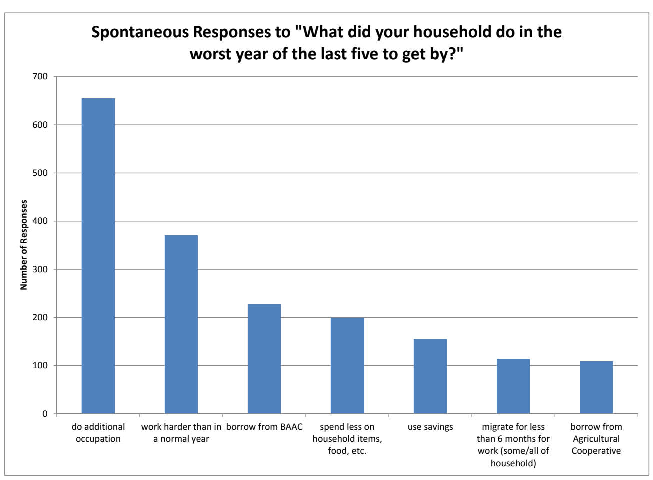
Note: M is the number of side activities; $\bar{F}_m$ is the threshold fixed cost below which a household moves from m to $m+1$ side activities. A household only enters many activities if the fixed cost it must pay for each is low. Then the number of activities predicts a household's costs. The cost of these extra activities appears in the error term of a regression of side revenue on number of activities. Thus the coefficient on the number of activities is biased upwards.

Figure 3
Number of Economic Activities, Rice Farmers and Non-Rice Farmers



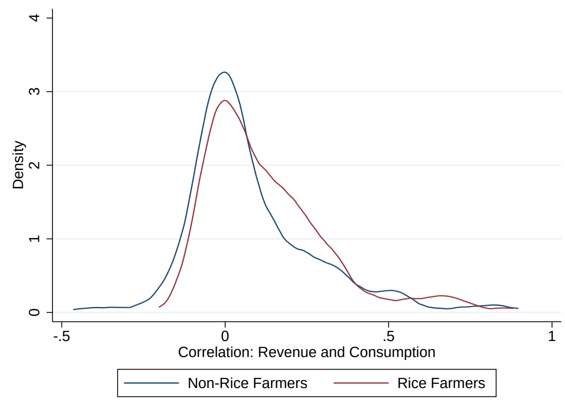
Note: The histogram shows the fraction of households with any number of economic activities in an arbitrarily chosen month. Rice farmers are more likely to have many activities.

Figure 4
Household Response to Negative Income Shocks



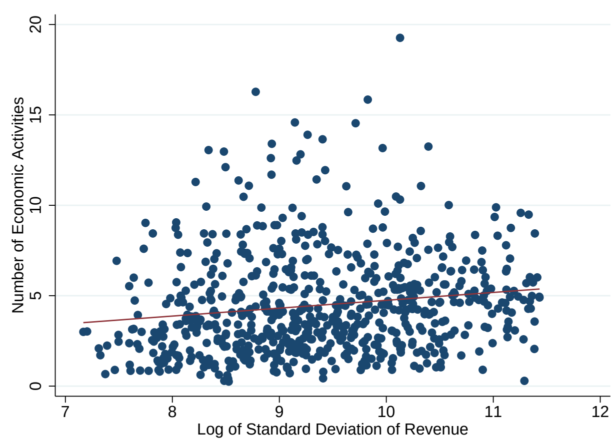
Note: The 1997 round of the Townsend Thai annual survey asks households how they coped during the the worst income year of the last five. They first gave spontaneous responses, which the project classified into categories. The graph reports the frequencies of the seven most popular responses. Many households work more or spend less to absorb income shocks rather than borrowing or using savings.

Figure 5
Correlation Between Monthly Revenue and Consumption



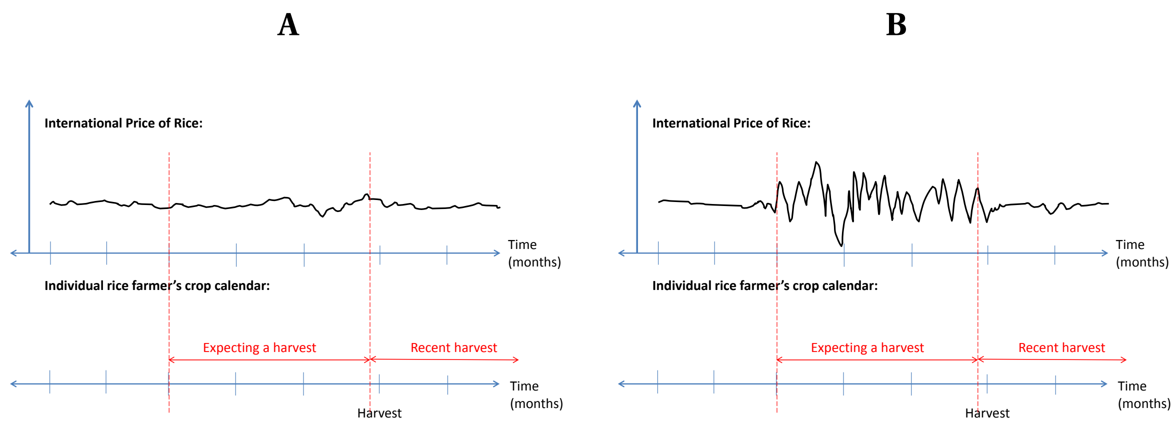
Note: For each household I compute the monthly correlation between total consumption and total revenue. I plot the density of the correlation for rice farmers versus non-rice farmers. Perfect insurance (whether self-insurance or otherwise) implies zero correlation. Almost all households have a positive correlation, meaning they consume less when their revenue falls.

Figure 6
Riskiness of Revenue and Number of Economic Activities



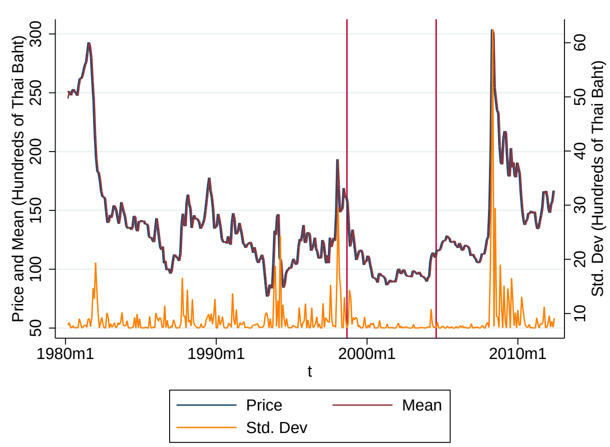
Note: I plot the log of the standard deviation of the household's monthly revenue against the average number of activities during the sample. I exclude the top and bottom 5 percent of standard deviations. Households with riskier revenue have more economic activities.

Figure 7
Response to Conditional Volatility



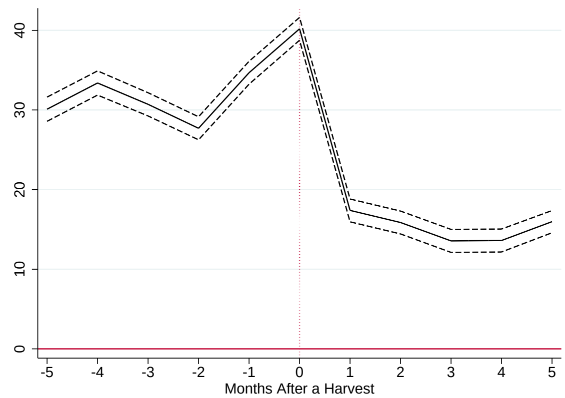
Note: Among rice farmers expecting a harvest I compare the response when rice prices are (**A**) stable to when they are (**B**) volatile. Since I use household fixed-effects I effectively compare each farmer to himself.

Figure 8
Rice Price and Predicted Mean and Conditional Standard Deviation



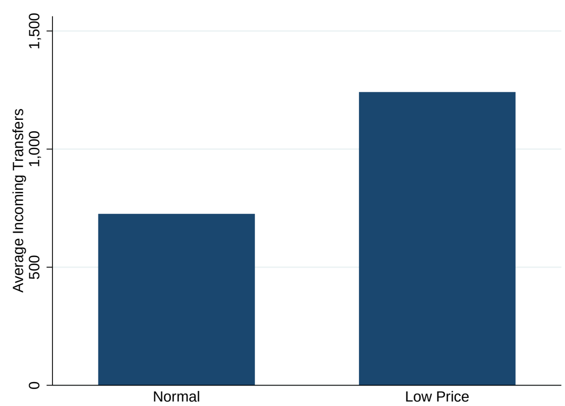
Note: I plot the actual rice price next to both the predicted rice price and the predicted volatility (square root of the predicted conditional variance) from the Autoregressive Conditional Heteroskedasticity (ARCH) model. The red lines mark the start and end of the panel data.

Figure 9
An Impending Rice Harvest Requires Labor



Note: The figure shows how many days the average household works in its fields in the months before and after a rice harvest. More precisely, I plot the coefficients of a regression of the number of days worked in the fields on dummies for periods before and after the harvest. The dashed lines cover 95 percent confidence intervals.

Figure 10
Households Receive More Transfers when Prices are Low at Harvest



Note: The first bar depicts average incoming transfers for households harvesting rice when the international rice price is "normal"—above the bottom quartile of all prices I observe in the period covered by the monthly panel. The second bar depicts the average transfers when prices are "low"—in the bottom quartile. Rice farmers receive more money when the value of their harvest is low.

B Detailed Data Appendix (For Online Publication)

B.1 Time Series Variables

- **Consumer Prices:** From Bank of Thailand monthly index, acquired from Global Financial Data database. Data were used with permission of Global Financial Data.

- **International Rice Price:** Acquired from IMF monthly commodity price data. Deflated using monthly consumer price index.

B.2 Panel Variables

- **Rice Harvest:** From module 7 (Crop Harvest) section of the monthly survey. Keep only un-milled rice (both sticky and non-sticky). Define rice harvest soon as a reported positive harvest of unmilled rice in the subsequent three months. Define rice harvest past as having had positive harvest of unmilled rice in the current or previous three months. Define rice farmer (or rice harvest ever) as having had a positive rice harvest at any point in the survey span.

- **Crop-Plots:** From module 5 (Crop Activities) section of the monthly survey. Make the monthly aggregate of "value transacted" for each households sale of each crop. This is the revenue from crops. For number of crop plots, I use the "projected harvest" table, which asks farmers to predict revenue for each productive crop. Every entry corresponds to a different perceived revenue stream for the farmer, so I take number of crop-plots as simply the count of these for each household in each month.

- **Aquaculture:** From module 10 (Fish-Shrimp) of the monthly survey. For each household, make monthly aggregates of the value of fish and shrimp output; this is the revenue from aquaculture. I compute whether a household does aquaculture as whether it reports raising fish/shrimp or having shrimp ponds in a given month.

- **Large Businesses:** From module 12 (Household Business) of the monthly survey. For each household, make monthly aggregates of the cash and in-kind revenue plus the value of products/services consumed by the household; this is the revenue from large businesses. Compute the number of businesses for each household as the number of entries in the household report of revenues.

- **Small/Miscellaneous Businesses:** From module 24 (Income) of the monthly survey. For each household, make monthly aggregates of the cash and in-kind revenue for each "other" income source; this is the revenue from miscellaneous businesses. Compute the number of miscellaneous activities for each household as the number of entries in the household report of revenues.

- **Number of Jobs:** From module 11 (Activities-Occupation). For each person and each job number in any month, mark if it was worked the previous two and the following two months (note that jobs are not assigned job numbers in their first months, so technically I only check the previous one month as it must have been worked the month before to have an ID). If so, it is a "steady job." I count each households total number of jobs and steady jobs each month, then compute the number of unsteady jobs as the difference. For each job and each month, sum the cash and in-kind payments and aggregate by household-month. This is the monthly job revenue.

- **Number of Activities:** I define number of activities as simply the sum of the number of crop plots, the number of livestock activities, the indicator for practice of aquaculture, the number of large businesses, the number of jobs, and the number of miscellaneous activities.

- **Total Revenue, Consumption, and Transfers:** Total revenue is the sum of revenue from crop activities, livestock activities, aquaculture, large businesses, jobs, and miscellaneous activities. Total consumption is the sum of all domestic expenditures by both cash and credit plus consumption of home-produced goods. Expenditures reported at a weekly rather than monthly frequency (in module 23W, Weekly Expenditures Update) are aggregated by month for each household and added to those reported at a monthly frequency (in module 23M, Monthly Expenditures Update). Transfers are defined as the household's net incoming transfers. More precisely, I aggregate by household-month the transfers from people inside and outside the village and subtract similarly aggregated transfers to people inside and outside the village (all found in module 13 on Remittances). I use only transfers not earmarked for a specific event because these unplanned transfers are more like insurance.

C Inference: The Two-Stage Bootstrap (For Online Publication)

The predicted mean and volatility are both generated regressors, so I must adjust my inference to account for their presence. It is easy to see that under my assumptions the full estimators match the conditions for Murphy and Topel (2002). Directly applying their analytic expressions is inconvenient and also problematic because small sample bias in the time series estimates might produce an abnormal small sample distribution for the estimated parameters. But the asymptotic normality their propositions guarantee also ensures the validity of bootstrapped confidence intervals and hypothesis tests.

[Figures 11-13 around here]

I implement the procedure as outlined in Figures 11-13. First, I prepare the time series of rice prices for resampling. I form "blocks" consisting of the contemporaneous price and however many lags I need to estimate the time series model. I then group every observation into one or more "blocks of blocks," contiguous interlocking sets of observations and their associated lags.

Next, I run the bootstrap replications. Each replication follows five intermediate steps. First, I sample with replacement the blocks of blocks of rice prices to construct a bootstrapped time series of equal length to the original time series. I estimate the parameters of the time series model on the bootstrapped data. I then resample with replacement households (together with all their monthly observations) from the panel to construct a bootstrapped panel with as many households as the original panel. (I resample households rather than villages because I find considerable variation in harvest times within villages. Village-month dummies explain only about half the variation in my "Expecting Harvest" indicator.) Then I use the estimated time series model to predict the conditional mean and variance of the international rice price for each household-month observation. Finally, I estimate the panel specification and record the resulting coefficients. I run 1000 replications for the risk specification and 2000 replications for the IV specifications.

The final step is to compute confidence intervals and p-values. To construct confidence intervals, I use the dataset of estimated parameters from bootstrap replications to find the 2.5th and 97.5th percentiles. These are the boundaries of the 95% confidence interval. To construct p-values, I compute the absolute t-statistic centered around the original parameter estimate for each replication. The fraction of these absolute t-statistics that is greater than the original t-statistic is the p-value.

D Other Tests of Robustness (For Online Publication)

[Table 9 around here]

[Table 11 around here]

E Alternative Model: Minimum Labor Inputs (For Online Publication)

Is it plausible that the kinds of activities a rice farmer can enter three months before his harvest would, as my model assumes, have a lumpy fixed cost? Finding casual labor or growing cassava may be easy if the farmer has already done so every time prices turned volatile in the past. In this appendix I build a model without fixed costs where risk still causes under-specialization. The prediction's robustness is why I emphasize that my model of risk and under-specialization is not *the* model, but just a convenient tool to formalize the intuition.

Let the household's utility function be as before and for simplicity consider the case of choosing between perfect specialization and one side activity. The household can costlessly enter a side activity but must allocate it at least $\underline{L} > 0$ units of labor. The lower-bound on labor choice captures the idea that it is not worth an employer's time to hire a worker for only a few hours per week, so even work that does not require paying a fixed cost does require a lumpy investment of time. I need the lumpiness to make specialization optimal for some degree of riskiness. Otherwise the household always has a side activity and only varies how much it works on the side activity instead of whether it has one at all. I also assume the average return to the side activity is strictly less than the average return to the primary activity—that is, $\bar{w}^p - \bar{w}^s = w^+ > 0$. The household faces the trade-off

	$M = 0$	$M = 1$
$\bar{C}$	$\bar{w}^p$	$\bar{w}^p - w^+(1 - L^p)$
V	σ_p^2	$(L^p)^2\sigma_p^2 + (1 - L^p)^2\sigma_s^2$

The opportunity cost of the side activity is $w^+(1 - L^p)$, and since it is no less than $w^+\underline{L} > 0$ the household still loses a discrete chunk of expected revenue when it diversifies. Although it does not literally pay a fixed cost the household's trade-off between the mean and variance of consumption is similar to the one it faced in the original model. They are not identical—for example, the cost of diversification is now uncertain—but similar enough for risk to cause under-specialization.

[Figure 14 around here]

Figure 14 gives the intuition. With perfect specialization the household's expected utility is maximized when the primary activity's returns have zero variance, but expected utility falls steeply as the variance rises. The household can flatten the utility-variance relationship by moving some labor from the primary activity to the side activity. Without a lower bound on labor devoted to the side activity, the household would always move ε units of labor to the side activity and be happier without perfect specialization. But with a lower bound the household must accept a discretely lower and flatter utility-variance relation. If the variance of the side activity is low, the household prefers specialization. But when the variance exceeds a critical threshold the household prefers to diversify. If w^+ has a nondegenerate distribution the average number of activities will rise continuously with the variance. Then the lower bound model makes the same prediction $\frac{d\mathbb{E}[M]}{d\sigma_p^2} > 0$ as the fixed cost model from the main text.

Table 8
Modeling the Rice Price as a Random Walk

	(1)
	P_t
P_{t-1}	0.995***
	(0.00)
N	389
R^2	0.995

Note: The random walk specification describes the data well. It models the current price of rice as the previous month's price plus a random innovation: $P_t = P_{t-1} + \varepsilon_t$.

Table 9
Robustness: Main Results Excluding Pre-Harvest Rice Sales

	(1) Activities	(2) Revenue
	(1)	(2)
	Activities	Revenue
Activities		-14195.18**
		[0.027]
Rice Farmer		
- × Mean	0.00	-86.06
	[0.395]	[0.307]
- × Volatility	-0.09**	-368.73
	[0.011]	[0.531]
Expecting Harvest		
- Main	1.37***	-238.27
	[0.000]	[0.953]
- × Mean	-0.01***	(Excluded Instrument)
	[0.000]	
- × Volatility	0.05*	(Excluded Instrument)
	[0.091]	
Recent Harvest		
- Main	-0.63	-27329.39*
	[0.263]	[0.077]
- × Mean	-0.01***	147.78
	[0.003]	[0.390]
- × Volatility	0.17***	1095.82
	[0.003]	[0.370]
Household Fixed-Effects	Yes	Yes
Month Fixed-Effects	Yes	Yes
Households	743	743
Observations	47395	47395
F-stat Exc. Inst.		10.054
Hansen's J Stat		0.015

Note: .I exclude observations when households claim they sold rice while still expected their harvest. Volatility still causes households to enter more activities (Column 1) and the extra activities are costly (Column 2).

Table 10
Main Regression with Only Rice Farmers

	Activities
Mean	0.00*
	[0.050]
Volatility	-0.28***
	[0.000]
Expecting Harvest	
- Main	1.82***
	[0.004]
- × Mean	-0.02***
	[0.000]
- × Volatility	0.18***
	[0.003]
Recent Harvest	
- Main	-0.76
	[0.482]
- × Mean	-0.03***
	[0.000]
- × Volatility	0.41***
	[0.000]
Household Fixed-Effects	Yes
Month Fixed-Effects	No
Households	354
Observations	23613

Note: I confirm the results hold when I exclude non-farmers. The bracketed values are p-values. I compute the p-values using a two-stage bootstrap that corrects for generated regressors and clusters by household (see Appendix C).

Table 11
Robustness: Regressor of Interest Does not Affect Wages

	(1)
	Median Wage
Mean	0.01
	[0.424]
Volatility	0.66
	[0.167]
Rice Farmer	
- Main	-43.31
	[0.281]
- × Mean	0.01
	[0.891]
- × Volatility	0.52
	[0.775]
Expecting Harvest	
- Main	24.62
	[0.478]
- × Mean	-0.06
	[0.444]
- × Volatility	-2.90
	[0.483]
Recent Harvest	
- Main	20.65
	[0.193]
- × Mean	-0.07
	[0.498]
- × Volatility	-1.80
	[0.477]
Village Fixed-Effects	Yes
Month Fixed-Effects	No
Villages	16
Observations	1152

Note: Suppose wages are correlated with volatility in the rice price. Then the extra jobs the household takes up may not be a response to risk but rather a response to better earnings in side activities. This table shows that median village wages are not correlated with the village averages of any of the regressors of interest. I run the analysis at the village rather than individual level because households might be willing to take lower paying jobs to hedge against risk, lowering the average wage of jobs held even though volatility has no confounding effect on wages.

Figure 11
Bootstrap, Step 1: Forming Blocks of Blocks

Make "blocks" of current obs and lags

$\vdots$

$\mathbf{t-3}: P_{t-3}$
$\mathbf{t-2}: P_{t-2}$
$\mathbf{t-1}: P_{t-1}$
$\mathbf{t}: P_t$
$\mathbf{t+1}: P_{t+1}$
$\mathbf{t+2}: P_{t+2}$
$\mathbf{t+3}: P_{t+3}$
$\mathbf{t+4}: P_{t+4}$

$\vdots$

$\mathbf{t-3}: (P_{t-3}\ P_{t-4}\ P_{t-5}\ P_{t-6})$
$\mathbf{t-2}: (P_{t-2}\ P_{t-3}\ P_{t-4}\ P_{t-5})$
$\mathbf{t-1}: (P_{t-1}\ P_{t-2}\ P_{t-3}\ P_{t-4})$
$\mathbf{t}: (P_t\ P_{t-1}\ P_{t-2}\ P_{t-3})$
$\mathbf{t+1}: (P_{t+1}\ P_t\ P_{t-1}\ P_{t-2})$
$\mathbf{t+2}: (P_{t+2}\ P_{t+1}\ P_t\ P_{t-1})$
$\mathbf{t+3}: (P_{t+3}\ P_{t+2}\ P_{t+1}\ P_t)$
$\mathbf{t+4}: (P_{t+4}\ P_{t+3}\ P_{t+2}\ P_{t+1})$

Make blocks of the blocks

1
$\mathbf{t-3}: (P_{t-3}\ P_{t-4}\ P_{t-5}\ P_{t-6})$
$\mathbf{t-2}: (P_{t-2}\ P_{t-3}\ P_{t-4}\ P_{t-5})$
$\mathbf{t-1}: (P_{t-1}\ P_{t-2}\ P_{t-3}\ P_{t-4})$

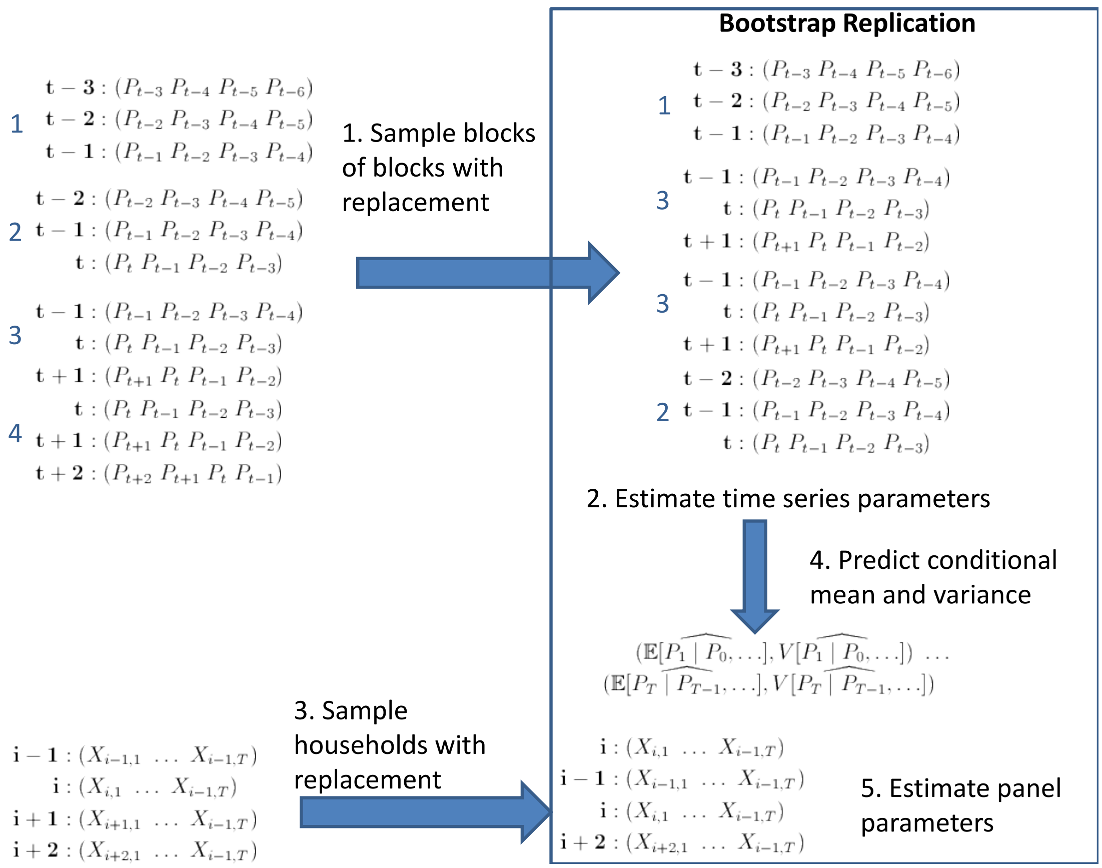
2
$\mathbf{t-2}: (P_{t-2}\ P_{t-3}\ P_{t-4}\ P_{t-5})$
$\mathbf{t-1}: (P_{t-1}\ P_{t-2}\ P_{t-3}\ P_{t-4})$
$\mathbf{t}: (P_t\ P_{t-1}\ P_{t-2}\ P_{t-3})$

3
$\mathbf{t-1}: (P_{t-1}\ P_{t-2}\ P_{t-3}\ P_{t-4})$
$\mathbf{t}: (P_t\ P_{t-1}\ P_{t-2}\ P_{t-3})$
$\mathbf{t+1}: (P_{t+1}\ P_t\ P_{t-1}\ P_{t-2})$

4
$\mathbf{t}: (P_t\ P_{t-1}\ P_{t-2}\ P_{t-3})$
$\mathbf{t+1}: (P_{t+1}\ P_t\ P_{t-1}\ P_{t-2})$
$\mathbf{t+2}: (P_{t+2}\ P_{t+1}\ P_t\ P_{t-1})$

Note: First, I prepare the time series of rice prices for resampling. I form "blocks" consisting of the current price and however many lags are needed to estimate the time series model. I then group every observation into one or more "blocks of blocks," adjacent interlocking sets of observations and their associated lags.

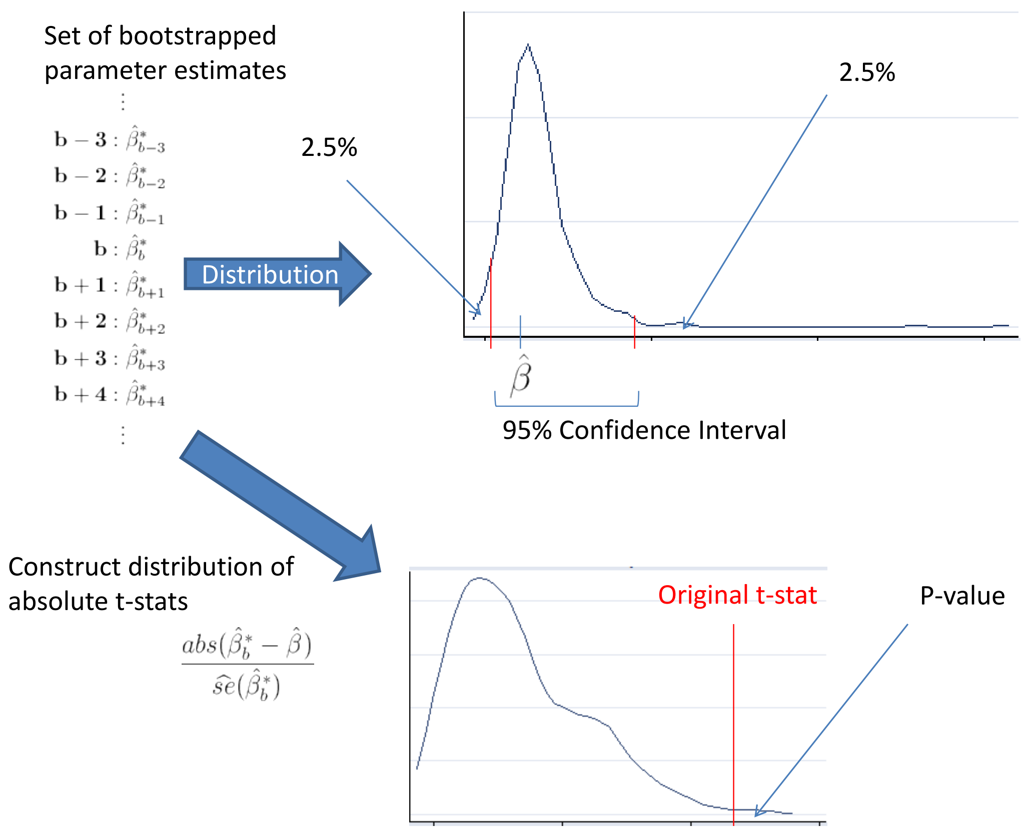
Figure 12
Bootstrap, Step 2: Bootstrap Replications



Note: Next I run the bootstrap replications. Each replication follows five steps. First I sample with replacement the blocks of blocks of rice prices to construct a bootstrapped time series of the same length as the original time series. Next I estimate the parameters of the time series model on the bootstrapped data. I then resample with replacement households (together with all their monthly observations) from the panel to construct a bootstrapped panel with as many households as the original panel. Then I use the estimated time series model to predict the conditional mean and variance of the international rice price for each household-month observation. Finally, I estimate the panel specification and record the resulting coefficients. I run 2000 replications for the risk specifications and 3000 replications for the IV specifications.

Figure 13

Bootstrap, Step 3: Constructing Confidence Intervals and P-Values



Note: I compute the absolute t-statistic centered around the original parameter estimate for each replication. The fraction of these absolute t-statistics that is greater than the original t-statistic is the p-value.

Figure 14
Intuition of the Alternative Model