Inverse Productivity or Inverse Efficiency? Evidence from Mexico

Justin Kagin
Kagin's Consulting

J. Edward Taylor
University of California, Davis

Antonio Yúnez-Naude
El Colegio de México

Paul Collier

The peasant life forces millions of ordinary people into the role of entrepreneur, a role for which most are ill suited…their mode of production is ill suited to modern agricultural production…Given the chance, peasants seek local wage jobs, and their offspring head to the cities.

Outline

- Productivity and Efficiency
- Testing the Inverse Relationship: A Panel Approach
- Stochastic Frontier Analysis
- Testing the Inverse Efficiency Hypothesis
- Conclusions

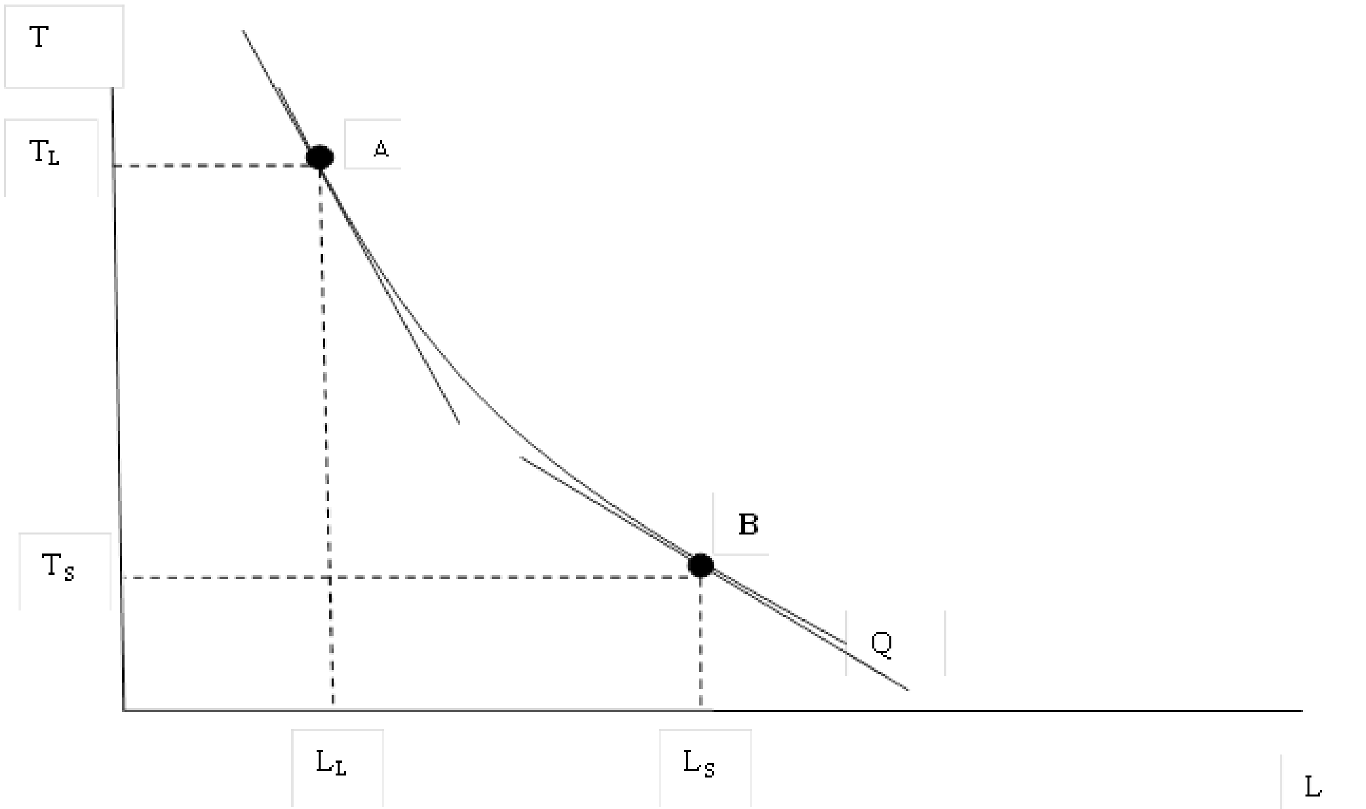
Inverse-Farm Size (IR) Productivity Relationship

- IR hypothesis: Small farms produce more output per hectare than large farms

- Reasons

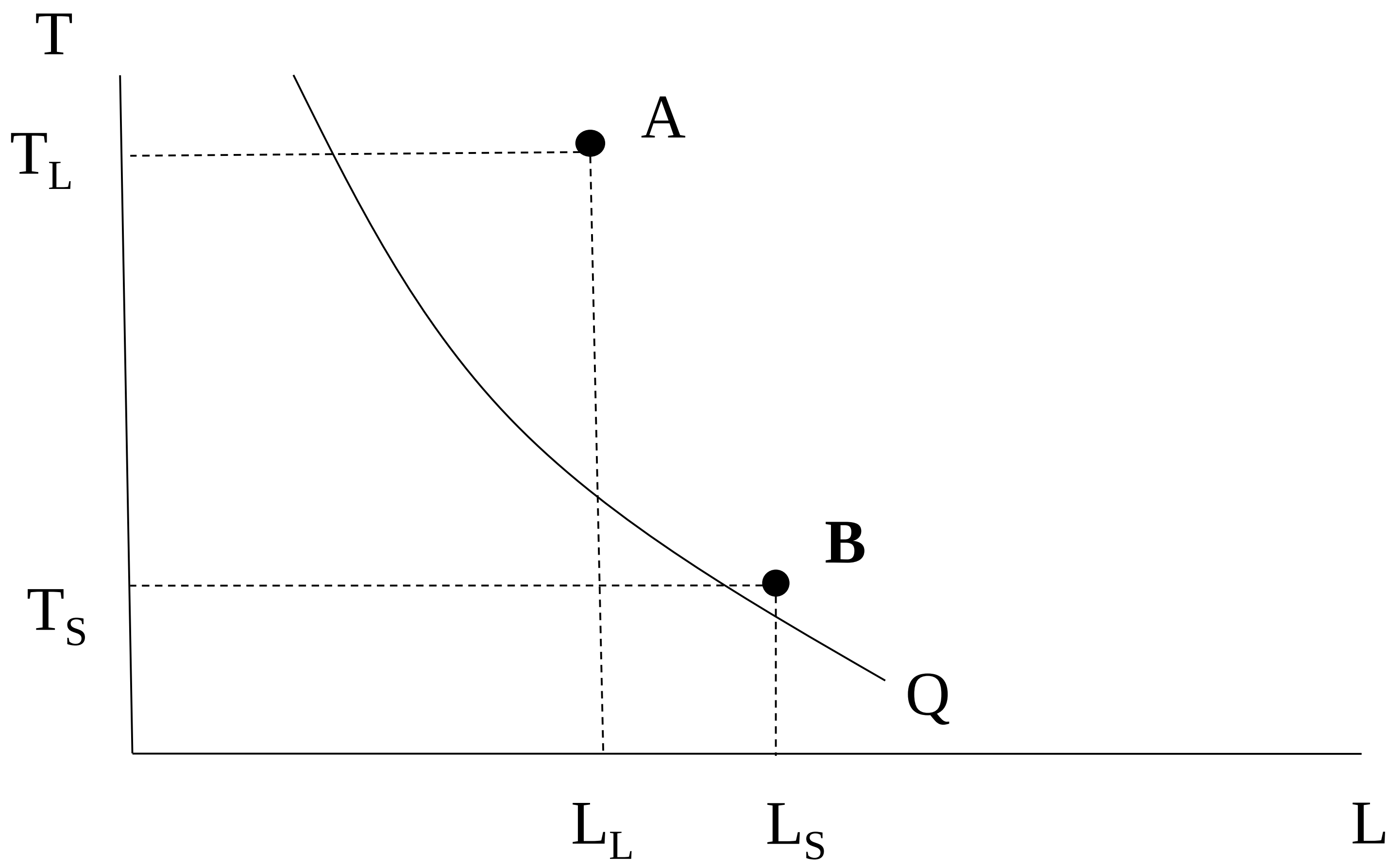
  - Risk
  - Missing Markets (labor, credit)
  - Unobserved Variables (Spurious Correlation)

Inverse Productivity does not Necessarily Imply Inverse Efficiency

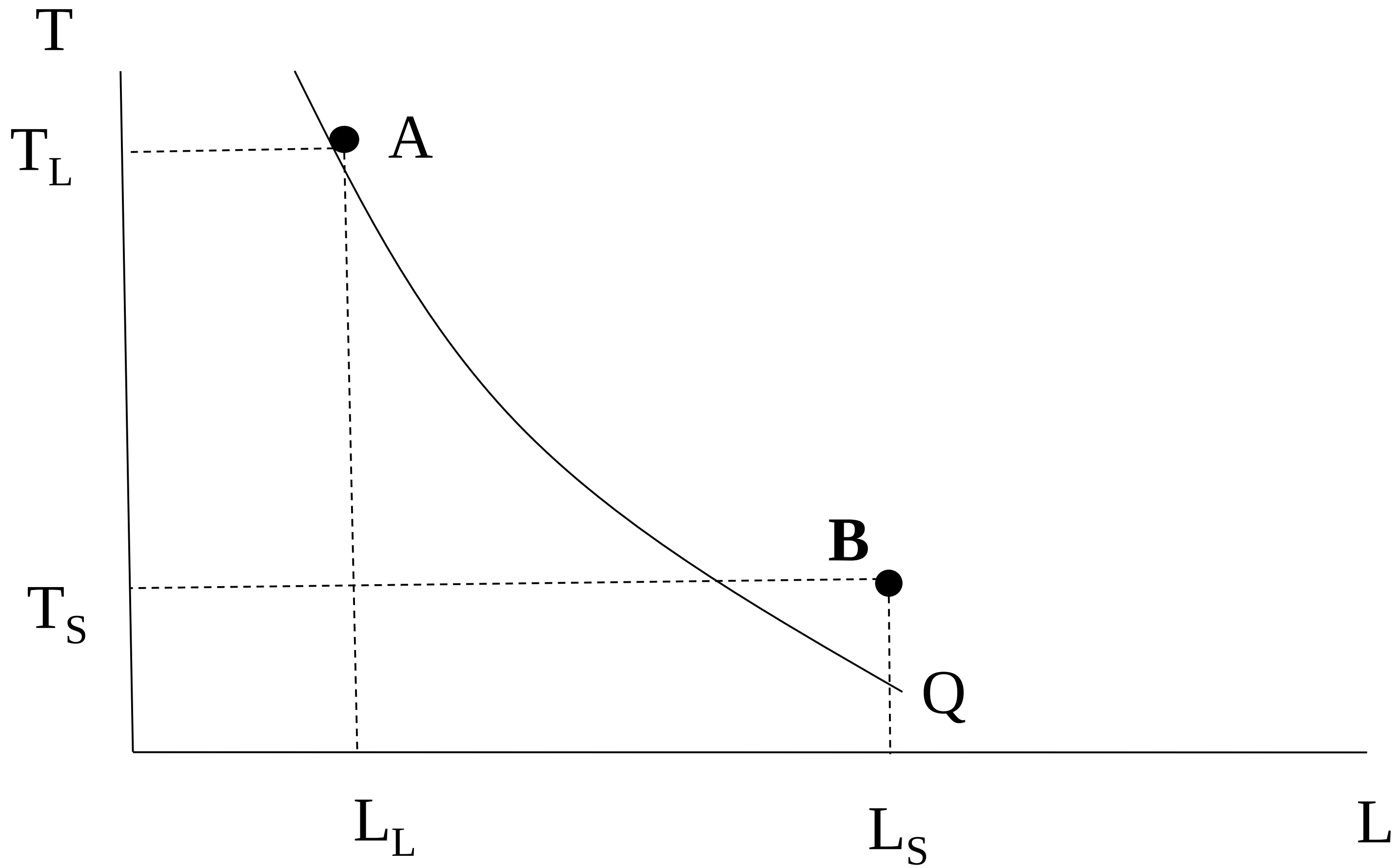


Two efficient producers (but with different relative prices)

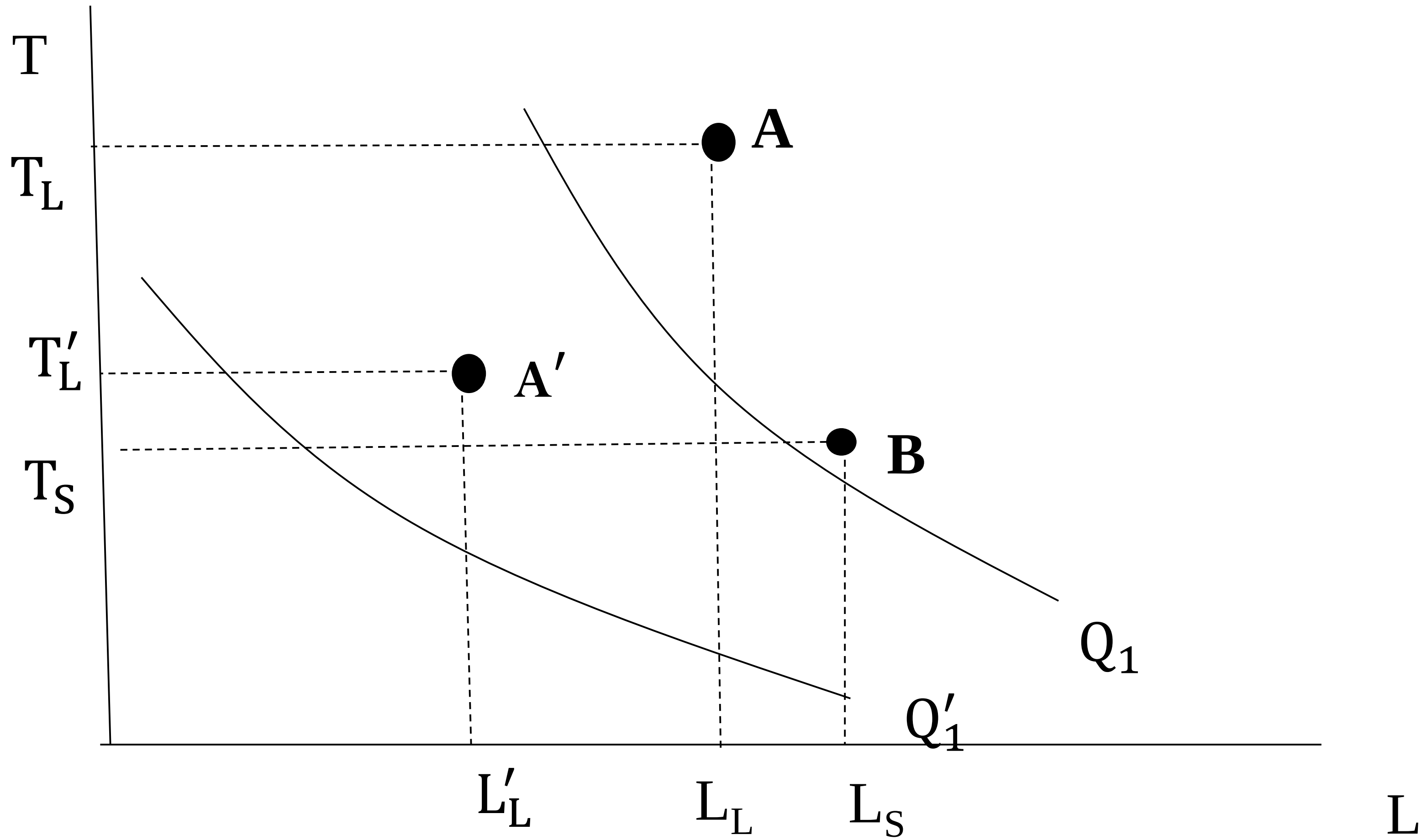
…but both can occur



...or not



Collier's Argument: Small Farms Left In the Dust



Technology changes increase relative productivity and efficiency of large farms

Small Producers and Technical Efficiency

- Scale effects are important: Is smallholder agriculture capable of initiating self-sustained economic growth?
  - It may be doubtful if missing markets or risk are the main contributors to the inverse productivity relationship
- On the other hand, if smallholders are both productive as well as efficient a heterogeneous structure of agricultural producers could promote growth.
  - An empirical question. . .

ENHRUM 2002 & ENHRUM 2007

- 2003 and 2008 rounds of the Mexico National Rural Household Survey (*Encuesta Nacional de Hogares Rurales de Mexico*, or ENHRUM).
- Five-year matched longitudinal data set on assets, socio-demographic characteristics, production, income sources, and migration from a nationally representative sample of rural households.
- The 2003 sample includes 1,782 households in 14 Mexican states; of these, 1,543 were successfully re-surveyed in 2008.

Descriptive Statistics

Variable	Small (<=3 hectares)	Large (>3 hectares)	Difference
Agricultural production per hectare (pesos)	10,047.1	4564.48	**
Ag land (hectares)	1.497244	9.82779	***
Ag labor (days)	84.86	143.46	***
Ag purchased inputs (pesos)	4163.76	14,216.98	***
Household head education	3.73	4	
Household head speaks an indigenous language	38%	24%	***
Distance to the US border	7.23	6.17	***
Only dirt roads	12.18%	6.45%	***
# of Households	574	788	

Testing the Inverse Relationship

$$\ln(y_{it}) = \alpha_i + \beta_1 \ln(T_{it}) + \beta_2 \ln(L_{it}) + \beta_3 \ln(K_{it}) + \beta_4 \ln(PI_{it}) + \epsilon_{it}$$

- y_{it}: Agricultural output value per hectare (constant pesos), person *i* and period *t*
- T_{it}: Farm size
- L_{it}: Farm labor (family and hired labor-days)
- K_{it}: Capital services (machinery and animal)
- PI_{it}: Purchased inputs (fertilizer, pesticides, seeds, etc.)
- α_i: Household fixed effects.

- In a constant-returns-to-scale economy with perfect factor markets, there should be no observed differences in productivity across farm sizes.

We find the Inverse Productivity Relationship in Mexico

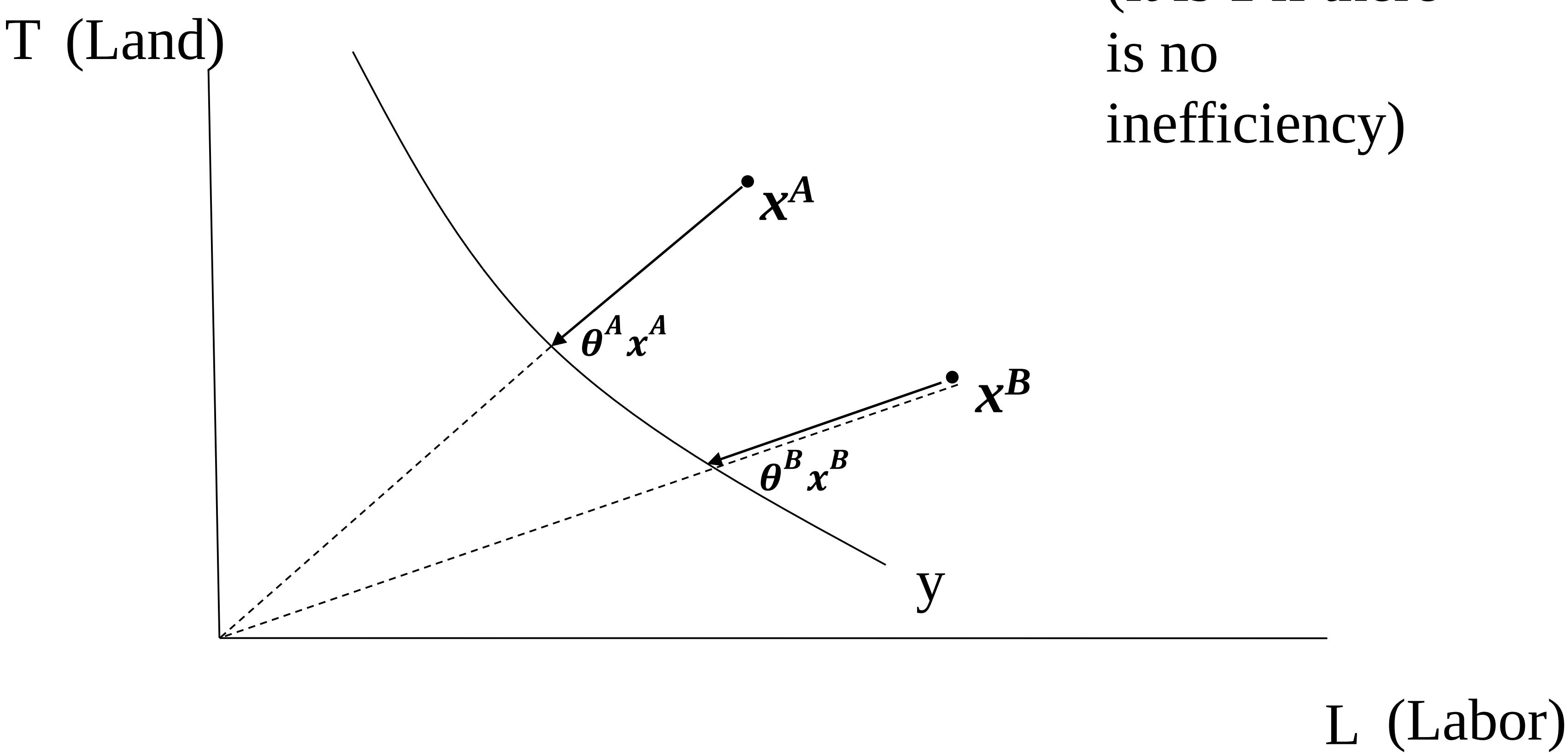
Variable	Value of Crop Output (Pesos) per Hectre RE	FE
Ln (Ag. Land)	-0.055	-0.226***
	(0.038)	(0.071)
Ln (Ag. Labor/Ag. Land)	0.195***	0.197***
	(0.030)	(0.046)
Ln (Ag. Capital/Ag. Land)	0.043***	-0.004
	(0.010)	(0.018)
Ln (Ag. Other Costs/Ag. Land)	0.156***	0.042*
	(0.015)	(0.023)
Year	0.168***	0.144**
	(0.059)	(0.065)
Constant	5.98***	6.86***
	(0.15)	(0.21)
Observations	1,361	1,361
Number of Households	842	842

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Stochastic Frontier Analysis

- $TE(y, x^A) = \frac{\theta^A x^A}{x^A}$ and $TE(y, x^B) = \frac{\theta^B x^B}{x^B} \leq 1$

(it is 1 if there is no inefficiency)



Empirical Formulation

$$y \leq f(\boldsymbol{x})$$

$$TE(y, \boldsymbol{x}) = \frac{y}{f(\boldsymbol{x})} \leq 1$$

$$y_{it} = f(\boldsymbol{x}_{it}, \boldsymbol{\beta}) TE_{it} e^{v_i}$$

$$\ln(y_{it}) = \boldsymbol{\beta} ln(\boldsymbol{x}_{it}) + \ln(TE_{it}) + v_{it},\ v_{it}\ iid \sim N(0, \sigma_v^2)$$

A measure of technical inefficiency is thus $u_{it} = -\ln(TE_{it})$ where $u_{it} \geq 0$ and assumed to be independent of v_{it}

$$Ln(Y_{it}) = \beta_0 + \beta_1 \ln(T_{it}) + \beta_2 \ln(L_{it}) + \beta_3 \ln(K_{it}) + \beta_4 \ln(C_{it}) + v_{it} - u_{it}$$

Inefficiency Estimate

$$V_{it} - U_{it}$$

- Observed deviations from the production function could arise from two sources:

  - (1) productive inefficiency, which necessarily would be negative (u_{it}); and
  - (2) idiosyncratic effects that are specific to the farm and that could enter the model with either sign (v_{it}).

- The specification of the error and efficiency terms we choose is that of Battese and Coelli (1992) with time varying inefficiency effects.

- The u_{it} are independently distributed as truncations at zero of the $N(m_{it}, \sigma_u^2)$

- We let $m_{it}=z_{it}'\delta$, where z_{it} is a vector of variables that may influence the efficiency of the farm, and δ is a vector of parameters to be estimated (Coelli et al., 1998).

Results

Productivity

Inefficiency

Model	(1)	(2)	(3)	(4)
Frontier Estimates	RE	RE	RE	FE
β_0	6.28***	6.35***	6.25***	6.25***
	(0.14)	(0.14)	(0.13)	(0.14)
ln (Ag. Land)	0.011	-0.098**	-0.102***	-0.254***
	(0.038)	(0.032)	(0.038)	(0.039)
ln (Ag. Labor/Ag. Land)	0.194***	0.175***	0.176***	0.167***
	(0.029)	(0.029)	(0.027)	(0.029)
ln (Ag. Capital/Ag. Land)	0.046***	0.0326***	0.0326***	0.019*
	(0.010)	(0.0010)	(0.0098)	(0.010)
ln (Ag. Cost/Ag. Land)	0.189***	0.164***	0.187***	0.059***
	(0.015)	(0.015)	(0.014)	(0.015)
Year		0.200***	0.187***	0.090**
		(0.066)	(0.068)	(0.044)
Good Land (% of total cultivated)		0.095	0.088	
		(0.075)	(0.071)	
Flat Land (% of total cultivated)		0.292***	0.291***	
		(0.077)	(0.075)	
Irrigated Land (# hectares)		0.083***	0.082***	
		(0.010)	(0.010)	
Household Head Education			0.029***	
			(0.010)	
Farm Fixed Effects				X
Inefficiency Estimates				
δ_0	-21.89***	-18.31**	-17.44	-3.36***
	(8.34)	(9.09)	(11.05)	(0.12)
ln (Ag. Land)	1.01***	1.23	1.24	0.16**
	(0.26)	(0.76)	(1.04)	(0.064)
σ_s^2	13.68***	11.86**	11.25*	1.35***
	(4.46)	(4.74)	(5.74)	(0.031)
γ	0.900***	0.897***	0.891***	0.99998***
	(0.032)	(0.043)	(0.058)	(0.00028)
Observations	1363	1363	1361	1361
Number of Households	843	843	841	841

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Explaining Inefficiency

Frontier Estimates	Model 5 FE
β_0	6.47***
	(0.15)
ln (Ag. Land)	-0.211***
	(0.043)
ln (Ag. Labor)	0.129***
	(0.031)
ln (Ag. Capital)	-0.007
	(0.010)
ln (Ag. Cost)	0.045***
	(0.015)
Year	1.20***
	(0.10)
Farm Fixed Effects	x
Inefficiency Estimates	
δ_0	-2.75***
	(0.26)
ln (Ag. Land)	0.257***
	(0.078)
Only dirt roads	0.89***
	(0.20)
International Migration	-1.57***
	(0.24)
Household Head Indigenous	0.83***
	(0.21)
Year	2.38***
	(0.25)
σ_s^2	0.728***
	(0.024)
γ	0.99999***
	(0.00022)
Observations	1348
Number of Households	836

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The efficient frontier and the inverse productivity relationship

The inverse efficiency relationship

Significant determinants of inefficiency related to transaction costs, transfer of knowledge; Indigenous producers disadvantaged with respect to information, access to resources; Value traditional varieties that are not highly tradable

Conclusions

- Theory is generally unclear about the relationship between productivity and efficiency
- We estimate an inverse relationship for efficiency and productivity
- Access to high quality land, schooling, and unobserved variables shape the efficiency frontier
- Market access, migration, and being indigenous influence efficiency, controlling for the other variables
- A more or less optimistic portrait of the capacity of small producers in Mexico
- Policies to increase efficiency:
  - Invest in market infrastructure
  - Supporting indigenous producers

The End

Supplemental Slides

Inefficiency Equation

$$U_{it} = \delta_0 + \delta_1 \ln(T_{it}) + \delta_2(D_{it}) + \delta_3(R_{it}) + \delta_4(I_{it})$$

- Vector of controls reflecting households' ability or incentives to efficiently transform inputs into output
  - Transaction costs (proxied by R_{it}, a dummy variable indicating whether a village has only dirt roads)
  - Ethnicity (I_{it}, a dummy for whether the household head speaks an indigenous language
  - Migration (instrumented by D_{it}, the distance to the US border by train (km/100), as in Demirgüç-Kunt et al. (2007) and Pfeiffer, et al. (2009)

Efficiency Frontier Equation

$$Ln(Y_{it}) = \beta_0 + \beta_1 \ln(T_{it}) + \beta_2 \ln(L_{it}) + \beta_3 \ln(K_{it}) + \beta_4 \ln(C_{it}) + \beta_5 Z_{it} + V_{it} - U_{it}$$

- Elements of Z_{it}:
  - Human capital (the education of the household head)
  - Self-reported land quality and land slope (each scaled so that the higher its value the worse the land quality and slope)
  - Irrigation (0-1 indicator variable)

Inefficiency Equation – Full Specification

$$U_{it} = \delta_0 + \delta_1 \ln(T_{it}) + \delta_2(R_{it}) + \delta_3(M_{it}) + \delta_4(I_{it}) + W_{it}$$

- Vector of controls reflecting households' ability or incentives to efficiently transform inputs into output
  - Transaction costs (proxied by R_{it}, a dummy variable indicating whether a village has only dirt roads)
  - US Migration, instrumented by M_{it}, as in Demirgüç-Kunt et al. (2007) and Pfeiffer, et al. (2009)
  - Ethnicity (I_{it}, a dummy for whether the household head speaks an indigenous language

Likelihood Ratio Tests on Inefficiency Estimates

Model	(1)	(2)	(3)	(4)	(5)
Hypothesis	$H_0 : \gamma = \delta_0 = \delta_1 = 0$	$H_0 : \gamma = \delta_0 = \delta_1 = 0$	$H_0 : \gamma = \delta_0 = \delta_1 = 0$	$H_0 : \gamma = \delta_0 = \delta_1 = 0$	$H_0 : \gamma = \delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$
5% critical value*	7.05	7.05	7.05	7.05	11.91
Likelihood Ratio Statistic	10.09	14.75	14.24	789.45	377.45

*The likelihood ratio statistic follows a mixed Chi-squared distribution and thus critical values are obtained from Kodde and Palm (1986), Table 1 p. 1246. Degrees of freedom are equal to the number of parameters equal to zero.

Adjusted Coefficients

Model	(1)	(2)	(3)	(4)
Coefficient on Ag. Land	RE	RE	RE	FE
ln (Ag. Land)	-0.174	-0.347	-0.360	-0.299

Risk

- Sandmo vs. Finklestein and Chalfant
  - Income and Purchasing Power Risk for small farmers

Missing Markets

- Households derive utility from a consumption good, C, and leisure L^L
- It has an endowment of land, T, and of labor time, $\bar{L}$
- Time is divided between working on farm and leisure:

$$\bar{L} = L^F + L^L$$

- Production is $F(T, L^F)$, exhibits CRTS.

Missing Markets

- Assuming there is no land market, so that land is a fixed factor, than we can write farm output as:

$$TF(L^F/T, T/T) = Tf(l)$$

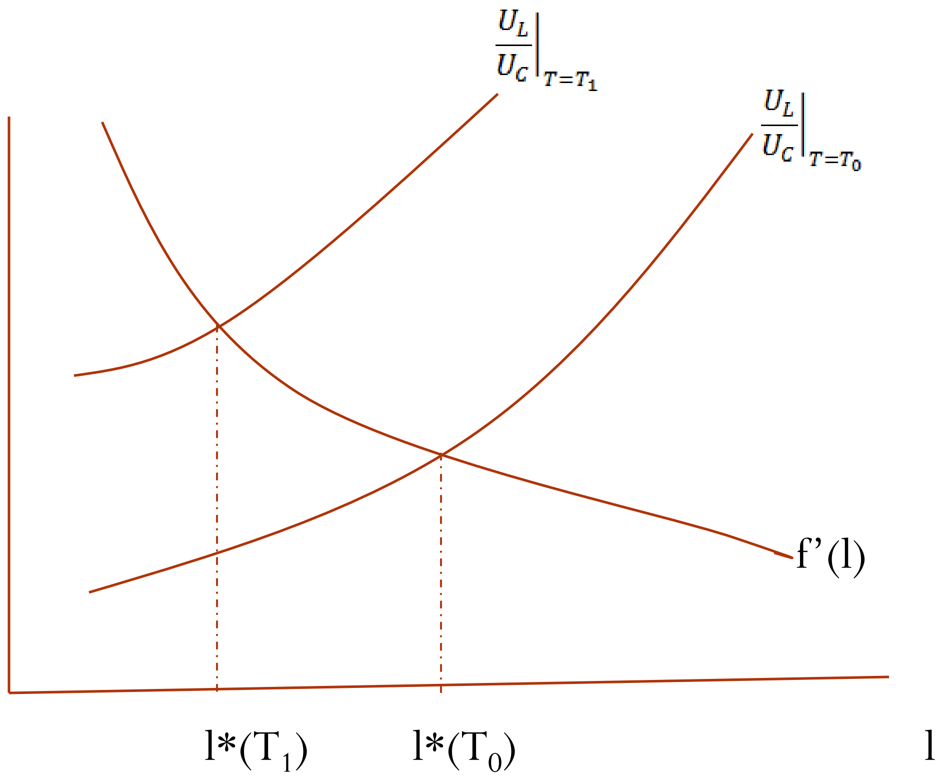
- The household's optimization problem is thus:

$$\max_{C, L^L} U(C, L^L) = \max_{l} U(Tf(l), \bar{L} - Tl)$$

- The FONC: $$\frac{dU}{dl} = 0 \rightarrow f' U_C = U_{L^L}$$

Missing Markets

- Marginal Rate of Substitution: $$f'(l) = \frac{U_{L^L}(C, L^L)}{U_C(C, L^L)}$$



Spurious Correlation

- Regress productivity, y_i on farm size, *T*:

$$y_i = \alpha + \beta T_i + \epsilon_i$$

- But only literate farmers are able to respond to the survey. u_i is unobserved, perhaps motivation to succeed.

$$L_i = \gamma + \delta T_i + u_i$$

- Let *R* be a binary variable taking value 1 if the farm responds (is literate) and 0 if not:

$$R_i = \begin{cases} 1 \; if \; L_i \geq 0 \rightarrow u_i \geq -(\gamma + \delta T_i) \\ 0 \; if \; L_i < 0 \rightarrow u_i < -(\gamma + \delta T_i) \end{cases}$$

Spurious Correlation

$$E[y_i|R_i = 1] = \alpha + \beta T_i + E[\epsilon_i|R_i = 1]$$

- But

$$E[\epsilon_i|R_i = 1] = E[\epsilon_i|u_i \geq -(\gamma + \delta T_i)]$$

- And it might be that $cov(\epsilon_i, u_i) \neq 0$

Spurious Correlation

- Need a high u_i to get into the sample