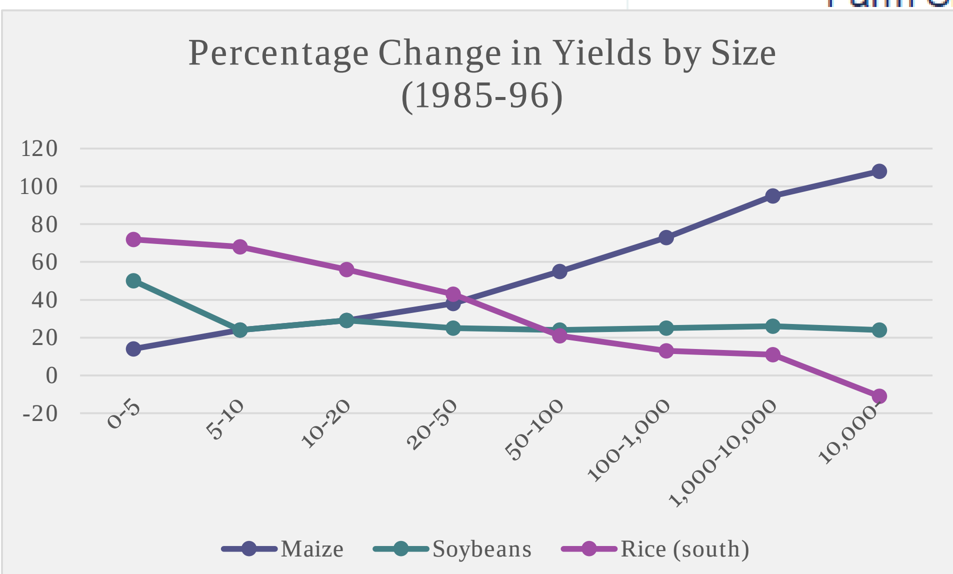
Large and Small Farms Excel in Brazil

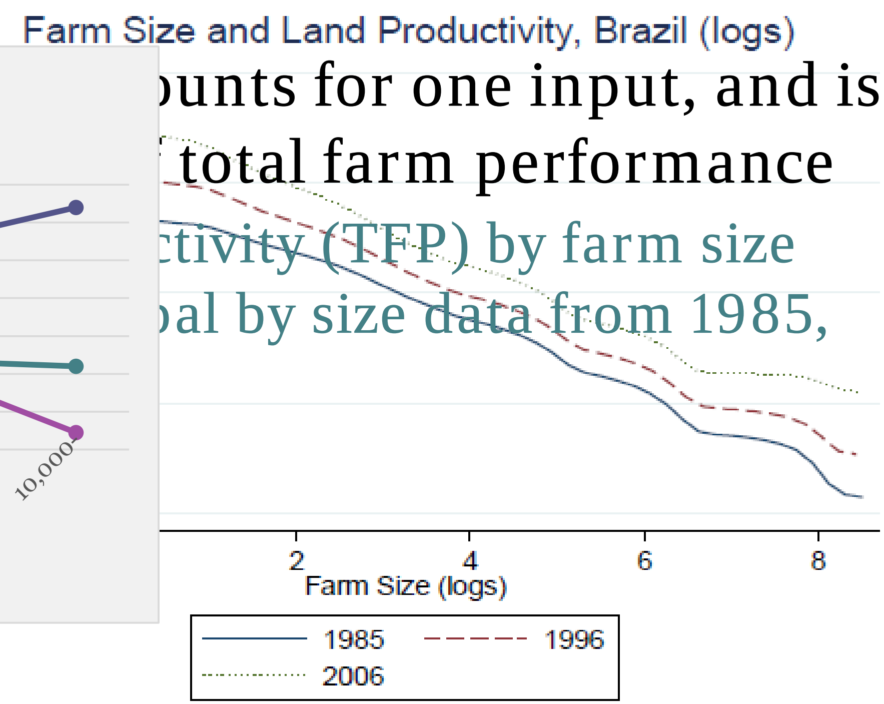
Steven M. Helfand, UCR
Nicholas E. Rada, ERS/USDA
Marcelo M. Magalhães, UNESP

USDA, February 2, 2017

1. Motivation and Objectives

- What is the farm size – productivity relationship over time in Brazil?
  - Defining productivity as crop yield (output/hectare), we find mixed results



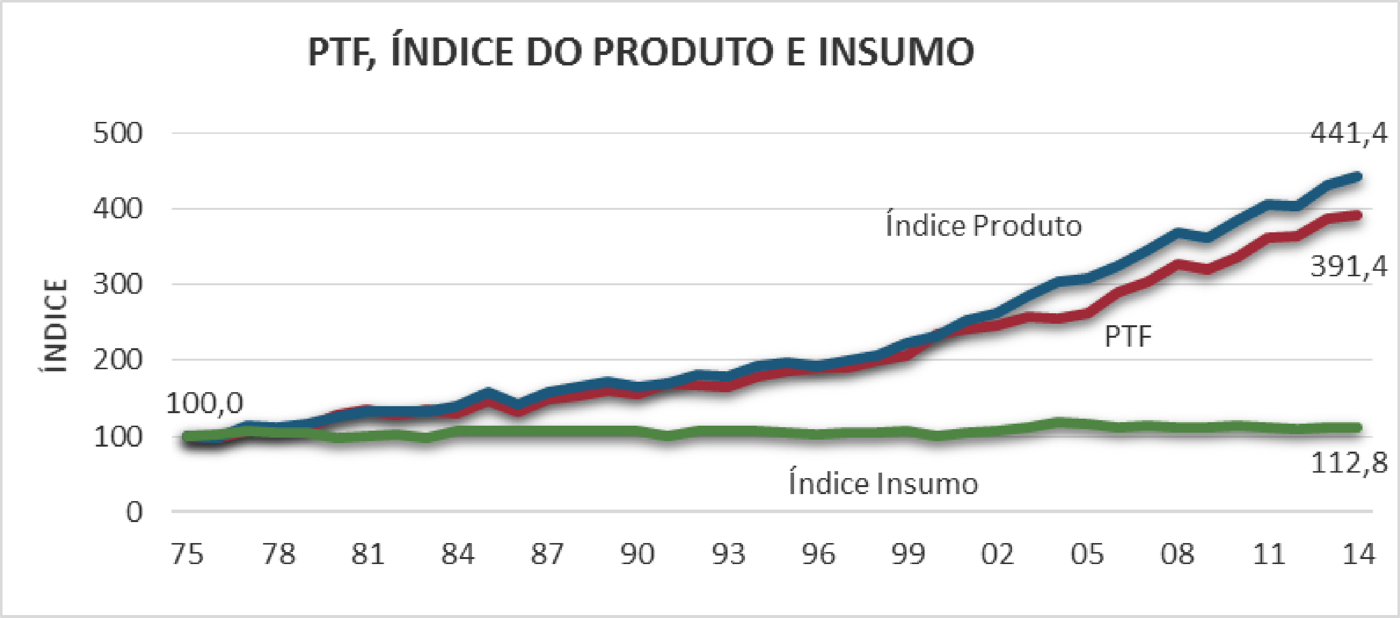


Note: Smoothed as a local polynomial regression with bandwidth of 1.25 and Epanechnikov kernel.

Source: Helfand and Taylor (2016); Fig. 1, pp. 4

…unts for one input, and is
… total farm performance
…ctivity (TFP) by farm size
…al by size data from 1985,

- Once we account for the spectrum of inputs, are Brazil's large farms the most productive? Can small farms continue to compete?
- There is little evidence on this issue. The overwhelming majority of evidence on TFP for Brazil is at the national or state levels.



Gasques, 2015.

- To answer this question we:
  - Estimate TFP in 1985, TFP growth in 1985-2006, and TFP in 2006 for five farm size classes
  - Decompose TFP growth beyond technical change and average efficiency change
    - We examine the full distribution of producer performance for each farm size class
  - Assess the role of output specialization and K-intensity in raising TFP growth
  - Evaluate how policies affect output, input, and TFP growth
    - Technical assistance, credit, electricity, and education

Outline

1. Motivation and objectives

2. Alternative models and challenges
   - The long and winding path to a relatively simple model
   - Little evidence to support a frontier (pseudo-panel)

3. Results
   - Descriptive
   - Model estimation
   - TFPC and its distribution and decomposition
   - Output specialization
   - Policies: credit, technical assistance, and education

4. Conclusions, next steps, remaining challenges

2. Alternative models and challenges

A. How to model inefficiency over time, and how to make it a fn of explanatory variables?
   - BC 92, BC 95, Wang 02
   - These models ignore unobserved heterogeneity =>
     - All deviations from frontier are inefficiency

B. How to model heterogeneity and inefficiency?
   - Greene (2005) TFE, TRE; CRE
   - Because we don't have farm level data, some differences in inefficiency—managerial ability—will cancel out in the process of aggregating
   - Thus, we are particularly concerned with heterogeneity

C. Models that allow for time varying inefficiency and try to separate time invariant effects into a) heterogeneity and b) persistent inefficiency
   - Kumbhakar et al. (2014), Colombi et al. (2014)

- In Stata and Limdep, we have estimated a number of these models

- General conclusions
  - Little evidence of skewness
  - Symmetric error dominates asymmetric error:
    - Vsigma often 10x or 15x larger than Usigma => lack of skew
  - Many models did not converge
  - When they did converge, TFP growth was not credible
    - TE too high/TEC too low/TFPC too low
  - Difficulty of capturing heterogeneity (FE, RE, CRE)
  - Often not able to include weights in frontier models

See evidence on following slides

APF w/ FEs and Environmental Shocks

```
Fixed-effects (within) regression               Number of obs      =     47526
Group variable: pid                             Number of groups   =     15937

R-sq:  within  = 0.5565                         Obs per group: min =         1
       between = 0.9298                                        avg =       3.0
       overall = 0.8736                                        max =         3

                                                F(52,31537)        =    761.11
corr(u_i, Xb)  = -0.3178                        Prob > F           =    0.0000
```

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
A	.2300766	.0507139	4.54	0.000	.1306753	.3294778
L	.7721255	.0798309	9.67	0.000	.6156539	.9285972
I	.5948066	.0374662	15.88	0.000	.5213713	.6682418
K	-.0717342	.0461056	-1.56	0.120	-.162103	.0186345
AA	.079134	.0083132	9.52	0.000	.0628399	.0954281
LL	.0416005	.0180606	2.30	0.021	.006201	.077
II	-.0490668	.0074277	-6.61	0.000	-.0636254	-.0345082
KK	.0269398	.0082173	3.28	0.001	.0108336	.043046
AL	-.0376496	.0127478	-2.95	0.003	-.0626359	-.0126634
AI	.0125246	.0047956	2.61	0.009	.003125	.0219242
AK	-.0490146	.0058436	-8.39	0.000	-.0604682	-.037561
LI	-.0442728	.0119448	-3.71	0.000	-.0676852	-.0208605
LK	.0003552	.0116378	0.03	0.976	-.0224553	.0231657
IK	.0326421	.0068066	4.80	0.000	.0193009	.0459833
T	.019541	.0039772	4.91	0.000	.0117455	.0273364
TT	.0022151	.0001906	11.62	0.000	.0018416	.0025886
AT	.0059797	.0003931	15.21	0.000	.0052092	.0067503
LT	-.001741	.001288	-1.35	0.176	-.0042654	.0007835
IT	.0002635	.000585	0.45	0.652	-.000883	.00141
KT	-.005284	.0005603	-9.43	0.000	-.0063821	-.0041858
_cons	2.641486	.2425521	10.89	0.000	2.166074	3.116897
sigma_u	.44593824					
sigma_e	.5090227					
rho	.4342273	(fraction of variance due to u_i)				

```
F test that all u_i=0:     F(15936, 31537) =     1.54          Prob > F = 0.0000
```

No Skew in error of the FE model

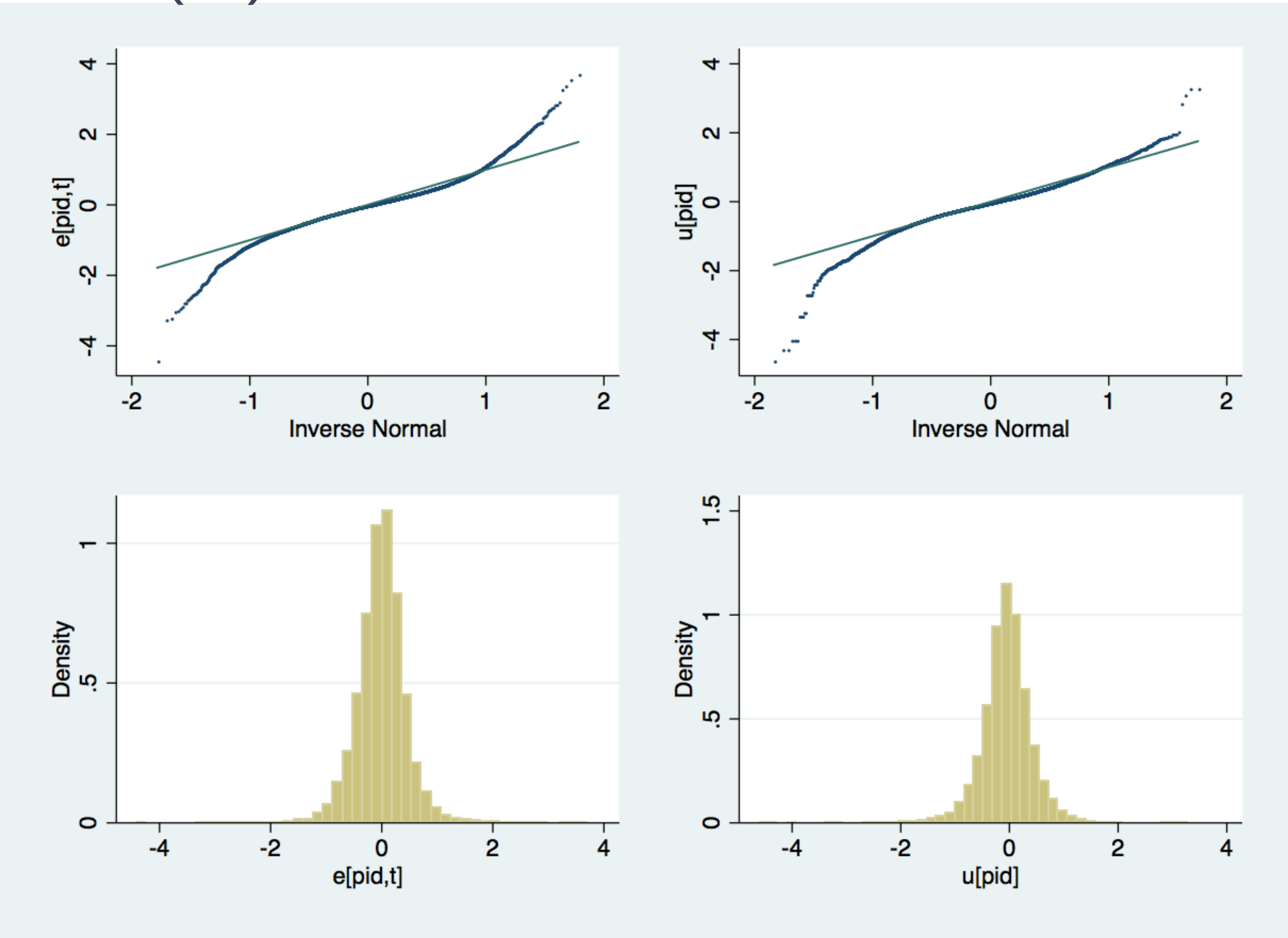
```
Error (err) and random/fixed effect (muf)

                          e[pid,t]
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -1.126145      -4.410871
 5%    -.7035062      -3.243076
10%    -.5120579      -3.200329       Obs               47526
25%    -.2408693      -3.010675       Sum of Wgt.       47526

50%      .009119                      Mean           3.68e-18
                        Largest       Std. Dev.      .4356894
75%     .2417121       3.291006
90%      .483008       3.393147       Variance       .1898252
95%     .6672699       3.571082       Skewness       .0185012
99%     1.174795       3.718479       Kurtosis       6.593565
```

No Skew in the fixed effects (U)

u[pid]

	Percentiles	Smallest		
1%	-1.276051	-4.60384		
5%	-.7557302	-4.274406		
10%	-.5452864	-4.274406	Obs	47526
25%	-.2725626	-4.004275	Sum of Wgt.	47526
50%	-.0379985		Mean	-.040353
		Largest	Std. Dev.	.4388224
75%	.2033705	2.86308		
90%	.4635404	3.11268	Variance	.1925651
95%	.66138	3.298158	Skewness	-.3128723
99%	1.091161	3.298158	Kurtosis	6.365477

Brazil (FE)


e[pid,t]
Inverse Normal
u[pid]
Inverse Normal
Density
e[pid,t]
Density
u[pid]

Stochastic Frontier Models: Convergence with Problematic Results

	BC92 + State FE	True RE	CRE: True RE + Mundlak	CRE: True RE + Soil + Rainfall (level) + Transport Cost + Von Thunen index + Mundlak
TC	0.0507	0.0012	0.0064	0.0065
TEC	-0.0355	0.0037	0.0010	0.0010
TFPC	0.0152	0.0049	0.0074	0.0075
Sigma2	0.4127	0.6331	0.2512	0.2509
Sigma_u2	0.1503	0.0778	0.0350	0.0365
Sigma_v2	0.2624	0.5552	0.2162	0.2144
Gamma	0.3642	0.1229	0.1392	0.1454
Lambda	0.5728	0.1401	0.1617	0.1702
N	50,247	50,247	47,631	47,631

- These models suggest no frontier (except with restrictive BC92)
  - Probably due to pseudo-panel data
  - Representative farms (300 farms on average)
- But we believe there is heterogeneity in performance much like what would be captured with a SFPF
  - The heterogeneity is across municipalities and sizes, not across individual farms
- We capture this heterogeneity with an average translog production function:
  - Can use fixed effects and weights
  - Recovers the entire distribution of TFPC
  - Permits decomposition of TFPC: TC and TEC
  - Permits analysis of marginal effects of policy on inputs, output and TFPC
  - Permits assessment of the role of specialization, or capital intensity, in raising TFPC

Empirical specification:

$$y_{it} = \alpha_i + \sum_{k=1}^{N} \beta_k x_{kit} + 0.5 \sum_{k=1}^{N} \sum_{h=1}^{N} \beta_{kh} x_{kit} x_{hit} + \sum_{k=1}^{N} \delta_k x_{kit} t + \lambda_1 t + 0.5 \lambda_{11} t^2 + \sum_{j=1}^{M} \theta_j z_{jit} + e_{it}$$

- y is the log of output; x is a vector of logged inputs; t is time; and z is a vector of policy variables and controls.
- α_i are fixed effects for approximately 16,000 representative farms.
- Subscript k=1,...,4 refers to family labor, land, purchased inputs, and capital.
- Subscripts j refer to a vector controls and policy variables that enter the production function linearly:
  - Temperature and rainfall shocks, technical assistance, credit and education.
- e_{it} is the error term.

- Note that TFPC, and all elasticities, are different for every observation and depend on the inputs

$$\frac{\partial y_{it}}{\partial t} = \lambda_1 + \lambda_{11} t + \sum_{k=1}^{N} \delta_k x_{kit}$$

$$e_{kit} = \frac{\partial y_{it}}{\partial x_{kit}} = \beta_k + \sum_{h=1}^{N} \beta_{kh} x_{hit} + \delta_k t$$

- This permits us to estimate the entire distribution of TFPC, elasticities, and other variables.

3A. Descriptive Statistics

Table 4

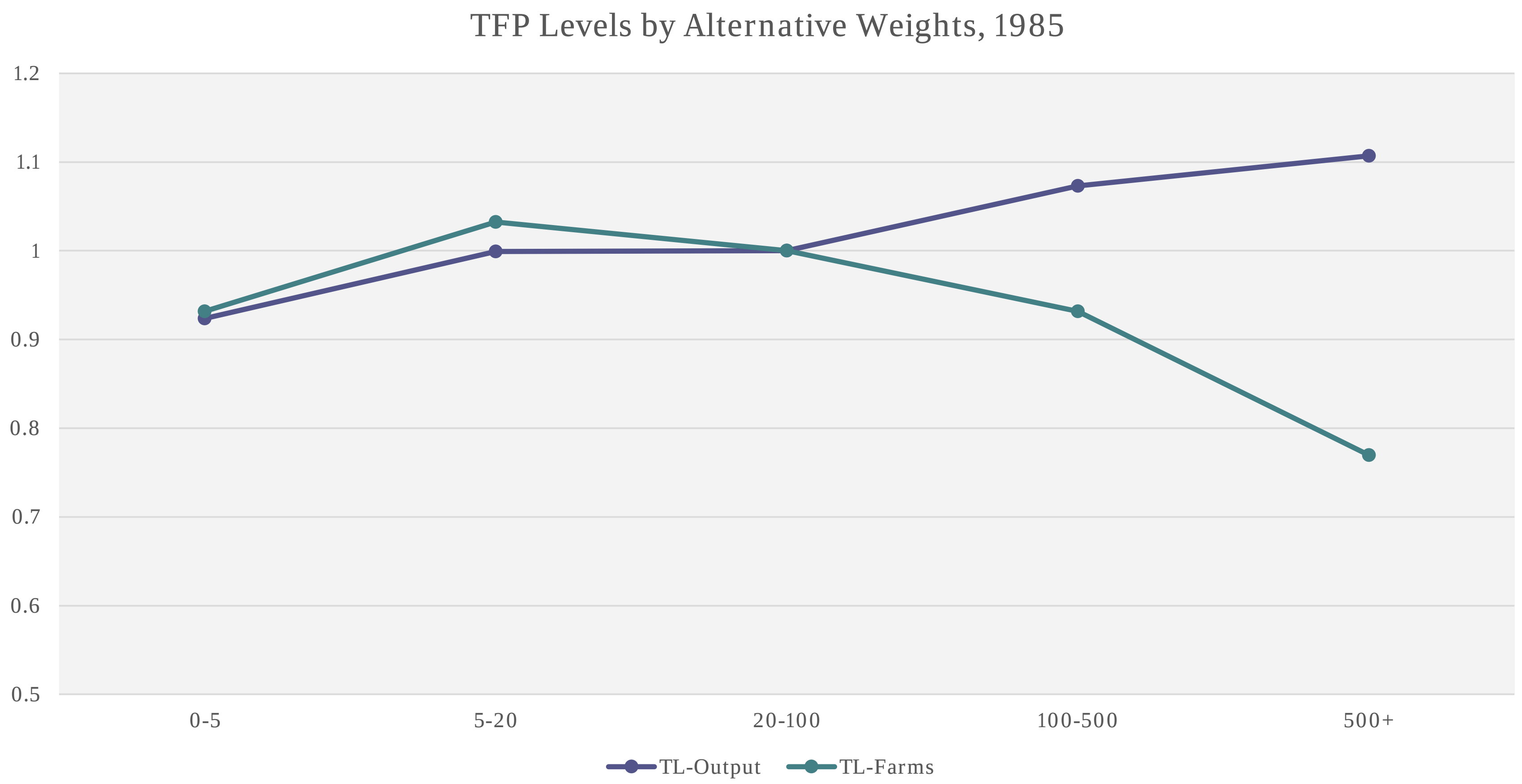
Percentage Change of Outputs and Inputs by Size: 1985-2006

Size (ha) / Brazil	Output	Land	Family Labor	Purchased Inputs			Capital Stock			
				Technology	Other	Total	Machines	Animals	Trees	Total
0-5	100	-37	-24	58	132	104	113	73	121	100
5-20	76	-27	-28	94	88	90	43	82	41	47
20-100	53	-23	-26	63	60	61	1	69	14	7
100-500	43	-21	-31	158	65	96	-10	26	16	-3
500-	174	-4	7	580	193	306	16	35	62	25
Brazil	85	-15	-26	223	110	149	10	50	29	17

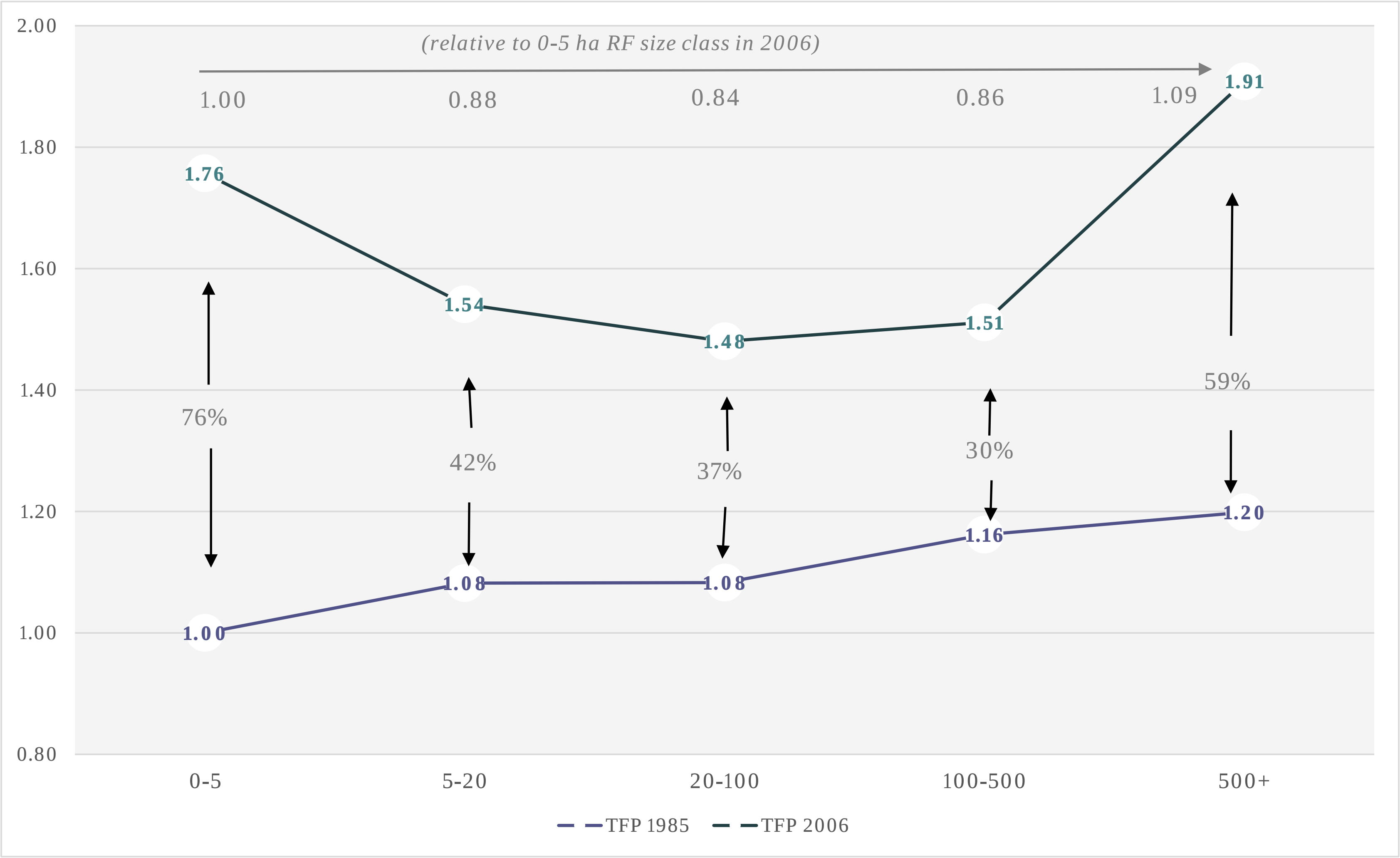
3B. Econometric Results

- Model for 1985 TFP
  - Translog w/Municipal FEs, CRS and Size Dummies
  - Role of weights
- Model for 1985-2006 TFP growth
  - Various specifications: role of controls and policy variables

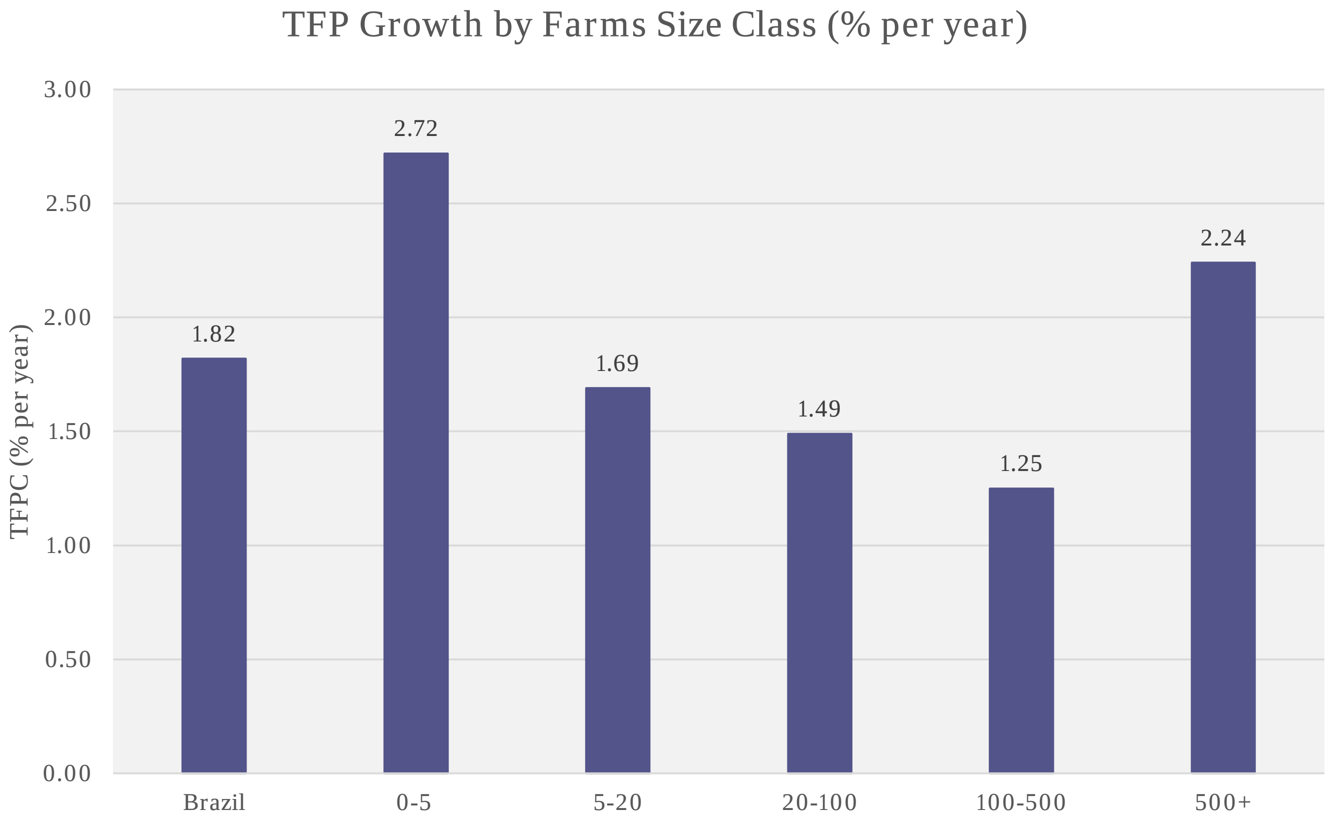
1985 TFP Levels & Role of Weights

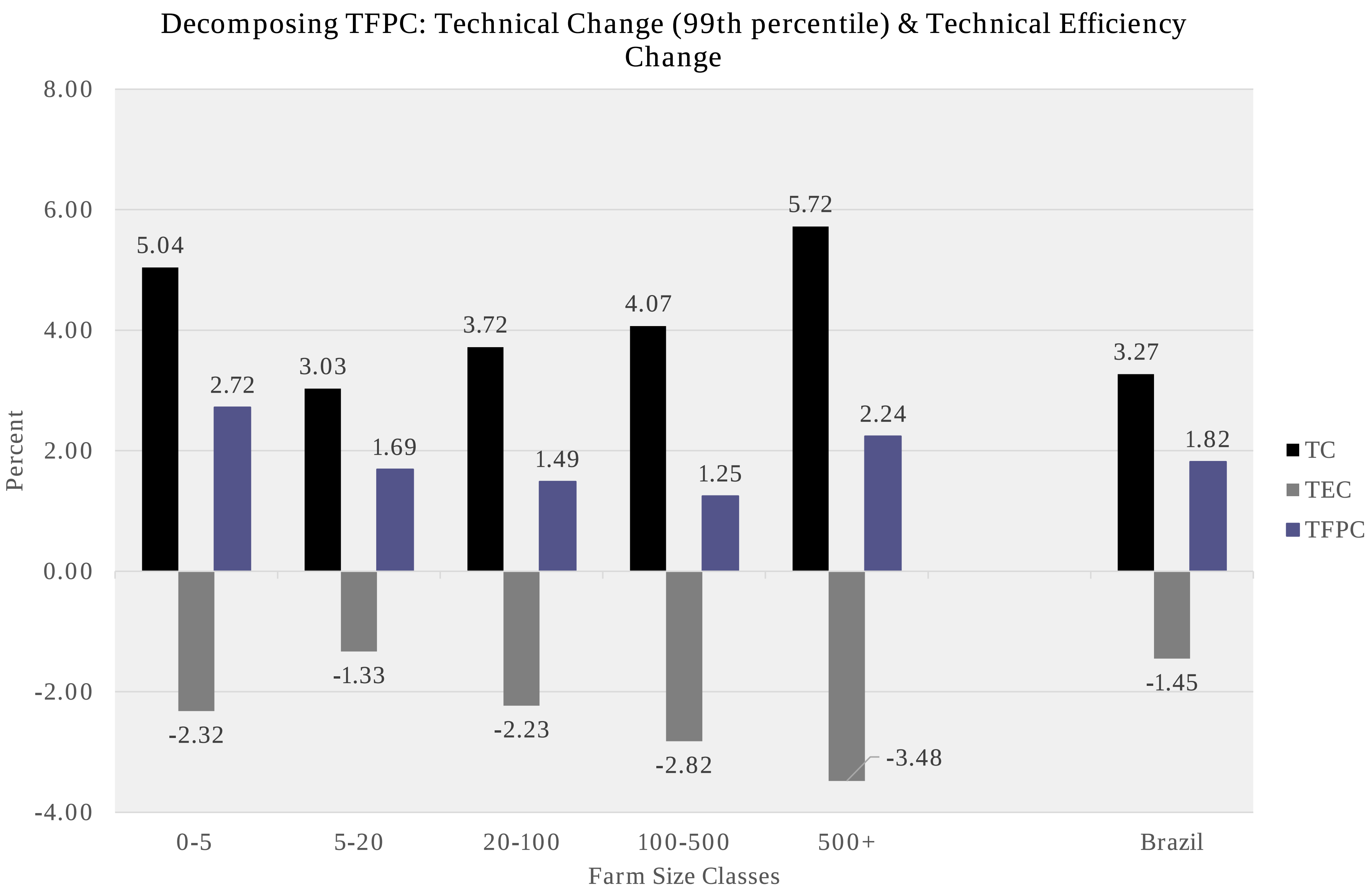


Brazil's Agricultural TFP by Farm Size, 1985 & 2006

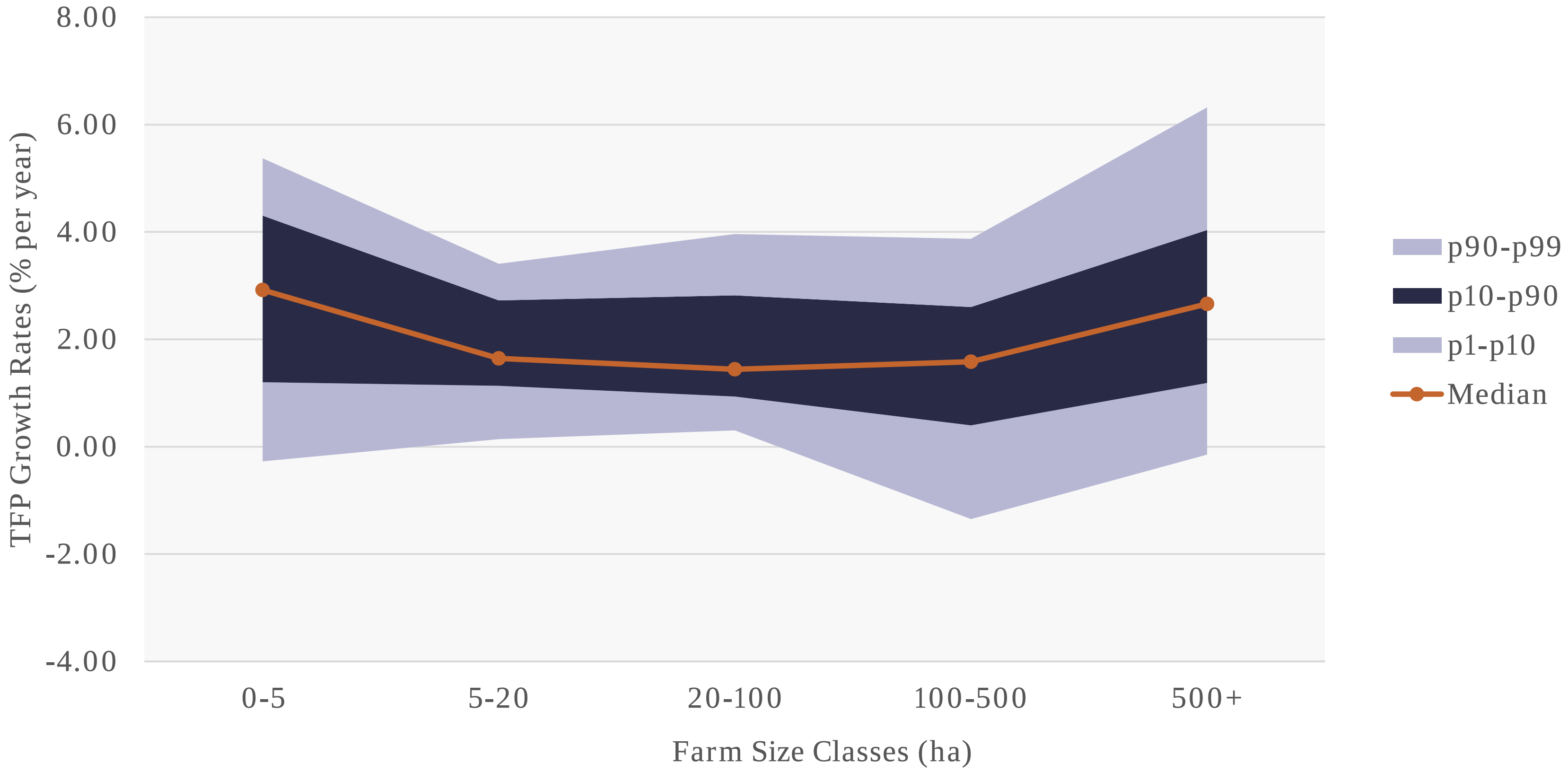


Mean TFPC by Farm Size

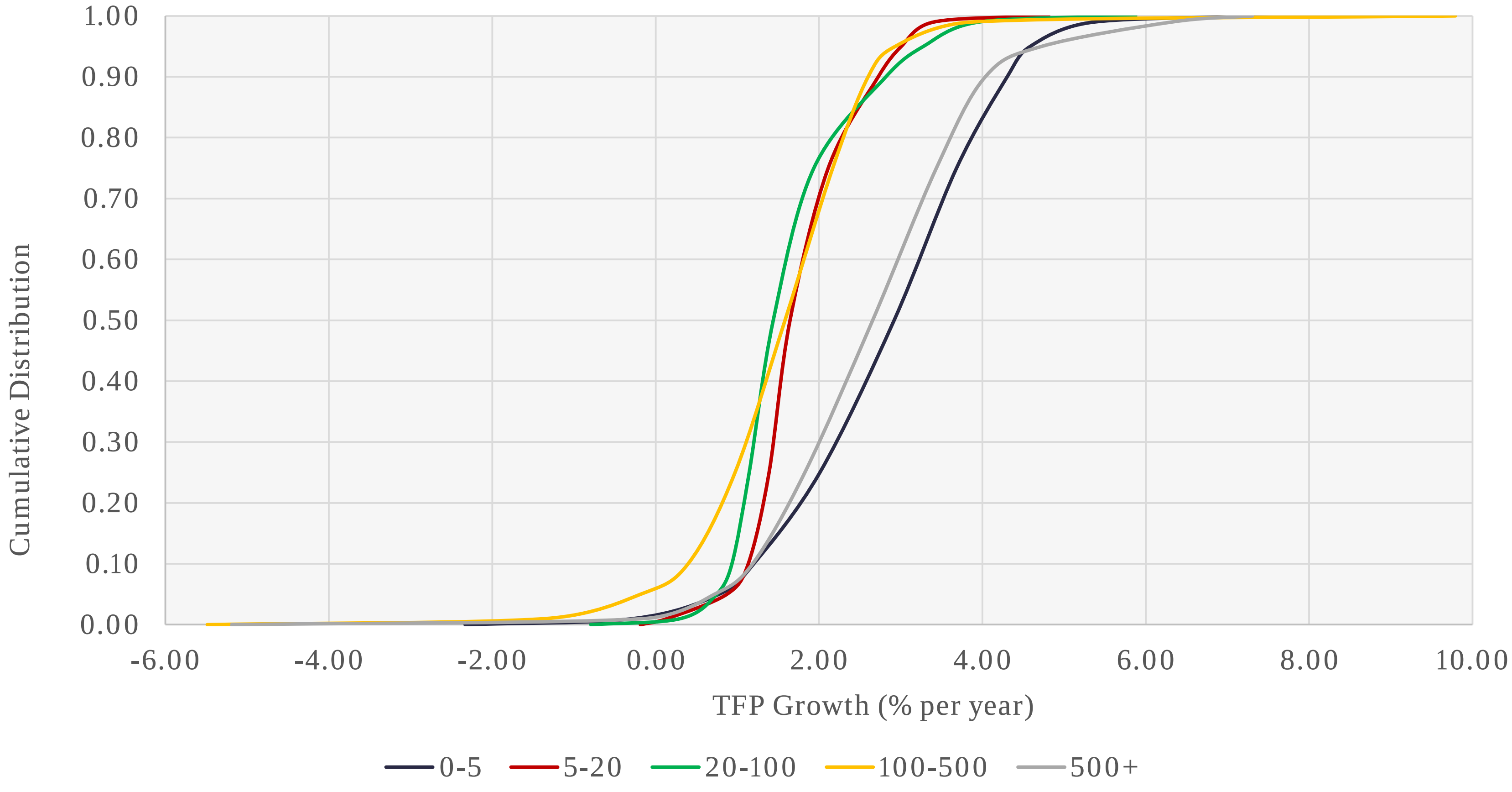



Decomposing TFPC: Technical Change (99th percentile) & Technical Efficiency Change
8.00
6.00
4.00
2.00
0.00
-2.00
-4.00
Percent
5.04
2.72
-2.32
3.03
1.69
-1.33
3.72
1.49
-2.23
4.07
1.25
-2.82
5.72
2.24
-3.48
3.27
1.82
-1.45
0-5
5-20
20-100
100-500
500+
Brazil
Farm Size Classes
TC
TEC
TFPC

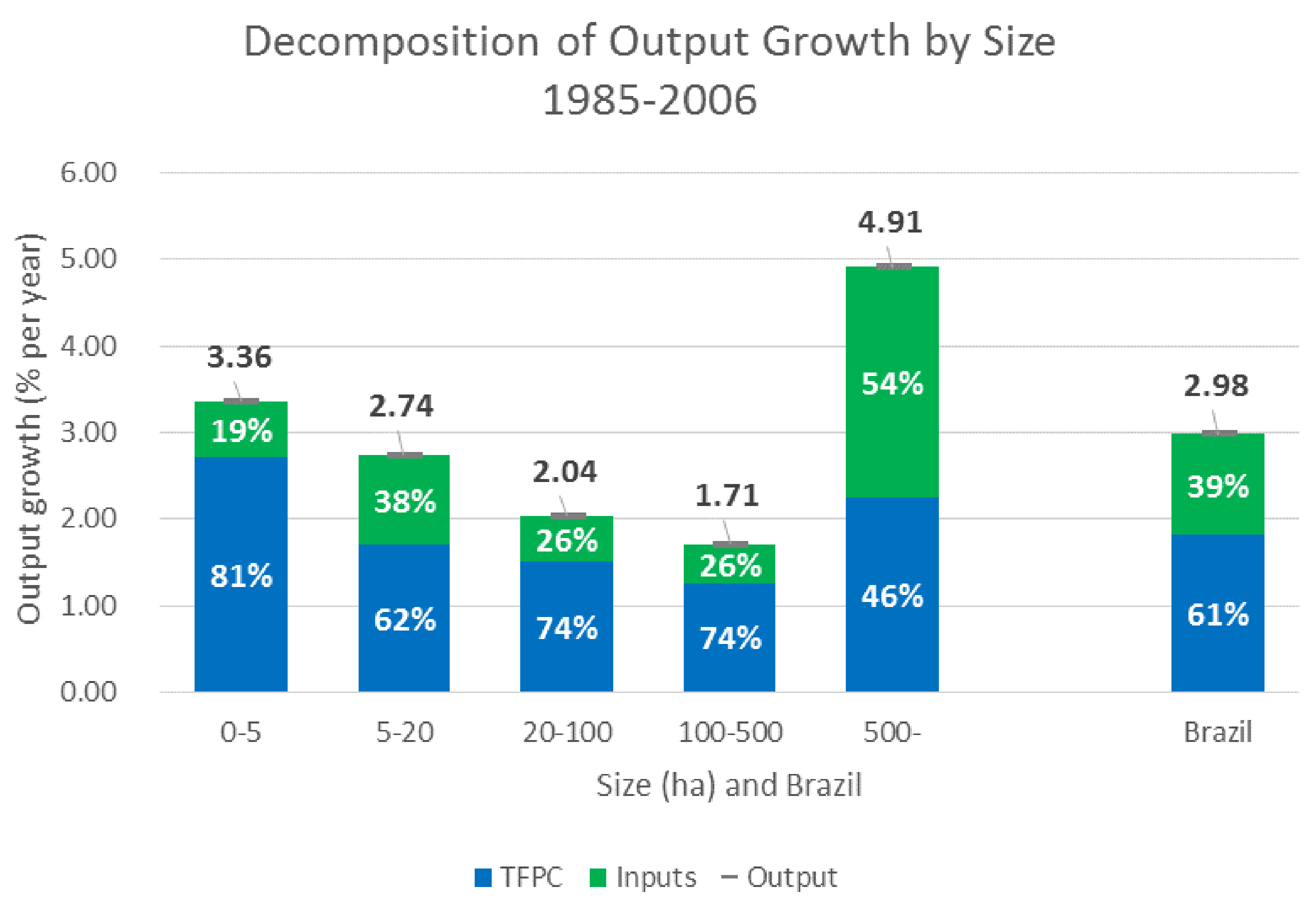
Dispersion of TFPC by Farm Size Class, 1985-2006



Cumulative Distribution of TFP growth in Brazil by Farm Size, 1985-2006

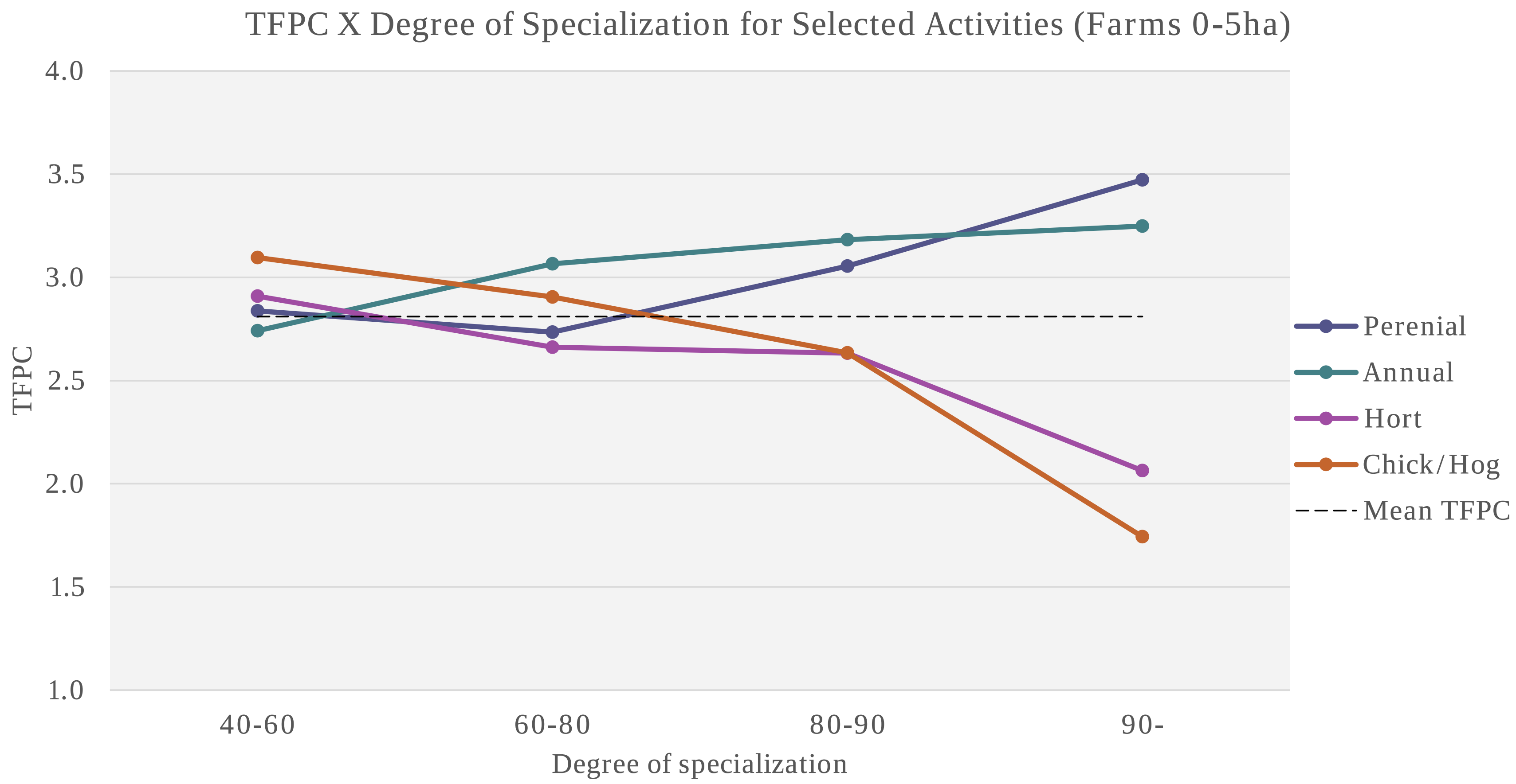


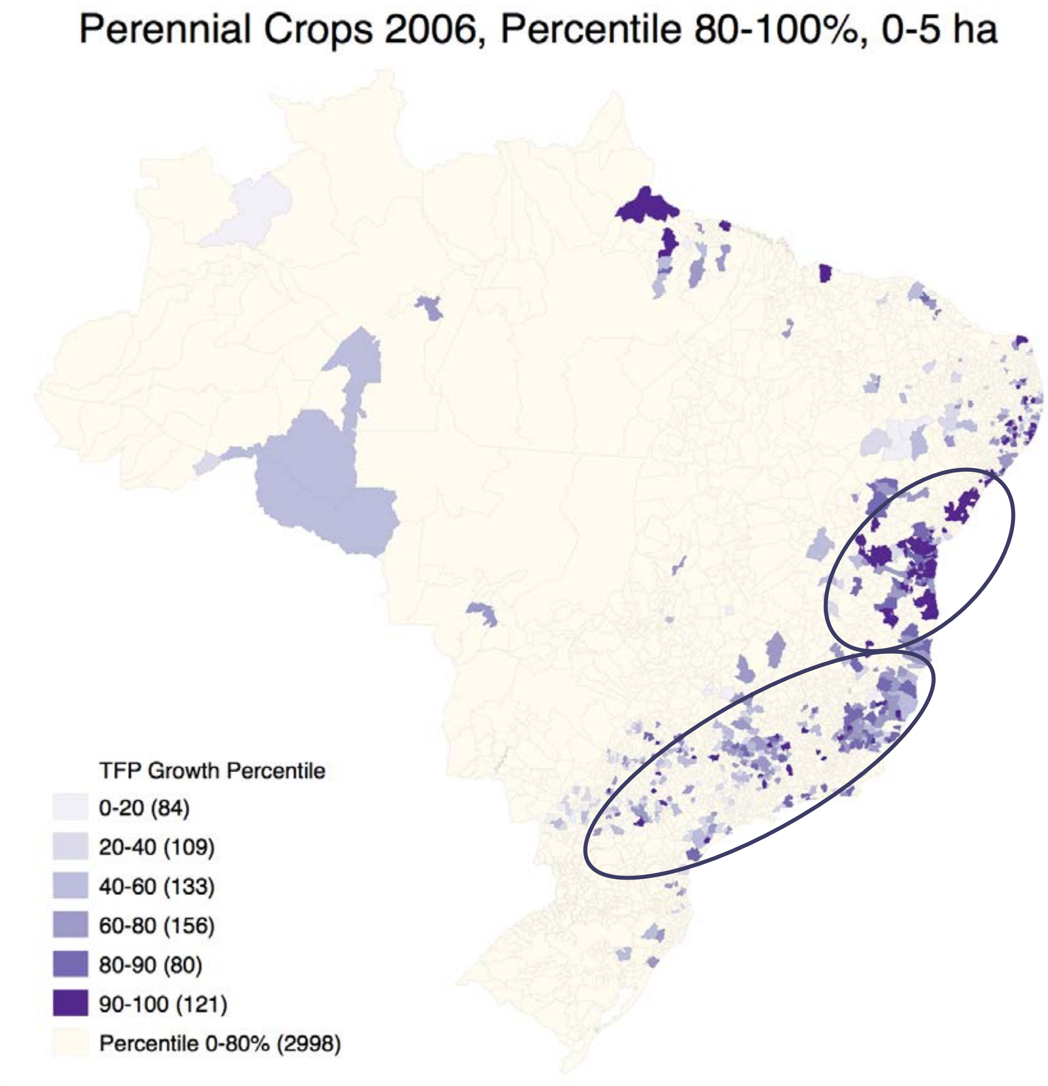
The Importance of TFPC for Output Growth



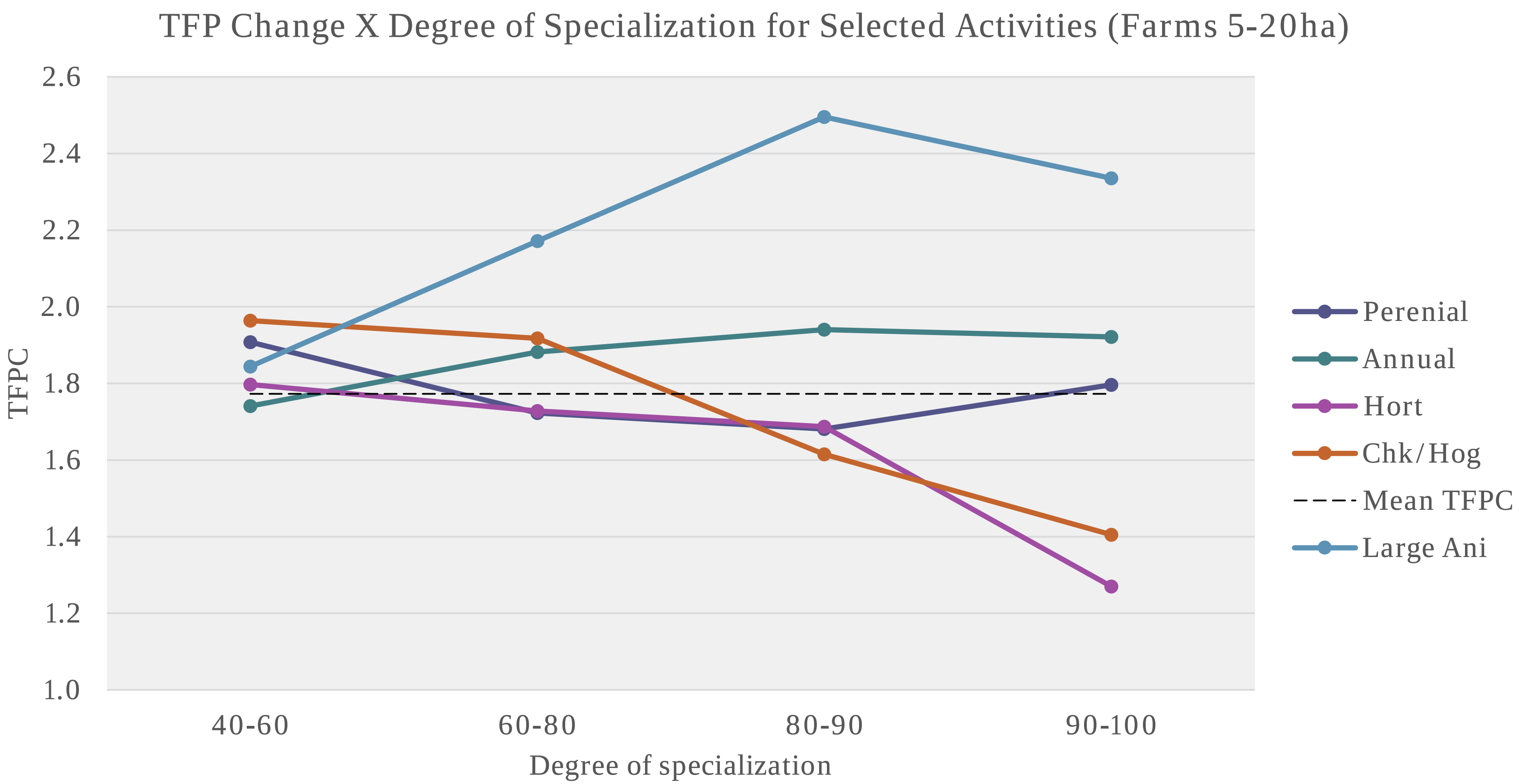
TFPC by Degree of Specialization

Specialization: Farms 0-5ha

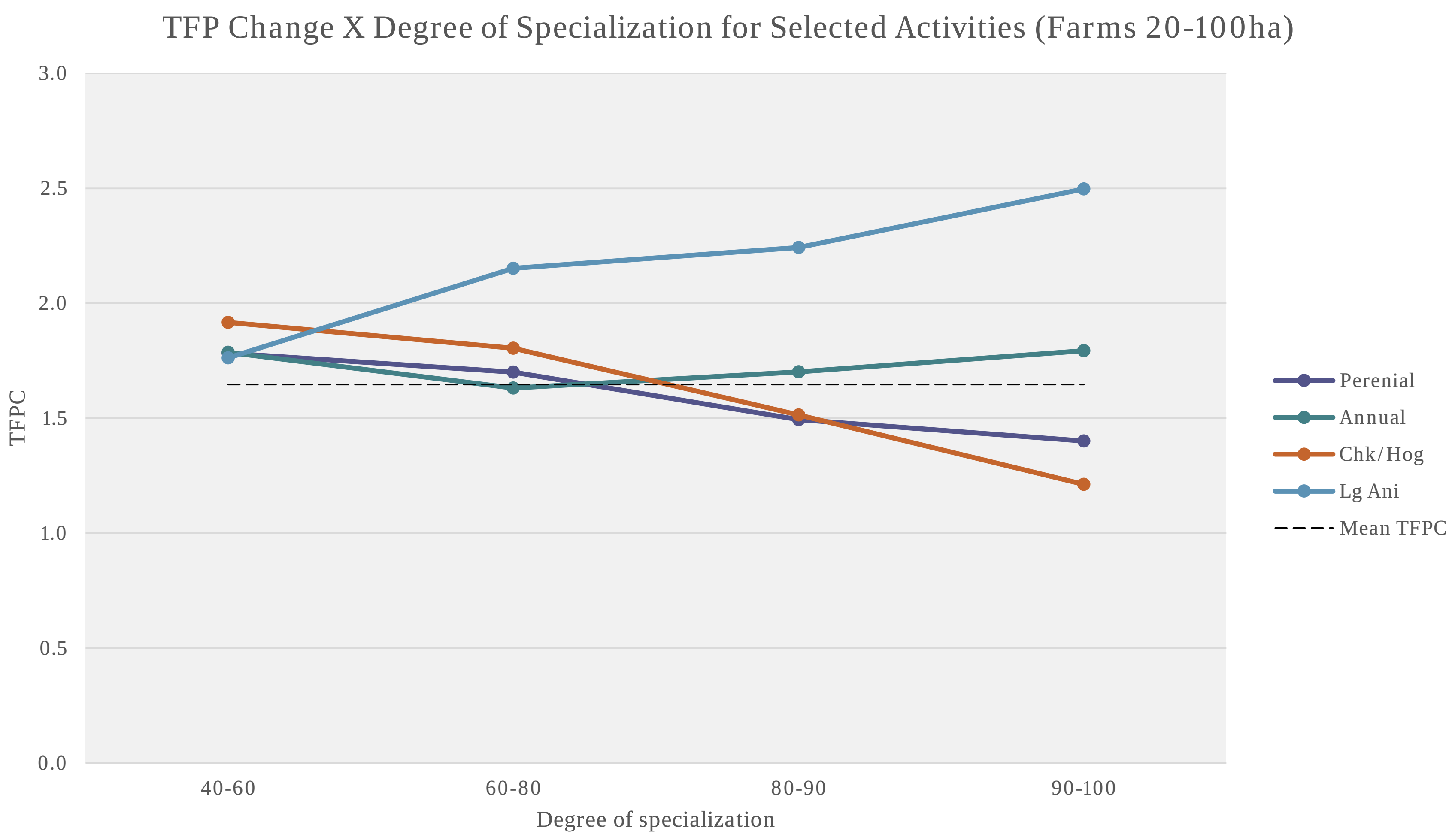



Perennial Crops 2006, Percentile 80-100%, 0-5 ha
TFP Growth Percentile
0-20 (84)
20-40 (109)
40-60 (133)
60-80 (156)
80-90 (80)
90-100 (121)
Percentile 0-80% (2998)

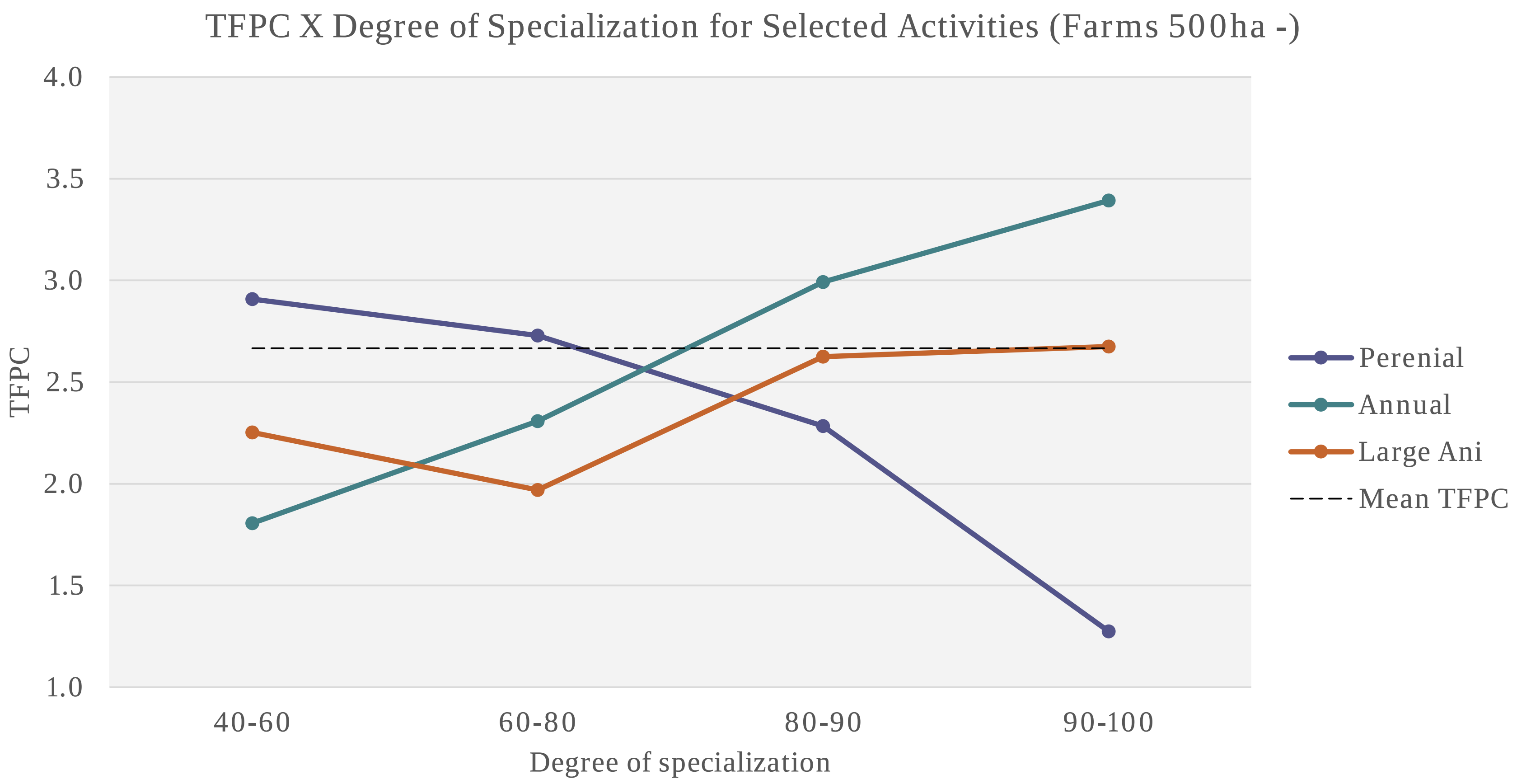
Specialization: Farms 5-20ha

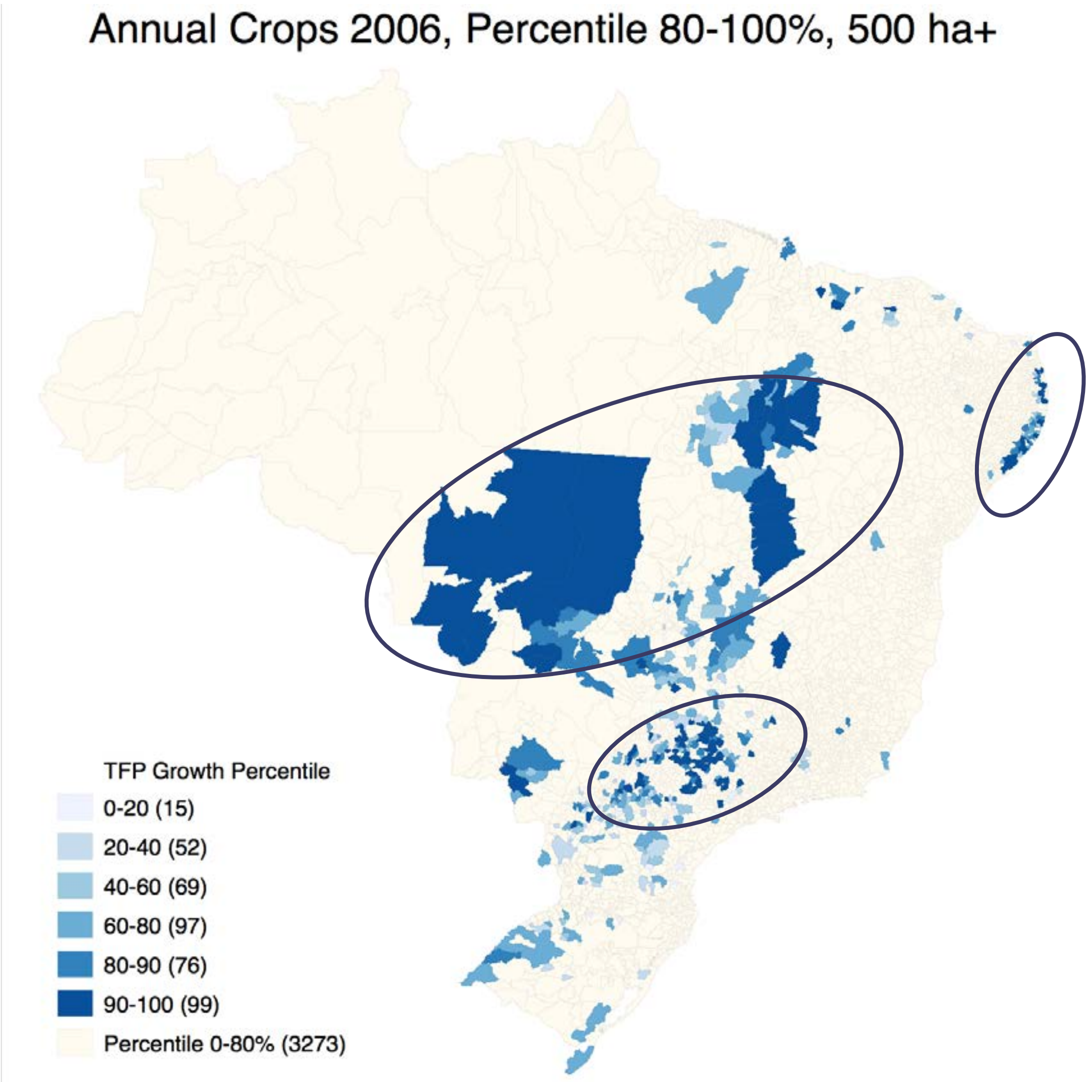


Specialization: Farms 20-100ha

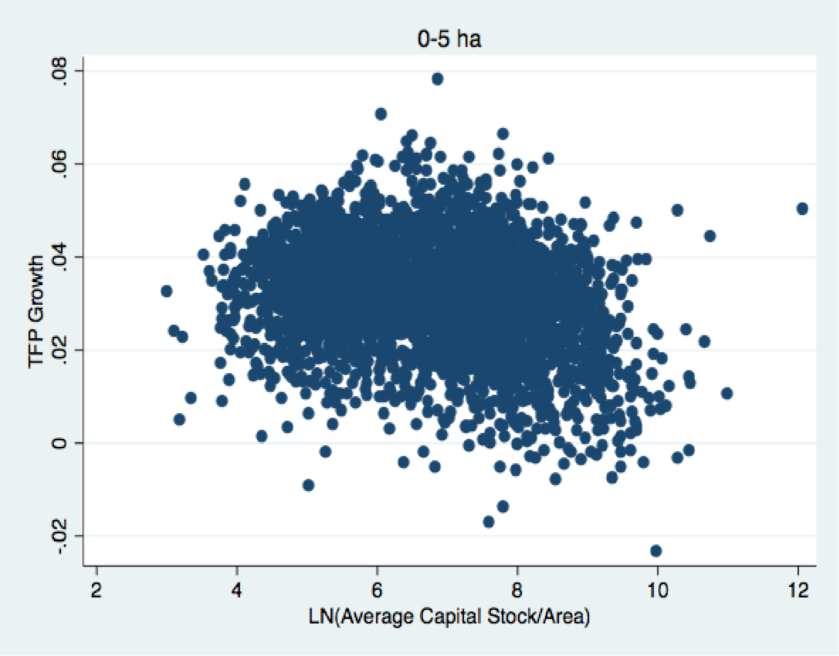


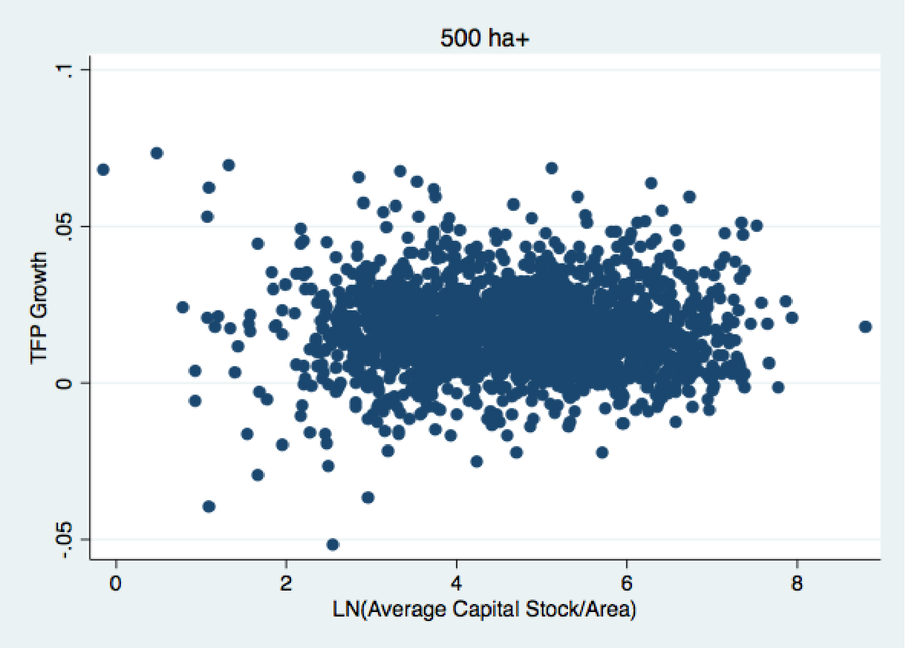
Specialization: Farms 500-ha



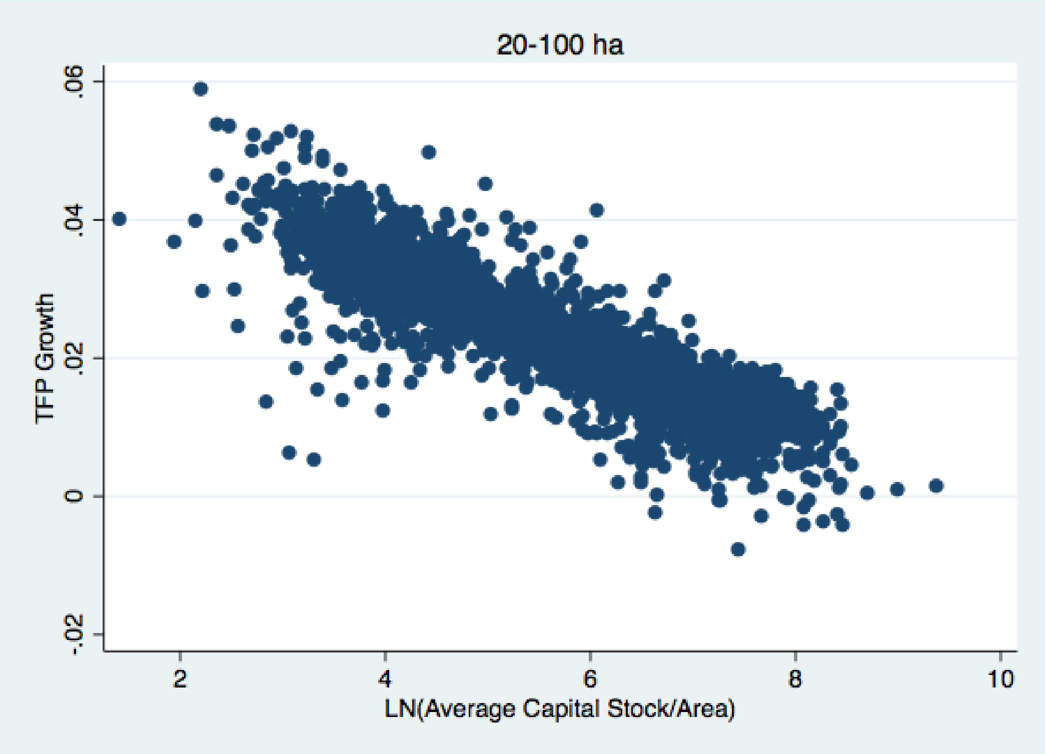

Annual Crops 2006, Percentile 80-100%, 500 ha+
TFP Growth Percentile
0-20 (15)
20-40 (52)
40-60 (69)
60-80 (97)
80-90 (76)
90-100 (99)
Percentile 0-80% (3273)

For small and large farms, K intensity (per ha) is fairly neutral w.r.t. TFPC:





But for mid-sized farms, k-intensity is associated with slower TFPC: Why?

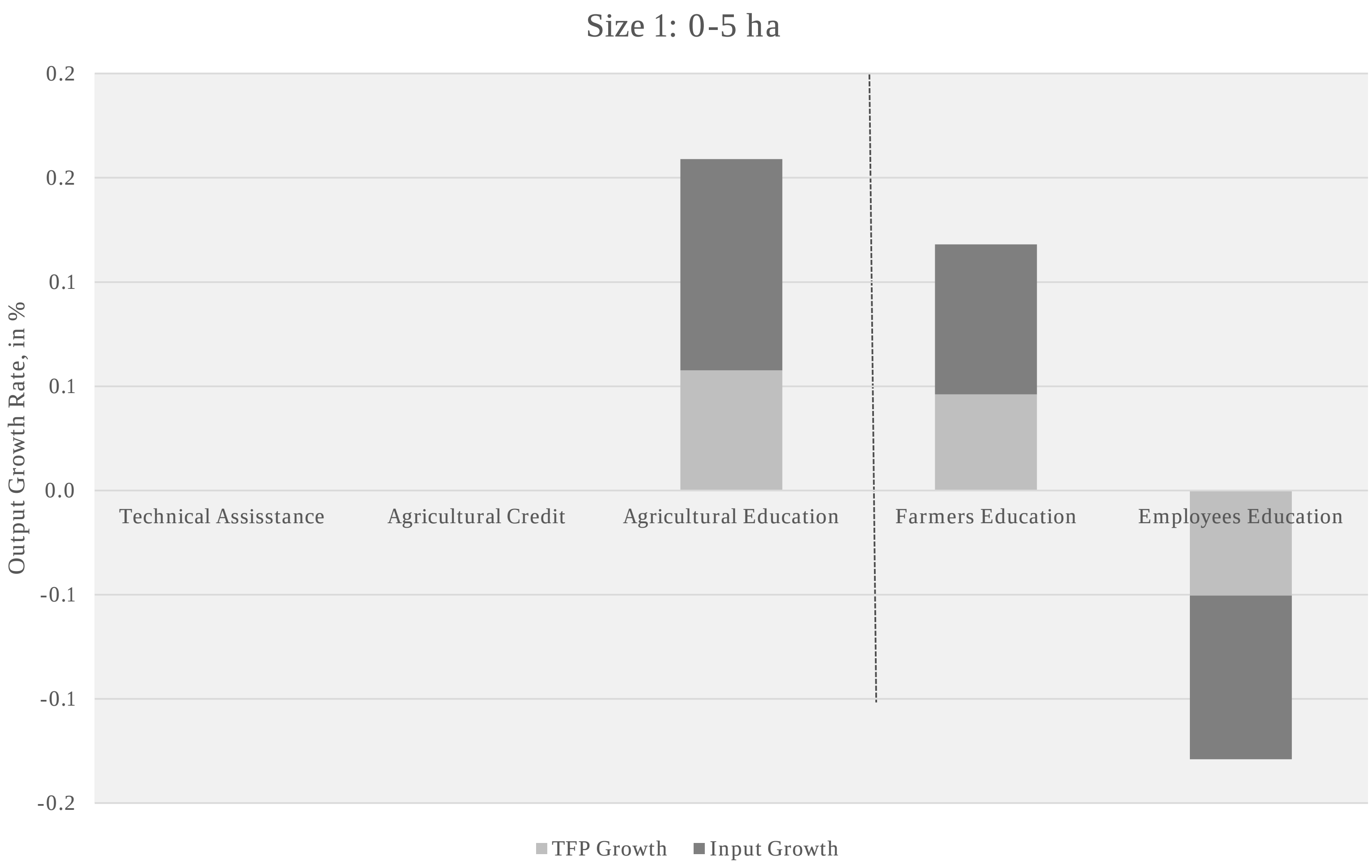


How much of TFPC do policies explain?

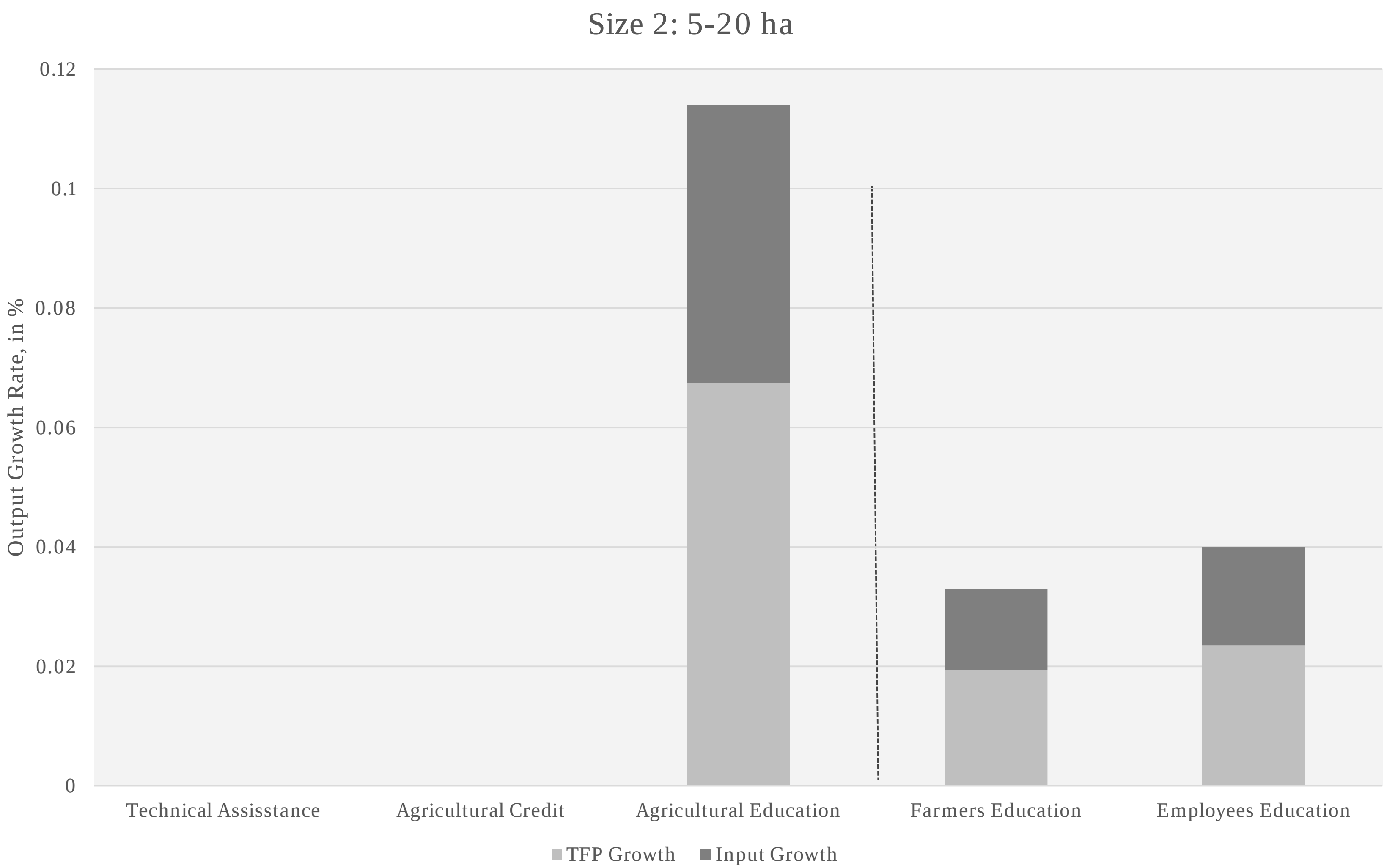
APF TL Brasil

	APF	Rainfall & temperature shocks	IV Tech Assistance (share)	IV Credit (share)	Lagged Ag. Schooling (years)	Lagged Farmers' Schooling (years)	Lagged Employees' Schooling (years)	All Policies (Ag. Schooling)	All Policies (F&E Schooling)
Tech. Assistance			0.096**					0.086**	0.092**
Credit				-0.048				-0.073*	
Ag. Schooling					0.15**			0.146**	
Farmers' Schooling						0.067**			0.053**
Farm Employees Schooling							0.049**		0.041**
TFPC (% per year)	1.92	1.82	1.74	1.83	1.76	1.80	1.81	1.69	1.71
								0.93	0.94
-	-	-	-	-	-	-	-	-	-

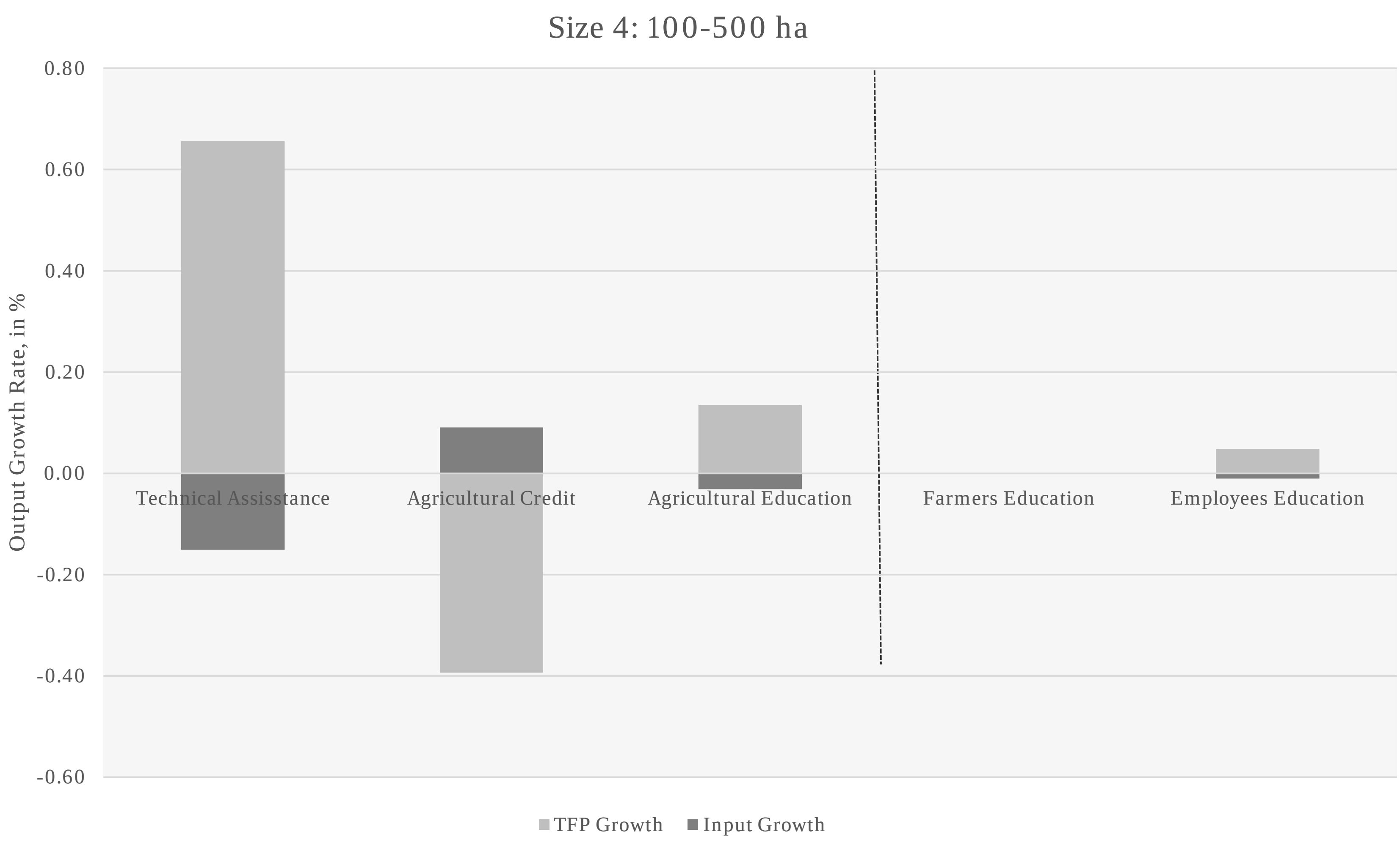
Marginal Policy Effects on Output, Input, and TFP Growth, 1985-2006



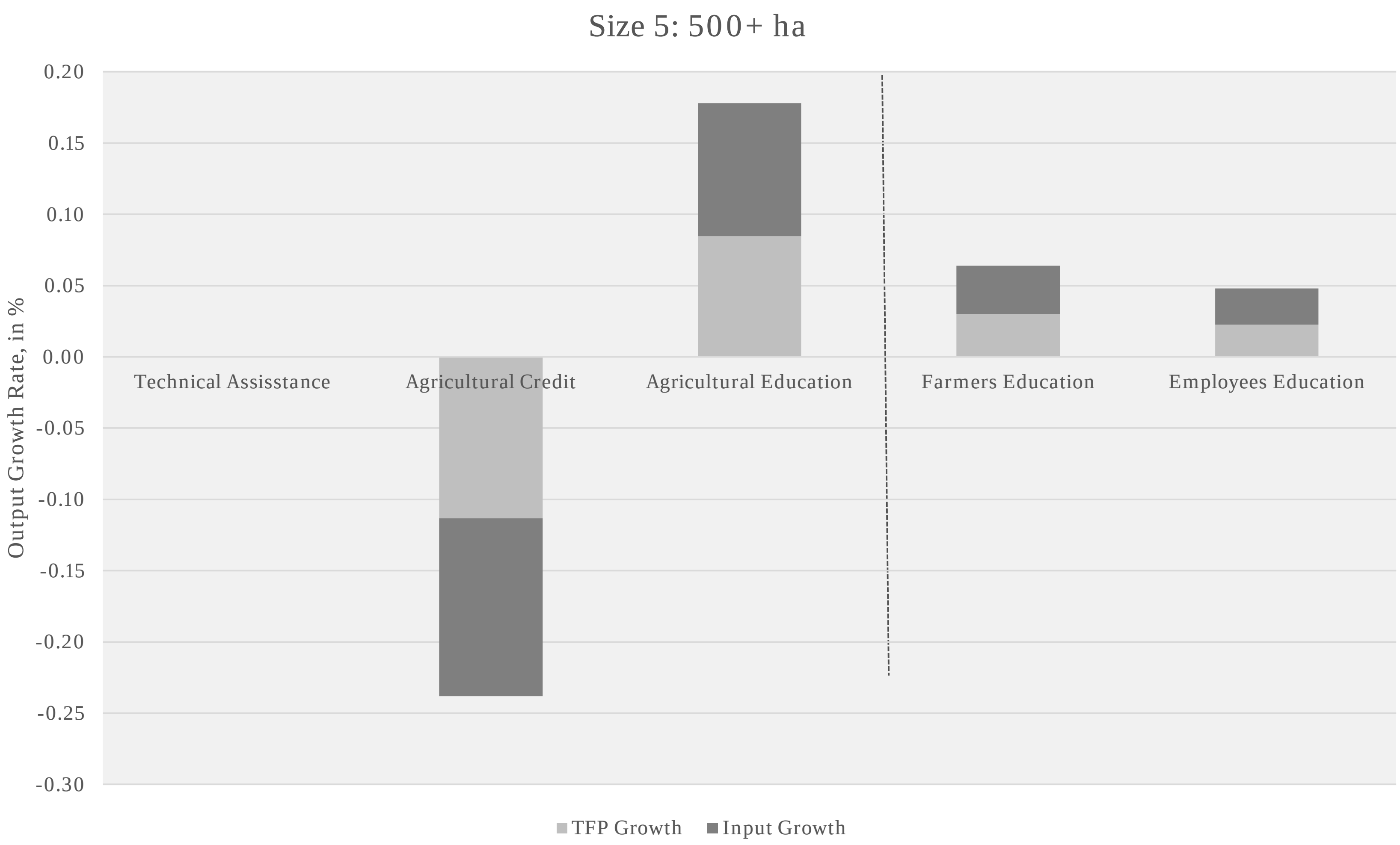
Marginal Policy Effects on Output, Input, and TFP Growth, 1985-2006



Marginal Policy Effects on Output, Input, and TFP Growth, 1985-2006



Marginal Policy Effects on Output, Input, and TFP Growth, 1985-2006



4. Conclusions, next steps, remaining challenges

- TFP growth has been fastest in the smallest and largest size classes, yet the middle has lagged.
  - Does the strong negative correlation between TFP growth and capital intensity provide a clue?
- Specialization in annual crops is associated with above average TFP growth in all size classes.
- Surprisingly, for farms 0-20ha, specialization in chickens/pigs/horticulture is **not** associated with faster TFP growth.
    - Too input intensive?
- Specialization in large animals/milk was important for TFP for farms 5-20ha and 20-100ha.

Conclusions

- Across all farms sizes, education boosts outputs more than inputs, leading to TFP growth.
  - However, for the smallest farmers, we see a tradeoff between education of farmers vs. employees. Opportunity cost?
- For Brazil as a whole, technical assistance has a positive impact on TFP, and credit a negative impact. Largely driven by farms 100-500ha and 500-ha.
  - Unique, size-related finding or an anomaly?
- Brazil could still greatly boost agricultural supply if it could improve the productivity performance of its lagging medium sized farms.