Migration, Specialization, and Trade: Evidence from the Brazilian March to the West

November 29, 2019

1 Thailand

These are regressions of migration flows on previous stock of workers for Thailand, using census 1970 and 1980. The geographical unit are provinces. Crops on these census are sufficiently disaggregated to do the analysis. We keep data on people that are farmers that work on the following crops: rice, corn, rubber, cassava, coconut, wood, fish, and hunting. For migration from origin to destination province, the notion of origin that we use is the province where the person was born. As in the original paper, we exclude cases where the origin province are equal to the destination province in the regressions. In Table 1 we construct L_{ikt-1} with the 1970 census, and L_{ijkt} with the 1980 census. In Table 2 we construct both L_{ikt-1} and L_{ijkt} from the 1970 census. In Table 3 we construct both L_{ikt-1} and L_{ijkt} from the 1970 census. Finally, in Table 4 we use both 1970 and 1980 census to construct both L_{ikt-1} and L_{ijkt}.

2 Thailand, balance check between 1970 and 1980 census

The 1970 census had a sample of 2% for a total of 772169 people, where district was the smallest geography in the sampling design. The 1980 census had a sample of 1% for a total of 388141 people, where provinces were in this case the smallest geography in the sampling design. In the ocassions where there is production of a crop in both census, the number are sufficiently close such that they could reflect structural change and not errors of some kind. There are ocassions where the production of some crop in some province disappears in or the production of a new crop appears in 1980, but in the majority of cases production of a crop happens in both census.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Migration flows, OLS									
Farmers in origin	0.058	0.058	0.148***	0.048	0.048	0.145***	0.053	0.053	0.161***
	(0.045)	(0.045)	(0.019)	(0.046)	(0.046)	(0.019)	(0.042)	(0.042)	(0.019)
R^2	0.833	0.833	0.250	0.835	0.835	0.257	0.832	0.832	0.251
Obs	921	921	921	871	871	871	982	982	982
Migration flows, PPML									
Farmers in origin	0.119***	0.104**	0.536***	0.115***	0.097**	0.561***	0.115***	0.116***	0.544***
	(0.025)	(0.045)	(0.092)	(0.026)	(0.044)	(0.095)	(0.023)	(0.039)	(0.087)
R^2	-	-	-	-	-	-	-	-	-
Obs	18559	921	18559	18559	871	18559	18559	982	18559
Dest-Crop-Year FE	Y	Y		Y	Y		Y	Y	
Orig-Dest-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Without zeros		Y			Y			Y	
Men HH heads				Y	Y	Y			
Men HH heads, 20-65 y/o							Y	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors are clustered at the destination-crop-year level, and are reported in parentheses. An observation is a cell at the origin-destination-crop-year level. Columns (1), (2), and (3) are based on a sample of 30-65 years old migrants. In columns (4), (5), and (6) the sample is comprised by men between 30-65 years old. In columns (7), (8), and (9) the sample is comprised by men between 20-65 years old. The covariate is the log of of agricultural workers in the same activity in the region of origin. The dependent variable is the log of migrant agricultural workers from an origin to a destination region working in an activity. The covariate is based on the 1970 census, while the dependent variable is based on the 1980 census. We exclude non-migrants from the sample.

Table 1: Regressions, 1970 is lag, 1980 is present

Table 2: Regressions, 1970 is both lag and present

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Migration flows, OLS									
Farmers in origin	0.057	0.057	0.093***	0.061	0.061	0.104***	0.077	0.077	0.114***
	(0.054)	(0.054)	(0.027)	(0.050)	(0.050)	(0.027)	(0.049)	(0.049)	(0.030)
R^2	0.869	0.869	0.526	0.862	0.862	0.526	0.869	0.869	0.518
Obs	839	839	839	803	803	803	895	895	895
Migration flows, PPML									
Farmers in origin	0.145***	0.210***	0.506***	0.133***	0.225***	0.505***	0.132***	0.218***	0.505***
	(0.032)	(0.073)	(0.114)	(0.031)	(0.066)	(0.114)	(0.029)	(0.063)	(0.116)
R^2	-	-	-	-	-	-	-	-	-
Obs	18559	839	18559	18559	803	18559	18559	895	18559
Dest-Crop-Year FE	Y	Y		Y	Y		Y	Y	
Orig-Dest-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Without zeros		Y			Y			Y	
Men HH heads				Y	Y	Y			
Men HH heads, 20-65 y/o							Y	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors are clustered at the destination-crop-year level, and are reported in parentheses. An observation is a cell at the origin-destination-crop-year level. Columns (1), (2), and (3) are based on a sample of 30-65 years old migrants. In columns (4), (5), and (6) the sample is comprised by men between 30-65 years old. In columns (7), (8), and (9) the sample is comprised by men between 20-65 years old. The covariate is the log of of agricultural workers in the same activity in the region of origin. The dependent variable is the log of migrant agricultural workers from an origin to a destination region working in an activity. Both the covariate and the dependent variable are based on the 1970 census. We exclude non-migrants from the sample.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Migration flows, OLS									
Farmers in origin	0.120***	0.120***	0.118***	0.121***	0.121***	0.113***	0.122***	0.122***	0.141***
	(0.038)	(0.038)	(0.020)	(0.043)	(0.043)	(0.020)	(0.042)	(0.042)	(0.020)
R^2	0.816	0.816	0.226	0.817	0.817	0.237	0.830	0.830	0.230
Obs	1012	1012	1012	960	960	960	1087	1087	1087
Migration flows, PPML									
Farmers in origin	0.199***	0.178***	0.626***	0.191***	0.183***	0.633***	0.187***	0.207***	0.624***
	(0.028)	(0.035)	(0.098)	(0.029)	(0.038)	(0.098)	(0.028)	(0.035)	(0.092)
R^2	-	-	-	-	-	-	-	-	-
Obs	20770	1012	20770	20770	960	20770	20770	1087	20770
Dest-Crop-Year FE	Y	Y		Y	Y		Y	Y	
Orig-Dest-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Without zeros		Y			Y			Y	
Men HH heads				Y	Y	Y			
Men HH heads, 20-65 y/o							Y	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors are clustered at the destination-crop-year level, and are reported in parentheses. An observation is a cell at the origin-destination-crop-year level. Columns (1), (2), and (3) are based on a sample of 30-65 years old migrants. In columns (4), (5), and (6) the sample is comprised by men between 30-65 years old. In columns (7), (8), and (9) the sample is comprised by men between 20-65 years old. The covariate is the log of of agricultural workers in the same activity in the region of origin. The dependent variable is the log of migrant agricultural workers from an origin to a destination region working in an activity. Both the covariate and the dependent variable are based on the 1980 census. We exclude non-migrants from the sample.

Table 3: Regressions, 1980 is both lag and present

Table 4: Regressions, 1970 + 1980

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Migration flows, OLS									
Farmers in origin	0.092***	0.092***	0.104***	0.093***	0.093***	0.108***	0.101***	0.101***	0.127***
	(0.032)	(0.032)	(0.017)	(0.033)	(0.033)	(0.017)	(0.032)	(0.032)	(0.018)
R^2	0.854	0.854	0.462	0.853	0.853	0.466	0.860	0.860	0.461
Obs	1851	1851	1851	1763	1763	1763	1982	1982	1982
Migration flows, PPML									
Farmers in origin	0.178***	0.187***	0.570***	0.169***	0.196***	0.572***	0.166***	0.210***	0.568***
	(0.021)	(0.033)	(0.078)	(0.021)	(0.033)	(0.079)	(0.020)	(0.031)	(0.077)
R^2	-	-	-	-	-	-	-	-	-
Obs	39329	1851	39329	39329	1763	39329	39329	1982	39329
Dest-Crop-Year FE	Y	Y		Y	Y		Y	Y	
Orig-Dest-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Without zeros		Y			Y			Y	
Men HH heads				Y	Y	Y			
Men HH heads, 20-65 y/o							Y	Y	Y

Notes: * / ** / *** denotes significance at the 10 / 5 / 1 percent level. Standard errors are clustered at the destination-crop-year level, and are reported in parentheses. An observation is a cell at the origin-destination-crop-year level. Columns (1), (2), and (3) are based on a sample of 30-65 years old migrants. In columns (4), (5), and (6) the sample is comprised by men between 30-65 years old. In columns (7), (8), and (9) the sample is comprised by men between 20-65 years old. The covariate is the log of of agricultural workers in the same activity in the region of origin. The dependent variable is the log of migrant agricultural workers from an origin to a destination region working in an activity. Both the covariate and the dependent variable are based on both the 1970 and 1980 census. We exclude non-migrants from the sample.

3 Thailand, regressions with migration flows at the person level [[THIS NEEDS MORE WORK]]

XXX

4 South Africa

These are the regressions for the 2007 census of South Africa. Location is at the province level, there are 9 provinces. There are 18 crops: grain and staple farming, vegetable farming, nursery farming, fruit farming, vineyards, sugar cane, cotton, cattle, chicken, horse, dairy, sheep, ostrich, goat, ocean fishing, inland fishing, fish farms, and mixed farming. For regression, we calculate average earnings by origin-destination-crop. The origin variable is determined by which province the person was born. In Table 5 we include the regressions of earnings on L_{ikt-1}. The first subtable is the baseline regression where earnings is calculated for head of HHs between 30 and 65 years old. The second subtable is for men head of HHs between 30 and 65 years old. The third subtable is for men head of HHs between 20 and 65 years old. Each subtable has four columns. The first column is OLS, the second one is PPML, the third one is PPML with only destination-crop FEs, and the fourth one is OLS with only destination-crop FEs. In Table 6 we present the same, but for migration flows. The only difference is that there is an extra middle column in each subtable, which is PPML but excluding rows where no migration flows into the destination province. Results are somewhat promising for migration flows, but find nothing for earnings.

Table 5: Regression (10), baseline + HH men + HH men 20-65

	(1)	(2)	(3)	(4)
L_iktlag_log	-0.117	-0.0673	-0.0434	-0.0804
	(0.145)	(0.164)	(0.104)	(0.0904)
Observations	265	266	284	283
R2	0.534			0.279
Pseudo R2		0.754	0.435	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(1)	(2)	(3)	(4)
L_iktlag_log	-0.172	-0.303*	-0.0237	-0.0969
	(0.150)	(0.171)	(0.111)	(0.0971)
Observations	242	242	258	258
R2	0.525			0.263
Pseudo R2		0.736	0.437	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(1)	(2)	(3)	(4)
L_iktlag_log	-0.118	-0.235	-0.00852	-0.0195
	(0.136)	(0.161)	(0.110)	(0.0948)
Observations	270	270	284	284
R2	0.530			0.274
Pseudo R2		0.746	0.451	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Regression (11), baseline + HH men + HH men 20-65

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0765	0.0760	0.0666	0.271***	0.194**
	(0.0915)	(0.0570)	(0.0559)	(0.0976)	(0.0764)
Observations	288	784	288	795	301
R2	0.779				0.380
Pseudo R2		0.842	0.880	0.436	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.116	0.103*	0.0684	0.247**	0.182**
	(0.0985)	(0.0555)	(0.0524)	(0.100)	(0.0858)
Observations	264	737	264	765	276
R2	0.740				0.357
Pseudo R2		0.814	0.855	0.422	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0420	0.0720	0.0426	0.256**	0.103
	(0.0883)	(0.0574)	(0.0631)	(0.103)	(0.0886)
Observations	290	751	290	780	298
R2	0.754				0.340
Pseudo R2		0.841	0.878	0.424	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$