Pellegrina + Sotelo: "Migration, Specialization, and Trade: Evidence from the Brazilian March to the West"

November 26, 2019

1 Regression (11) for 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. On each table, we report five types of estimators, one by column. The first one is OLS, the second is PPML, the third is PPML excluding zeros, the fourth one is PPML only including destination/crop/year fixed effects, and the fifth one is OLS with only destination/crop/year fixed effects. Each table reports four subtables: the first one is the case when the 1970 census is used to contruct L_{ikt-1}, and the 1980 census is used to construct L_{ijkt}; the second subtable is the case where both variables are constructed from the 1970 census; the third one is when both variables were constructed from the 1980 census; the fourth one is when both variables are constructed from each 1970 and 1980 census separately, then appended such that we introduce a clear notion of time.

Table 1 is the baseline scenario where we construct migration flows where we only keep the heads of the households that are aged between 30 and 65 years old. In Table 2 we show the same results but considering that the migrants are only heads of the households that are men aged between 30 and 65 years old. Finally, in Table 3 we show the same results but considering that the migrants are only heads of the households that are men aged between 20 and 65 years old.

Table 1: Baseline regressions

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0588	0.120***	0.105**	0.537***	0.149***
	(0.0459)	(0.0259)	(0.0452)	(0.0922)	(0.0197)
Observations	266	1295	266	7689	882
R2	0.834				0.250
Pseudo R2		0.893	0.894	0.384	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0574	0.146***	0.211***	0.506***	0.0931***
	(0.0543)	(0.0328)	(0.0738)	(0.114)	(0.0273)
Observations	222	1274	222	8081	797
R2	0.870				0.526
Pseudo R2		0.915	0.899	0.369	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.120***	0.199***	0.178***	0.627***	0.119***
	(0.0381)	(0.0284)	(0.0353)	(0.0982)	(0.0210)
Observations	375	1700	375	8691	971
R2	0.816				0.227
Pseudo R2		0.883	0.910	0.383	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0922***	0.179***	0.188***	0.570***	0.105***
	(0.0326)	(0.0212)	(0.0333)	(0.0785)	(0.0177)
Observations	597	2974	597	16772	1768
R2	0.855				0.462
Pseudo R2		0.897	0.907	0.376	

Standard errors in parentheses

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

Table 2: Head of the HH are men

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0480	0.116***	0.0970**	0.562***	0.145***
	(0.0468)	(0.0262)	(0.0448)	(0.0952)	(0.0191)
Observations	243	1178	243	7530	832
R2	0.836				0.258
Pseudo R2		0.890	0.895	0.383	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0616	0.134***	0.226***	0.506***	0.105***
	(0.0503)	(0.0312)	(0.0661)	(0.114)	(0.0277)
Observations	204	1202	204	7873	762
R2	0.863				0.527
Pseudo R2		0.912	0.872	0.362	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.122***	0.192***	0.183***	0.633***	0.114***
	(0.0432)	(0.0298)	(0.0386)	(0.0987)	(0.0206)
Observations	351	1570	351	8461	923
R2	0.817				0.238
Pseudo R2		0.879	0.910	0.378	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0938***	0.169***	0.196***	0.572***	0.109***
	(0.0335)	(0.0216)	(0.0337)	(0.0791)	(0.0178)
Observations	555	2772	555	16334	1685
R2	0.853				0.467
Pseudo R2		0.894	0.901	0.369	

Standard errors in parentheses

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

Table 3: Head of the HH are 20-65 men

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0538	0.115***	0.117***	0.271***	0.161***
	(0.0424)	(0.0236)	(0.0395)	(0.0417)	(0.0191)
Observations	294	1395	294	943	943
R2	0.832				0.251
Pseudo R2		0.897	0.897	0.370	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.0778	0.132***	0.219***	0.182***	0.114***
	(0.0492)	(0.0291)	(0.0637)	(0.0521)	(0.0307)
Observations	255	1369	255	855	855
R2	0.870				0.519
Pseudo R2		0.917	0.922	0.393	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.122***	0.188***	0.207***	0.259***	0.142***
	(0.0420)	(0.0285)	(0.0355)	(0.0434)	(0.0203)
Observations	414	1842	414	1046	1046
R2	0.830				0.230
Pseudo R2		0.884	0.917	0.352	

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)	(5)
L_iktlag_log	0.102***	0.167***	0.211***	0.221***	0.127***
	(0.0322)	(0.0208)	(0.0314)	(0.0357)	(0.0190)
Observations	669	3211	669	1901	1901
R2	0.861				0.461
Pseudo R2		0.899	0.920	0.373	

Standard errors in parentheses

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

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.

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

4 South Africa, regressions (10) and (11)

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 4: Baseline regressions

	(1)	(2)	(3)	(4)
L_iktlag_log	0.000209	0.0401**	0.000218	0.230***
	(0.000114)	(0.0130)	(0.000112)	(0.0291)
Observations	3499	4328	3499	10635
R2	0.987			
Pseudo R2		0.393	0.0227	0.394

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)	(2)	(3)	(4)
L_iktlag_log	-0.0533*	0.0191	-0.0171	0.227***
	(0.0232)	(0.0174)	(0.0109)	(0.0398)
Observations	3429	4305	3429	10993
R2	0.721			
Pseudo R2		0.620	0.482	0.448

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)	(2)	(3)	(4)
L_iktlag_log	-0.000112	0.0619***	-0.000108	0.277***
	(0.000210)	(0.0164)	(0.000200)	(0.0365)
Observations	3671	4799	3671	11729
R2	0.988			
Pseudo R2		0.408	0.0234	0.391

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	(1)	(2)	(3)	(4)
L_iktlag_log	-0.0190*	0.0461***	-0.00521	0.254***
	(0.00836)	(0.0122)	(0.00314)	(0.0275)
Observations	7100	9104	7100	22722
R2	0.737			
Pseudo R2		0.519	0.384	0.419

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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

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