

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 three types of estimators, one by column. The first one is OLS, the second is PPML, and the third is PPML excluding zeros. Each table reports four subtables: the first one is the case when the 1970 census is used to construct 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)
L_iktlag_log	0.0588 (0.0459)	0.120*** (0.0259)	0.105* (0.0452)
Observations	266	1295	266
R2	0.834		
Pseudo R2		0.893	0.894

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

Table 2: Head of the HH are men

	(1)	(2)	(3)
L_iktag_log	0.0480 (0.0468)	0.116*** (0.0262)	0.0970* (0.0448)
Observations	243	1178	243
R2	0.836		
Pseudo R2		0.890	0.895

Standard errors in parentheses

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

	(1)	(2)	(3)
L_iktag_log	0.0616 (0.0503)	0.134*** (0.0312)	0.226*** (0.0661)
Observations	204	1202	204
R2	0.863		
Pseudo R2		0.912	0.872

Standard errors in parentheses

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

	(1)	(2)	(3)
L_iktag_log	0.122** (0.0432)	0.192*** (0.0298)	0.183*** (0.0386)
Observations	351	1570	351
R2	0.817		
Pseudo R2		0.879	0.910

Standard errors in parentheses

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

	(1)	(2)	(3)
L_iktag_log	0.0938** (0.0335)	0.169*** (0.0216)	0.196*** (0.0337)
Observations	555	2772	555
R2	0.853		
Pseudo R2		0.894	0.901

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)
L_iktlag_log	0.0538 (0.0424)	0.115*** (0.0236)	0.117** (0.0395)
Observations	294	1395	294
R2	0.832		
Pseudo R2		0.897	0.897

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

Standard errors in parentheses

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

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

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 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.