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

November 18, 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 four types of estimators, one by column. The first one is OLS, the second is PPML, the third is PPML excluding zeros, and the fourth one is PPML only including destination/crop/year fixed effects. This last column we do it since it seems that the origin/destination/year fixed effects are soaking up too much variation, which reduces the sample size tremendously. 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 separatedly, 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)
L_iktlag_log	0.0588	0.120***	0.105*	0.537***
	(0.0459)	(0.0259)	(0.0452)	(0.0922)
Observations	266	1295	266	7689
R2	0.834			
Pseudo R2		0.893	0.894	0.384

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

Table 2: Head of the HH are men

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

	(1)	(2)	(3)	(4)
L_iktlag_log	0.0538	0.115***	0.117**	0.545***
	(0.0424)	(0.0236)	(0.0395)	(0.0879)
Observations	294	1395	294	7836
R2	0.832			
Pseudo R2		0.897	0.897	0.393

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)
L_iktlag_log	0.0778	0.132***	0.219***	0.505***
	(0.0492)	(0.0291)	(0.0637)	(0.116)
Observations	255	1369	255	7970
R2	0.870			
Pseudo R2		0.917	0.922	0.367

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)
L_iktlag_log	0.122**	0.188***	0.207***	0.625***
	(0.0420)	(0.0285)	(0.0355)	(0.0930)
Observations	414	1842	414	8917
R2	0.830			
Pseudo R2		0.884	0.917	0.389

Standard errors in parentheses

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

	(1)	(2)	(3)	(4)
L_iktlag_log	0.102**	0.167***	0.211***	0.569***
	(0.0322)	(0.0208)	(0.0314)	(0.0777)
Observations	669	3211	669	16887
R2	0.861			
Pseudo R2		0.899	0.920	0.378

Standard errors in parentheses

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

2 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 Regressions with migration flows at the person level

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

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