

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

December 8, 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 occasions 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 occasions 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.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Migration flows, OLS</i>									
Farmers in origin	0.058 (0.045)	0.058 (0.045)	0.148*** (0.019)	0.048 (0.046)	0.048 (0.046)	0.145*** (0.019)	0.053 (0.042)	0.053 (0.042)	0.161*** (0.019)
R <sup>2</sup>	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
<i>Migration flows, PPML</i>									
Farmers in origin	0.119*** (0.025)	0.104** (0.045)	0.536*** (0.092)	0.115*** (0.026)	0.097** (0.044)	0.561*** (0.095)	0.115*** (0.023)	0.116*** (0.039)	0.544*** (0.087)
R <sup>2</sup>	-	-	-	-	-	-	-	-	-
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 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 2: Regressions, 1970 is both lag and present

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Migration flows, OLS</i>									
Farmers in origin	0.057 (0.054)	0.057 (0.054)	0.093*** (0.027)	0.061 (0.050)	0.061 (0.050)	0.104*** (0.027)	0.077 (0.049)	0.077 (0.049)	0.114*** (0.030)
R <sup>2</sup>	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
<i>Migration flows, PPML</i>									
Farmers in origin	0.145*** (0.032)	0.210*** (0.073)	0.506*** (0.114)	0.133*** (0.031)	0.225*** (0.066)	0.505*** (0.114)	0.132*** (0.029)	0.218*** (0.063)	0.505*** (0.116)
R <sup>2</sup>	-	-	-	-	-	-	-	-	-
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	
MerHH 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 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.

Table 3: Regressions, 1980 is both lag and present

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Migration flows, OLS</i>									
Farmers in origin	0.120*** (0.038)	0.120*** (0.038)	0.118*** (0.020)	0.121*** (0.043)	0.121*** (0.043)	0.113*** (0.020)	0.122*** (0.042)	0.122*** (0.042)	0.141*** (0.020)
R <sup>2</sup>	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
<i>Migration flows, PPML</i>									
Farmers in origin	0.199*** (0.028)	0.178*** (0.035)	0.626*** (0.098)	0.191*** (0.029)	0.183*** (0.038)	0.633*** (0.098)	0.187*** (0.028)	0.207*** (0.035)	0.624*** (0.092)
R <sup>2</sup>	-	-	-	-	-	-	-	-	-
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	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 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 4: Regressions, 1970 + 1980

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Migration flows, OLS</i>									
Farmers in origin	0.092*** (0.032)	0.092*** (0.032)	0.104*** (0.017)	0.093*** (0.033)	0.093*** (0.033)	0.108*** (0.017)	0.101*** (0.032)	0.101*** (0.032)	0.127*** (0.018)
R <sup>2</sup>	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
<i>Migration flows, PPML</i>									
Farmers in origin	0.178*** (0.021)	0.187*** (0.033)	0.570*** (0.078)	0.169*** (0.021)	0.196*** (0.033)	0.572*** (0.079)	0.166*** (0.020)	0.210*** (0.031)	0.568*** (0.077)
R <sup>2</sup>	-	-	-	-	-	-	-	-	-
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 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

Still working on this, but I've been thinking expanding the datasets according to the person weights and then run probit regressions since we assume that the error terms in the transmission of knowledge regression distributes lognormal. Regarding the first point, that is, if a row has a person weight of 38, which means that that observation represents 38 people in ZA, then I repeat that row 38 times. This will yield millions of observations. Regarding the second point, this would be a probit regression where the dependent variable is the dummy variable  $L_{wijkl}$ , which equals 1 when the person  $w$  lives in region  $j$ , was born in region  $i$ , grows crop  $k$ , 0 otherwise. Given that in the equation  $\log(u^{migration})$  distributes normal, then the probability that  $L_{wijkl} = 1$  is  $\Phi(\iota_{jkt} + \iota_{ijt} + \kappa\beta\log L_{ijkl-1})$ .

### 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 the regressions of wages of migrant farmers on  $L_{ikt-1}$ , we calculate average earnings by origin-destination-crop. The origin variable is determined by which province the person was born. In Table 5 we construct both  $L_{ikt-1}$  and  $L_{ijkl}$  based on the 2007 census. In Table 6 we construct both  $L_{ikt-1}$  and  $w_{ijkl}$  from the 2007 census.

### 5 Brazil

These are regressions of migration flows and earnings on previous stock of workers for Brazil. Dependent variables are from census 2000 and 2010, and the main covariate is lagged by thirty years, so it is derived from censuses 1970 and 1980. For both migration flows  $L_{ijkl}$  and earnings  $w_{ijkl}$ , origin is the state where the person was born, and destination is the mesoregion the person resides at the time of the census. We consider 13 crops: banana, cassava, chicken, cocoa, coffee, cotton, corn, fish, livestock, rice, soy, sugarcane, and tobacco. In contrast to the original paper, we exclude fruits and horticulture since it is unclear how to code this from IPUMS's categorization. Table 7 present results for both OLS and PPML regressions. Results are very similar to those in the original paper, but higher values. This might be due to many reasons, including these regressions not including other covariates, the location variable not being as granular (in the original paper, origin and destination are both at the mesoregion), not including fruits and horticulture, our sample of migrants being comprised by 30-65 year old head of households, etc.

Table 5: Migration Flows Regression for ZA, 2007

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Migration Flows, OLS</i>									
Farmers in origin	0.076 (0.091)	0.076 (0.091)	0.194** (0.076)	0.116 (0.098)	0.116 (0.098)	0.181** (0.085)	0.042 (0.088)	0.042 (0.088)	0.102 (0.088)
R <sup>2</sup>	0.778	0.778	0.380	0.740	0.740	0.357	0.754	0.754	0.340
Obs	329	329	329	306	306	306	327	327	327
<i>Migration Flows, PPML</i>									
Farmers in origin	0.075 (0.057)	0.066 (0.055)	0.270*** (0.097)	0.102* (0.055)	0.068 (0.052)	0.246** (0.100)	0.071 (0.057)	0.042 (0.063)	0.255** (0.103)
R <sup>2</sup>	-	-	-	-	-	-	-	-	-
Obs	1208	329	1208	1208	306	1208	1208	327	1208
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 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 2007 census. We exclude non-migrants from the sample.

Table 6: Earnings Regression for ZA, 2007

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Earnings, OLS</i>						
Farmers in origin	-0.116 (0.144)	-0.080 (0.090)	-0.172 (0.149)	-0.096 (0.097)	-0.117 (0.135)	-0.019 (0.094)
R <sup>2</sup>	0.533	0.278	0.525	0.263	0.529	0.274
Obs	311	311	288	288	312	312
<i>Earnings, PPML</i>						
Farmers in origin	-0.067 (0.163)	-0.043 (0.104)	-0.302* (0.171)	-0.023 (0.111)	-0.235 (0.161)	-0.008 (0.109)
R <sup>2</sup>	-	-	-	-	-	-
Obs	312	312	288	288	312	312
Dest-Crop-Year FE	Y		Y		Y	
Orig-Dest-Year FE	Y	Y	Y	Y	Y	Y
Men HH heads			Y	Y		
Men HH heads, 20-65 y/o					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) and (2) are based on a sample of 30-65 years old migrants. In columns (3) and (4) the sample is comprised by men between 30-65 years old. In columns (5) and (6) the sample is comprised by men between 20-65 years old. The covariate is the log of agricultural workers in the same activity in the region of origin. The dependent variable is the log of average wages 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 2007 census. We exclude non-migrants from the sample.



Table 7: The Influence of the Region of Origin on Earnings and Employment of Agricultural Workers in their Destination Region

	OLS			PPML		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>A. Migration Flows</i>						
Farmers in origin	0.150*** (0.013)	0.150*** (0.013)	0.290*** (0.015)	0.166*** (0.012)	0.159*** (0.015)	0.742*** (0.030)
R <sup>2</sup>	0.886	0.886	0.264	-	-	-
Obs	5700	5700	5700	65212	5700	65212
<i>B. Earnings</i>						
Farmers in origin	0.039*** (0.012)	0.039*** (0.012)	0.055*** (0.007)	0.042*** (0.013)	0.042*** (0.013)	0.125*** (0.036)
R <sup>2</sup>	0.756	0.756	0.530	-	-	-
Obs	5468	5468	5468	5700	5700	5700
Dest-Crop-Year FE	Y	Y		Y	Y	
Orig-Dest-Year FE	Y	Y	Y	Y	Y	Y
Without zeros		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. For migration flows, origin is the state where the person was born, and destination is the mesoregion where the person currently lives. The sample is comprised by 30-65 year old migrants. Columns (1), (2), and (3) show results using OLS estimators; while columns (4), (5), and (6) use PPML estimators. The covariate is the log of of agricultural workers in the same activity in the region of origin lagged by thirty years. We include the census of 2000 and 2010 in our regressions. We exclude non-migrants from the sample.