The Deeper Roots of Human Capital Formation and Economic Development in Southeast Asia, ca. 1900 – 2000

Alexandra M. de Pleijt[1] Ewout Frankema[2]

[1]Wageningen University

[2]Wageningen University and CEPR

Economic History Seminar, Yale University

Motivation: Why study human capital in SEA?

- Formal education is key for Smithian and Schumpeterian growth: educated workers sustain processes of commercialisation, specialisation and innovation (Barro 1991, Barro 2001, Romer 1990, Galor 2011)
- Historical literature on contribution of human capital to modern economic growth has focused mainly on Western Europe and the US
- How important was mass education really? (Mokyr 1990, Mitch 1999, Crafts 1996, Clark 2005, Reis 2004, Allen 2003)
- Highly skilled and educated 'knowledge elites' (artisans, mechanics, engineers) important for technical leaders (Squicciarini and Voigtländer 2015, Feldman and van der Beek 2016, de Pleijt et al 2020, Mokyr et al 2022, Kelly et al 2023, Curtis and de la Croix 2023)
- But larger role for mass education in countries catching up with the technological frontier (e.g. Prussia and France) (Becker et al. 2011, Diebolt et al 2021, Franck and Galor 2022) and in later stages of the Industrial Revolution (Goldin and Katz 1998, Diebolt et al. 2021)

Motivation: Why study human capital in SEA? (cont.)

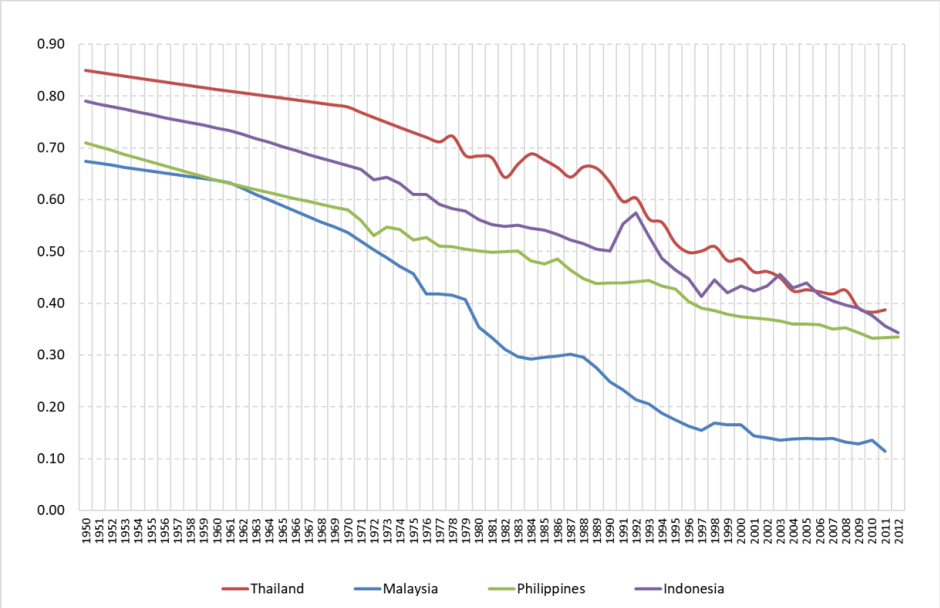


Figure: Share of the labour force in agriculture, 1950-2012

- Such questions are also relevant for the Global South!
- Late industrialising economies → technology adopters
- Development in SEA, 1950 - 2018:
  - 10-fold increase in GDP per capita
  - Poverty declined from 82% to 5%
  - Rapid structural change
- Research so far focused on national level, or very recent times (Crayen and Baten 2010, Barro and Lee 2015)

Deeper roots of modern economic growth?

1. Roots are not so deep:
   - Post-colonial/post-war economic policies (pro-rural, pro-poor, macro-economic stability) combined with agricultural technology shock (Green Revolution) (World Bank 1993, Bevan et al 1999, Berendsen et al 2013, Henley 2012, 2015)
2. Roots in colonial era:
   - (Extractive) agricultural commercialisation (Booth 1998, Dell and Olken 2020)
   - Rice and labour market integration (Huff and Caggiano 2007, Marks 2010)
   - Proto-industrial development (Huff and Angeles 2011, van Nederveen Meerkerk 2017)
   - General rise in real wages and living standards (Bassino and van der Eng 2022)

→ **We add a fifth dimension: human capital formation**

Our approach

- IPUMS microdata for Cambodia, Indonesia, Laos, Malaysia, Papua New Guinea, Philippines, Thailand and Vietnam:
  - 32 censuses; 111 million individuals (age 20 - 90)
  - 8 countries, 279 regions (province-level)
  - Birth-cohort approach: 1920s - 1960s
- Education attainment levels going back to birth-cohort of 1920s
- Most of these people are educated *before* the post-1970 take-off, therefore not influenced by the demand for schooling as a result of modern economic growth
- IPUMS information on the place of birth, allowing us to study impact of migration on the geographic concentration of human capital

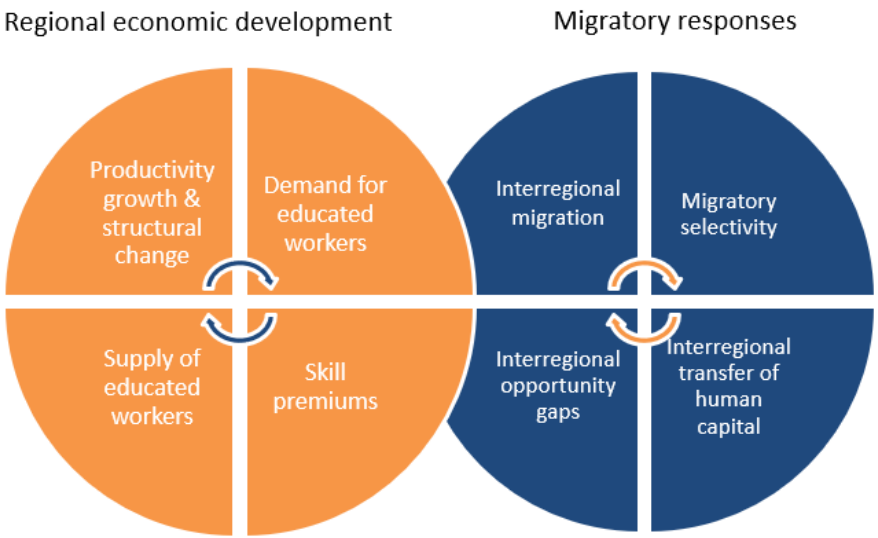
Our approach (cont.)

- Data is related to 21st century development levels:
  - Share of the working age population employed outside agriculture in 2000 (primary sector) (IPUMS)
  - Night light density in 2000 as proxy of regional GDP (Henderson et al 2018)

- We do not use an IV approach and do not claim causality, we think in terms of a path dependent development process with multiple feedback loops:
  - Historical roots of the schooling revolution are too diverse
  - Instead, we include an unusually broad set of historical and geographical control variables

Feedback loops in regional economic development and interregional migration

- The long run intertwining between education and development can best be seen as a path dependent process with multiple feedback loops:



1. Accumulation of human capital is both a response to, and a pre-condition for structural change
2. Skill-premiums create opportunity gaps that provoke migratory responses
3. Migratory selectivity into dynamic regions is a second driver of human capital accumulation

Key findings and contribution

1. Robust and significant relationship between 20th century schooling levels and economic development outcomes in 2000
    - Despite major economic and political disruptions and large-scale violence associated with decolonization in the 1940s - 1970s

2. No association between mass education and economic development, but a strong link with higher education (completed secondary and tertiary)

3. Migrants selected positively into the most dynamic regions and contributed ca. 2-3% of higher educational attainment over and above the levels of the non-migrant resident population

Main finding:

→ **Findings suggest that higher-education was crucial for adoption of complex modern/foreign technologies**

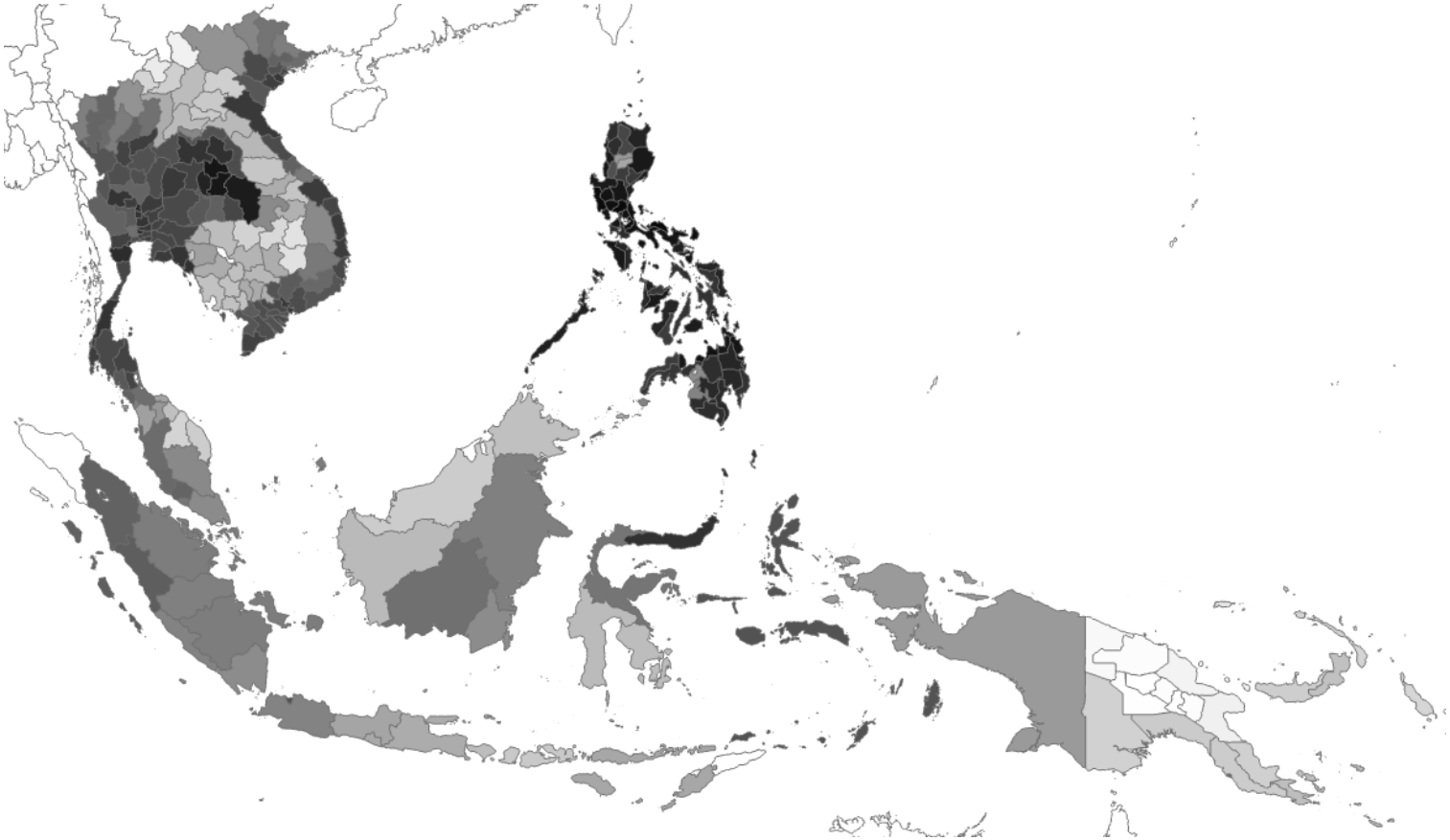
Overview of remainder of presentation

1. Data: Measuring human capital in the long run
2. The schooling revolution
3. Empirical analysis: Human capital and development
4. Exploring the contribution of migration
5. Conclusion

Data from IPUMS

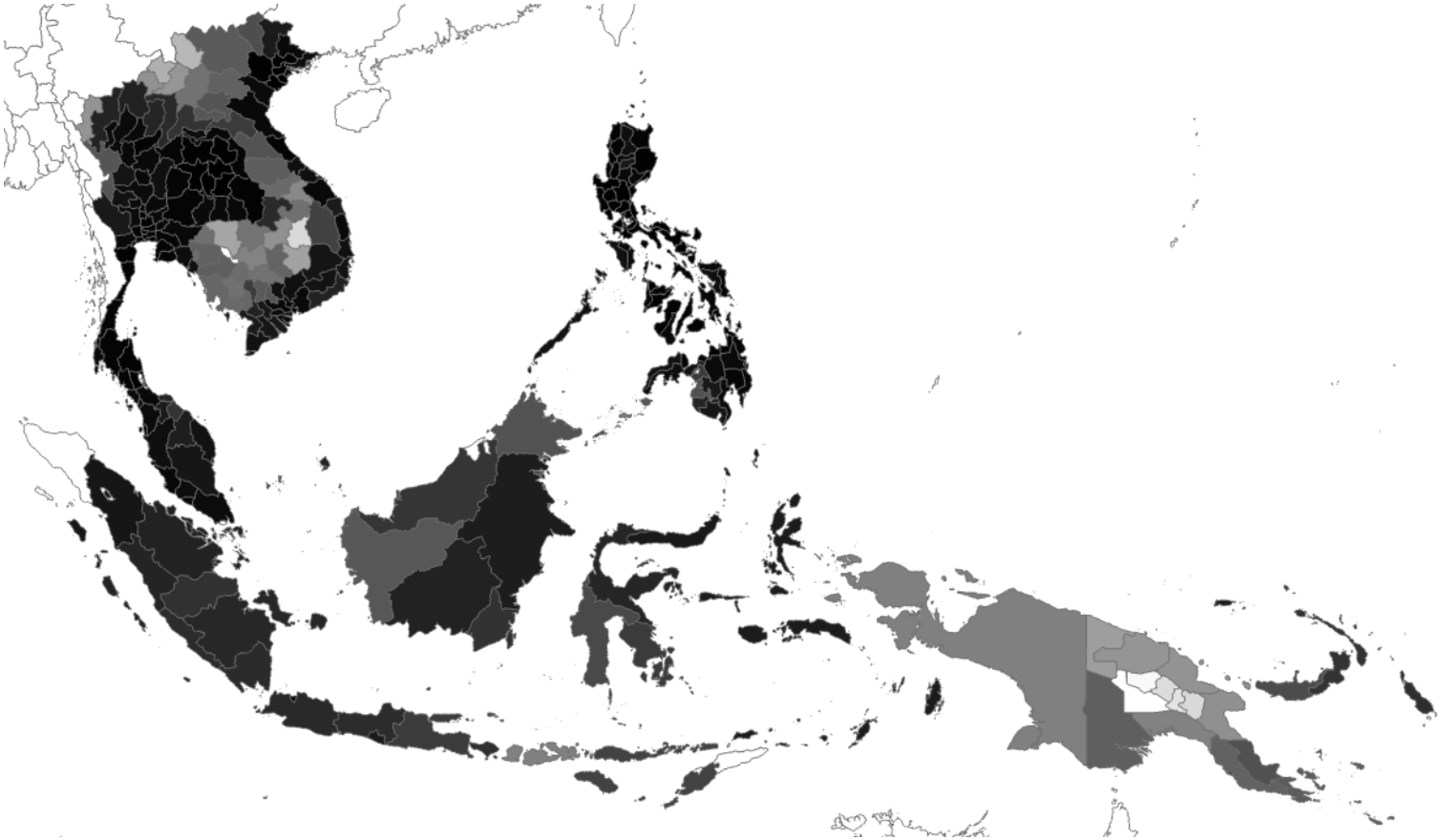
- Randomised samples from 32 population censuses and 8 countries:
  - Cambodia: 1998, 2004, 2008, 2013
  - Indonesia: 1971, 1976, 1980, 1985, 1990, 1995, 2000, 2005, 2010
  - Laos: 2005
  - Malaysia: 1970, 1980, 1991, 2000
  - Papua New Guinea: 1980, 1990, 2000
  - Philippines: 1990, 1995, 2000, 2010
  - Thailand: 1970, 1980, 1990, 2000
  - Vietnam: 1989, 1999, 2009
- Use evidence on: education levels, age, place of residence, province of birth
- For each birth-cohort at province level (279, *geolev1*):
  - % share with some formal education vs no education
  - % share of completed secondary education
  - % share of completed tertiary education
- Adjustments for migration (not reported in maps: similar results)
- Episodes of violence may bias our estimates ▸ Survivors' bias

The share of the 1925 birth cohort with some formal education



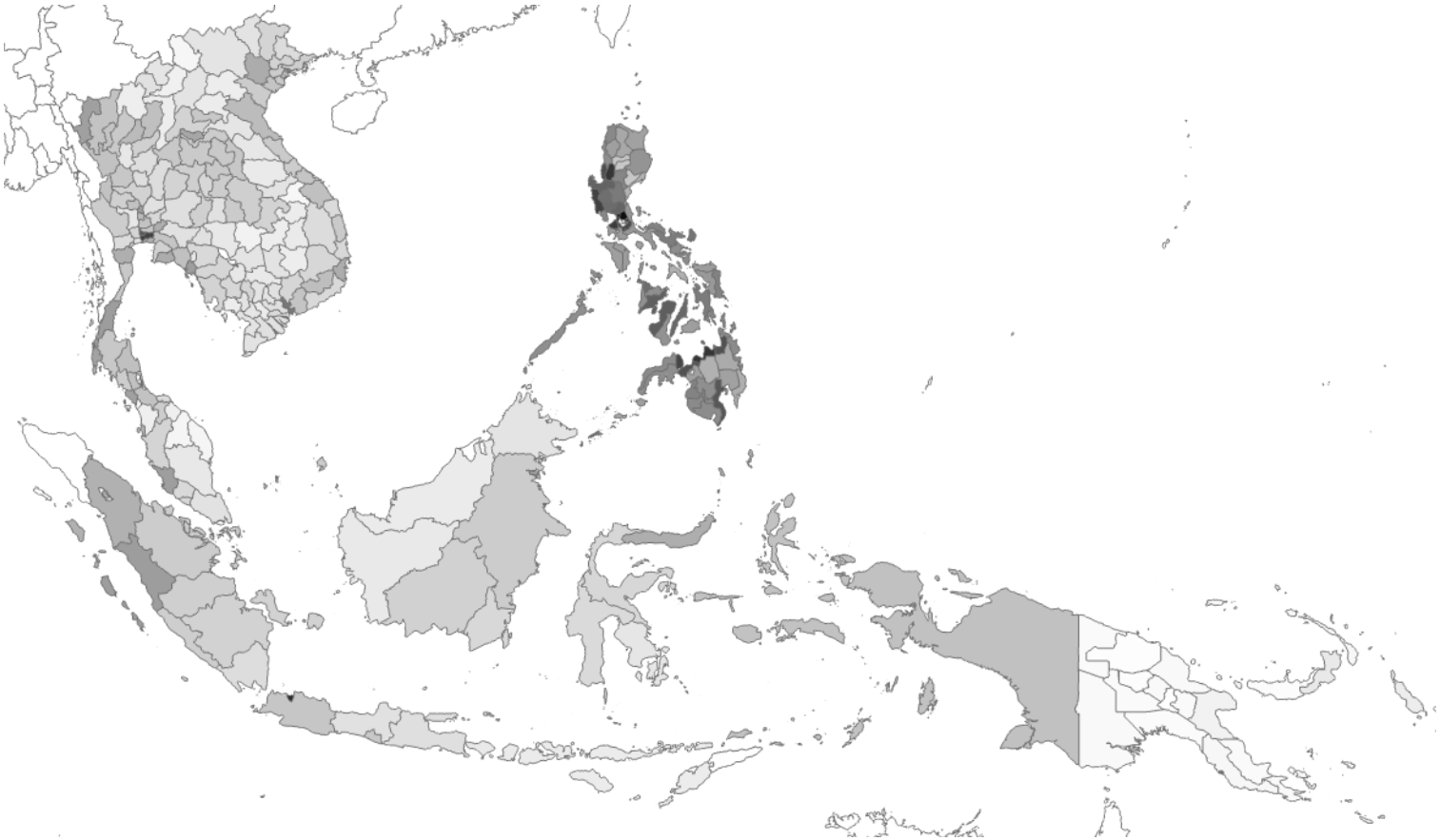
Notes: Darker is higher. The highest value is 0.97; the lowest value is 0.01; the mean is 0.58.

The share of the 1965 birth cohort with some formal education



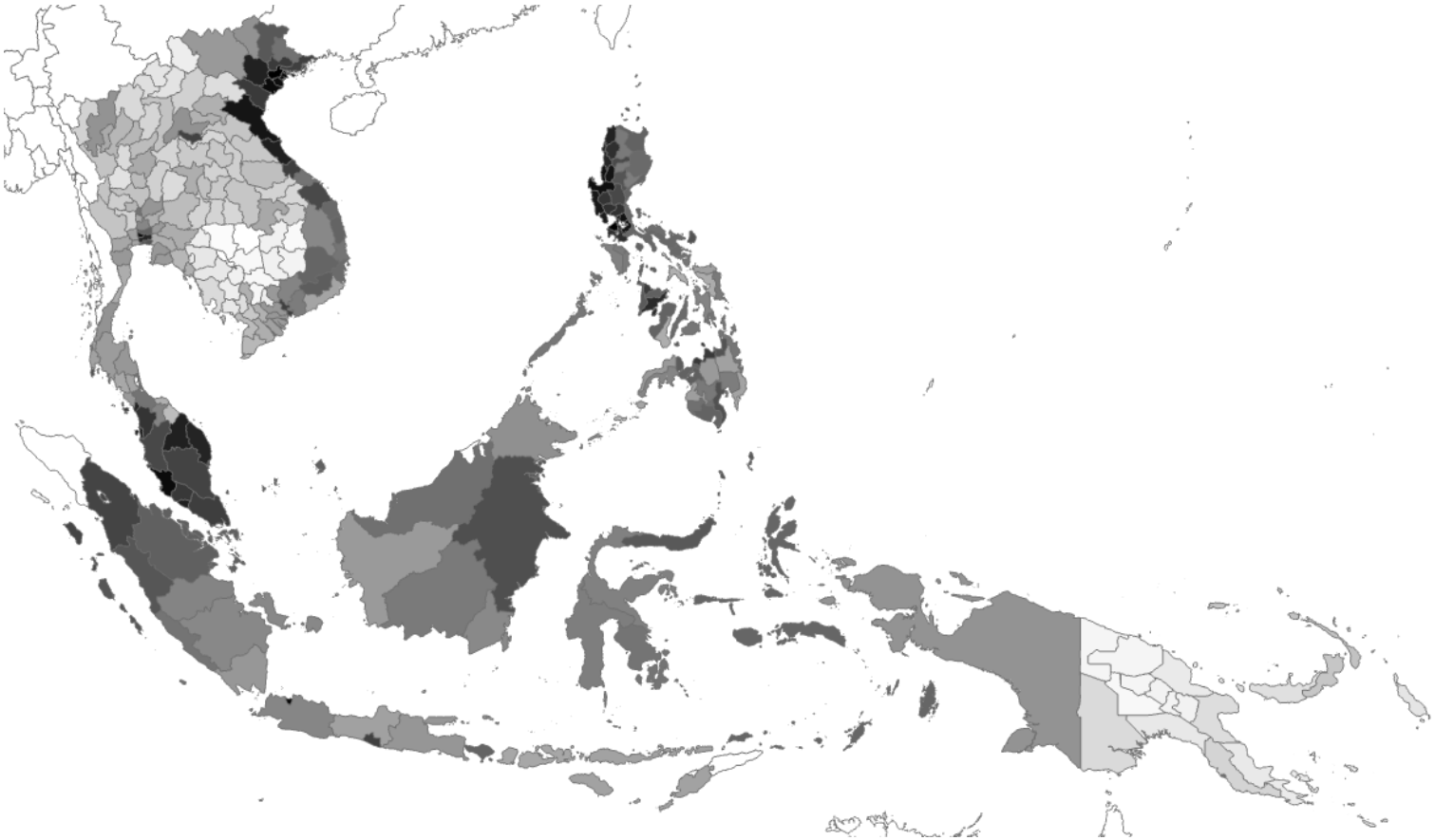
Notes: Darker is higher. The highest value is 0.99; the lowest value is 0.25; the mean is 0.88.

The share of the 1925 birth cohort with completed secondary education



Notes: Darker is higher. The highest value is 0.55; the lowest value is 0.00; the mean is 0.08.

The share of the 1965 birth cohort with completed secondary education



Notes: Darker is higher. The highest value is 0.79; the lowest value is 0.06; the mean is 0.35.

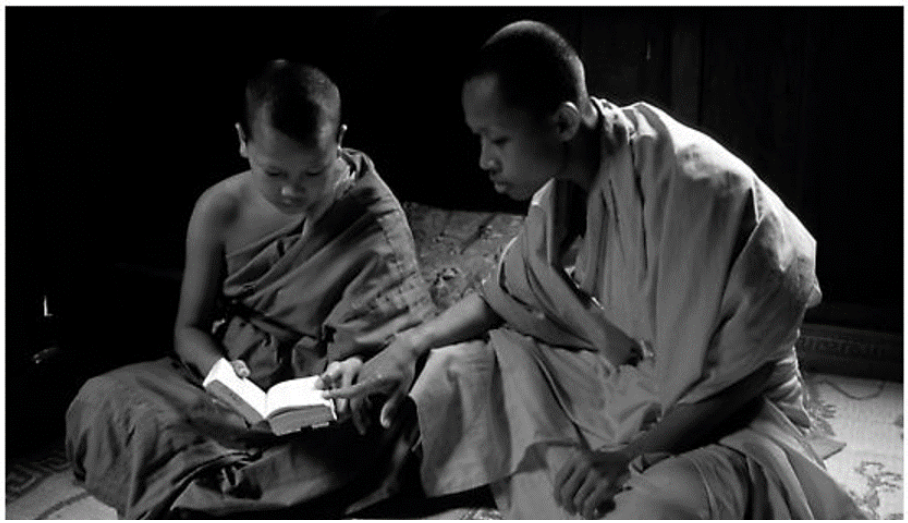
The roots of the schooling revolution

What were the roots of the schooling revolution in SEA?

Ho Chi Minh visiting a communist 're-education' center

- Colonial policies:
  - US saw mass-schooling as a pre-requisite for transfer of sovereignty; Dutch and French much more conservative
  - Training of a class of civil servants in capital cities
- Anti-colonial ideologies:
  - Ho Chi Minh re-educates the masses: education goes up from 20 to 90% between 1940s and 1950s
  - Thailand keen to preserve political autonomy

The roots of the schooling revolution (cont.)



Buddhist learning cultures

- Missionary influences and responses:
  - Tensions between 'local' and 'colonial' suppliers of education, and especially between Islamic and Christian schools
- (Local) religious roots:
  - Islam
  - Buddhism

Religion and human capital investments: Thailand

Table: Some formal education

	Education BP, 1925		Education BP, 1955		Education BP, 1965	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Priests per person, 1952	0.09*		0.02		0.01	
	(0.05)		(0.02)		(0.01)	
Monasteries per person, 1952		0.22***		0.08**		0.06*
		(0.08)		(0.04)		(0.03)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67	67	67	67	67	67
R-squared	0.74	0.75	0.55	0.60	0.52	0.56

Table: Higher education

	Higher education BP, 1925		Higher education BP, 1955		Higher education BP, 1965	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Priests per person, 1952	0.01		0.00		0.01	
	(0.01)		(0.01)		(0.02)	
Monasteries per person, 1952		0.03*		0.08**		0.13***
		(0.02)		(0.03)		(0.03)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67	67	67	67	67	67
R-squared	0.65	0.65	0.74	0.76	0.76	0.80

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

The roots of the schooling revolution (cont.)

- Economic geography:
  - Suitability for sedentary agriculture
  - Disease environments and climatic conditions
- Child labour:
  - Decreasing incentives to send children to school
  - E.g. tobacco and rubber in Indonesia

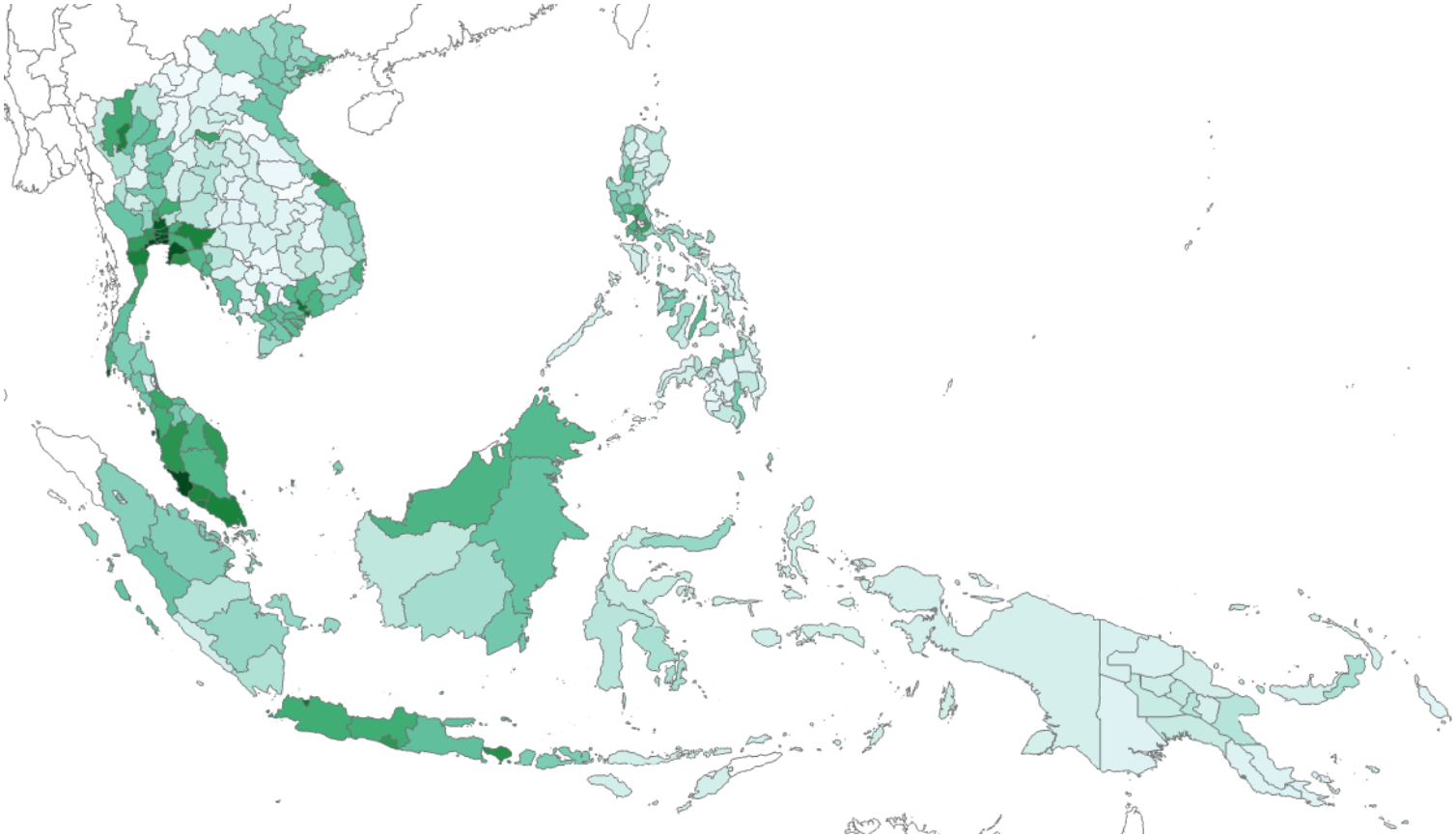
Empirical model

What was the association between human capital formation and economic development in SEA?

$$\text{Econ Dev}_r = \alpha + \beta_1 \text{HC}_{bc,r} + \text{X}'_r \mu + \Omega_c + \varepsilon_r, \tag{1}$$

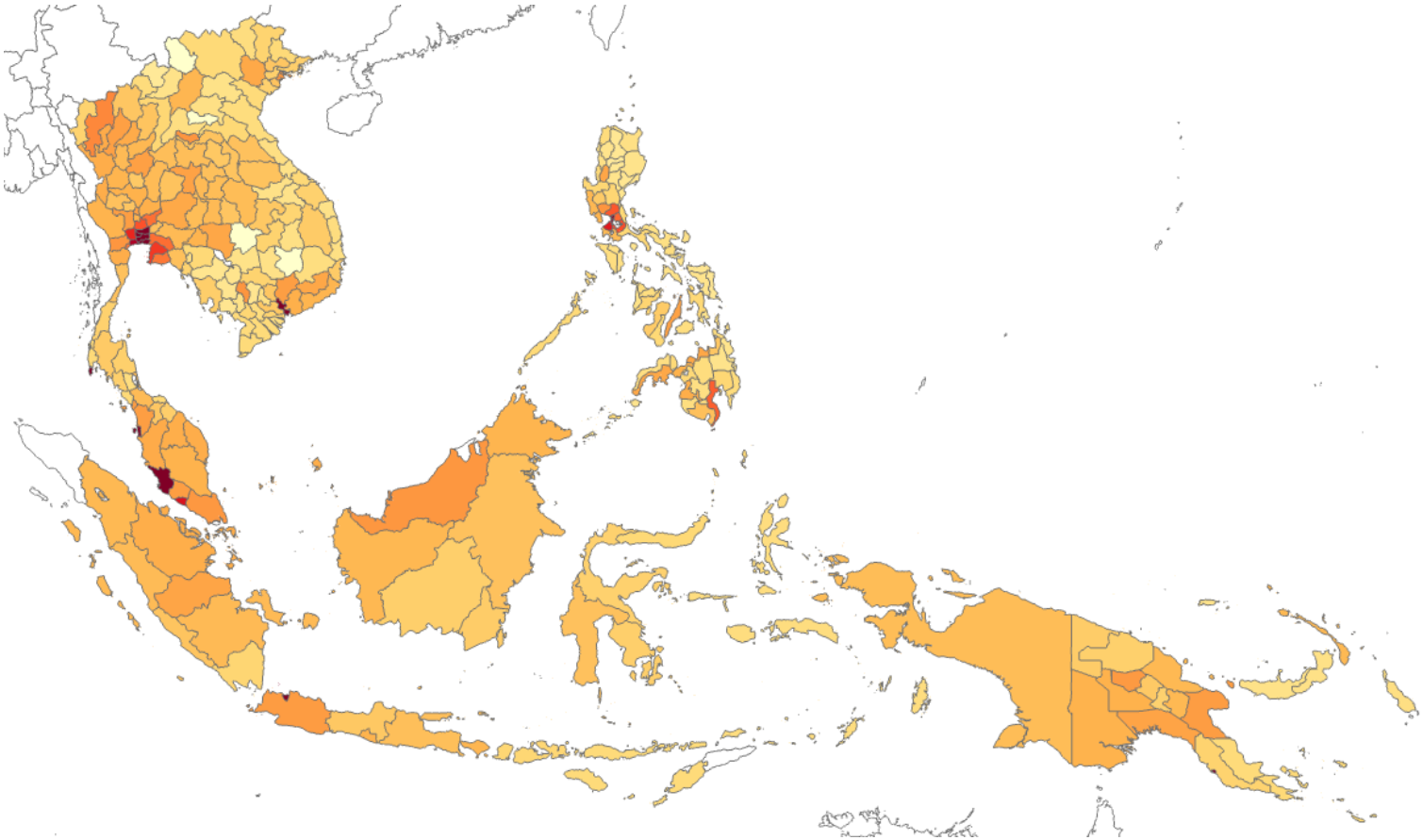
- Econ Dev_r is either the log of the share of the working age population employed outside of the primary sector in region r ca. 2000, or the log of night light intensity in region r in 2000
- $\text{HC}_{bc,r}$ is the log of share of birth-cohort bc residing in region r with a given education level
- Ω_c are country fixed effects
- ε_r is the error term
- X'_r includes a set of geographical controls as well as the initial levels of development

Measuring development: Share of the working age population employed outside of agriculture, ca. 2000



Notes: Darker is higher. The highest value is 0.67; the lowest value is 0.07; the mean is 0.30.

Measuring development: Lights at night, 2000



Notes: Darker is higher. The highest value is 63; the lowest value is 0; the mean is 9.41.

Control variables and robustness

- Geographical controls:
  - Latitude, longitude, temperature, rainfall, elevation
  - Quality of the land (rice suitability), suitability for cash-crops (tobacco, rubber, sugar, coffee), mining (coal, tin, oil)
  - Size of the region, access to sea, access to main rivers
  - Dummy indicating if the border is shared with a neighboring country
- Initial levels of development:
  - Log of the distance to the nearest city with more than 5,000 inhabitants in 1900
  - Alternatively: Log of population density in 1920 (HYDE-dataset)

- Robustness checks:
  - Spatial autocorrelation
  - Migration adjusted samples
  - Dropping countries

Human capital and development: Results 1925 birth-cohort

Table: Share of the 1925 birth-cohort with some education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Education, 1925	0.056	0.068	0.063	1.076***	1.020***	1.044***
	(0.050)	(0.043)	(0.046)	(0.355)	(0.327)	(0.344)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	279	279	279	279	279	279
R-squared	0.69			0.54		

Table: Share of the 1925 birth-cohort with higher education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Higher education, 1925	0.512***	0.428***	0.466***	6.259***	5.536***	6.564***
	(0.089)	(0.082)	(0.087)	(0.506)	(0.562)	(0.763)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	279	279	279	279	279	279
R-squared	0.72			0.65		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

▸ Results for Tertiary

Human capital and development: Results 1965 birth-cohort

Table: Share of the 1965 birth-cohort with some education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Education, 1965	0.085	0.108	0.154*	0.116	0.374	0.466
	(0.085)	(0.080)	(0.083)	(0.626)	(0.587)	(0.618)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	279	279	279	279	279	279
R-squared	0.69			0.52		

Table: Share of the 1965 birth-cohort with higher education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Higher education, 1965	0.446***	0.402***	0.454***	2.730***	2.386***	2.588***
	(0.061)	(0.059)	(0.059)	(0.406)	(0.370)	(0.388)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	279	279	279	279	279	279
R-squared	0.75			0.60		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

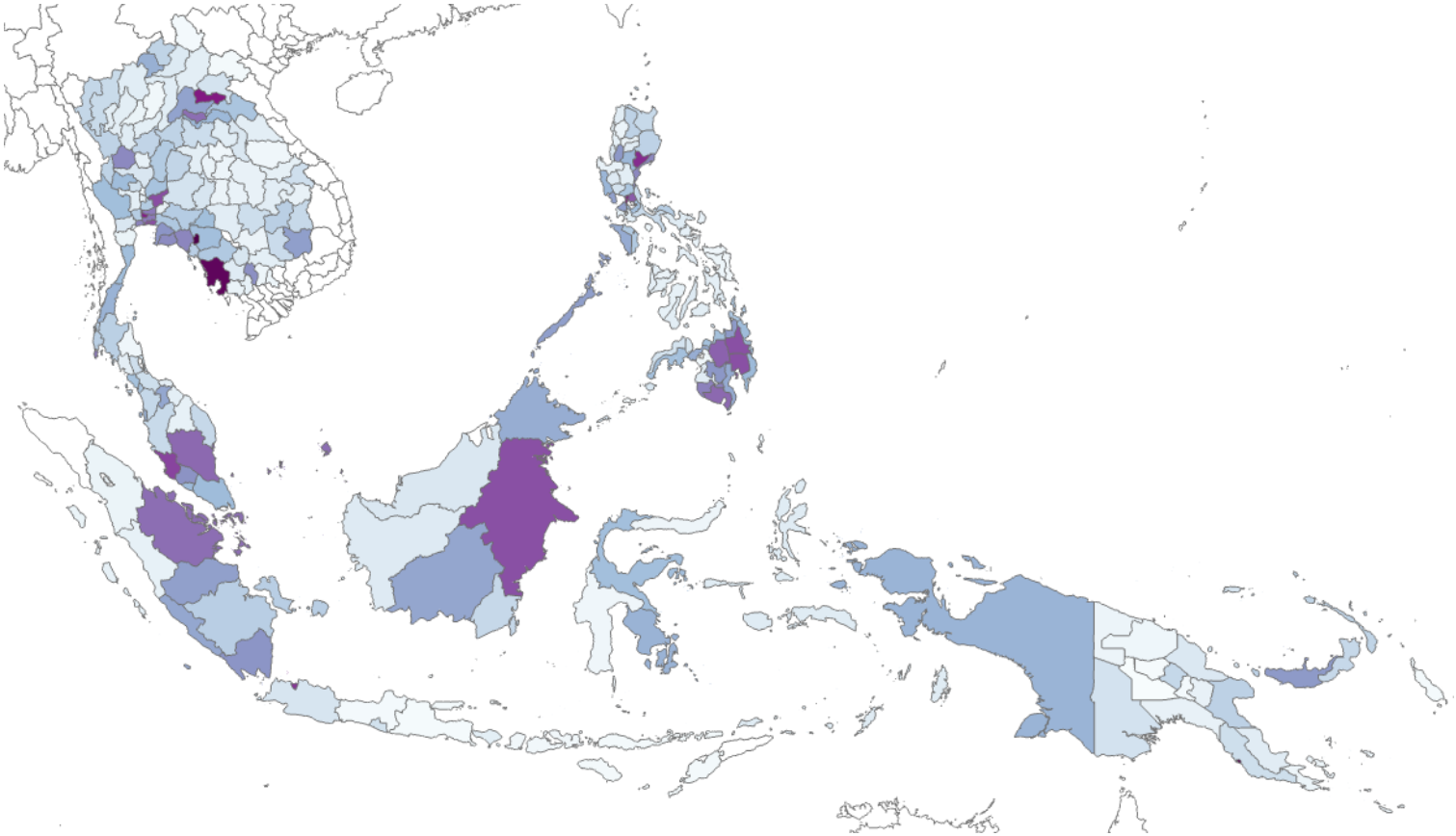
▸ Results for Tertiary

Interpretation of results so far

- No significant effect of mass education on structural change, only a strong association with higher education
- No claim for causality, but if we interpret the coefficients, one standard deviation increase in share of population with higher education in 1925 (7%) leads to:
  - 0.298 standard deviations increase in the share out of agriculture, i.e. a 3.1% increase (relative to a mean of 29.7%)
  - 0.667 standard deviations increase in lights at night, i.e. 45.7% increase (relative to a mean of 9.41 and 8.77 standard deviation)

- Exploring the contribution of immigration:
  1. Regressions robust to controlling for immigration
  2. Stylized fact of immigration at country level
  3. Association between immigration and development outcomes
  4. Link among educational differences and structural change

The share of immigrants, ca. 2000



Notes: Purple regions indicate areas with relatively high shares of immigrants in ca. 2000. The highest value is 0.98; the lowest value is 0.01; the mean is 0.20.

▸ Migrant origins

Human capital and development: Adjusting for immigration

Table: Share of the 1925 birth-cohort with higher education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Higher education BP, 1925	0.603***	0.441***	0.502***	5.685***	4.746***	6.373***
	(0.106)	(0.093)	(0.101)	(0.632)	(0.692)	(0.864)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241
R-squared	0.73			0.65		

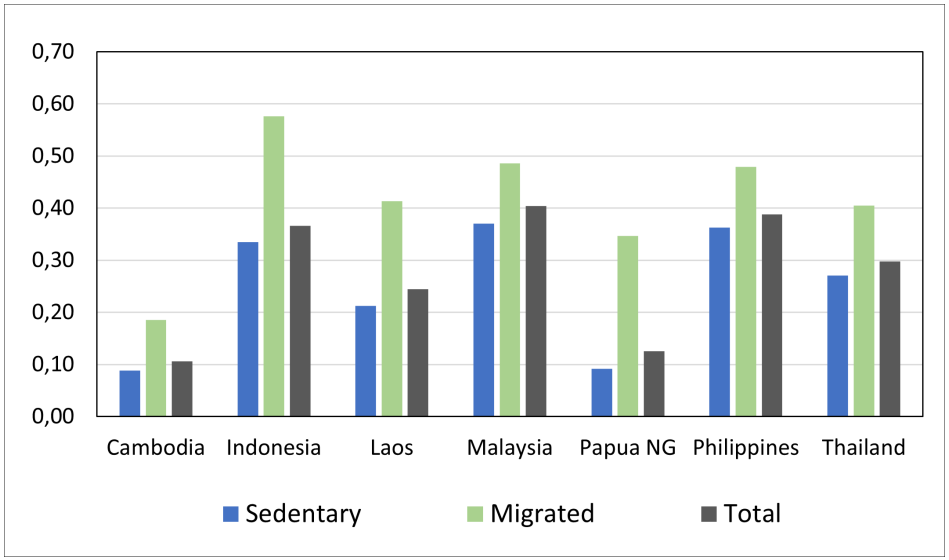
Table: Share of the 1965 birth-cohort with higher education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Higher education BP, 1965	0.506***	0.422***	0.493***	2.566***	2.198***	2.476***
	(0.073)	(0.070)	(0.068)	(0.558)	(0.521)	(0.535)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241
R-squared	0.76			0.61		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

▸ Results for Tertiary

The share of the sedentary and migratory working age population with completed secondary education in SEA, ca. 2000



- Migrant selectivity: interregional migrants more likely to be higher educated
- Education gap is larger in lesser developed economies (PNG, Laos, Cambodia) than in more advanced economies (e.g. Thailand and Malaysia)
- On average, immigration adds 2 - 3 percentage points to the stock of higher education

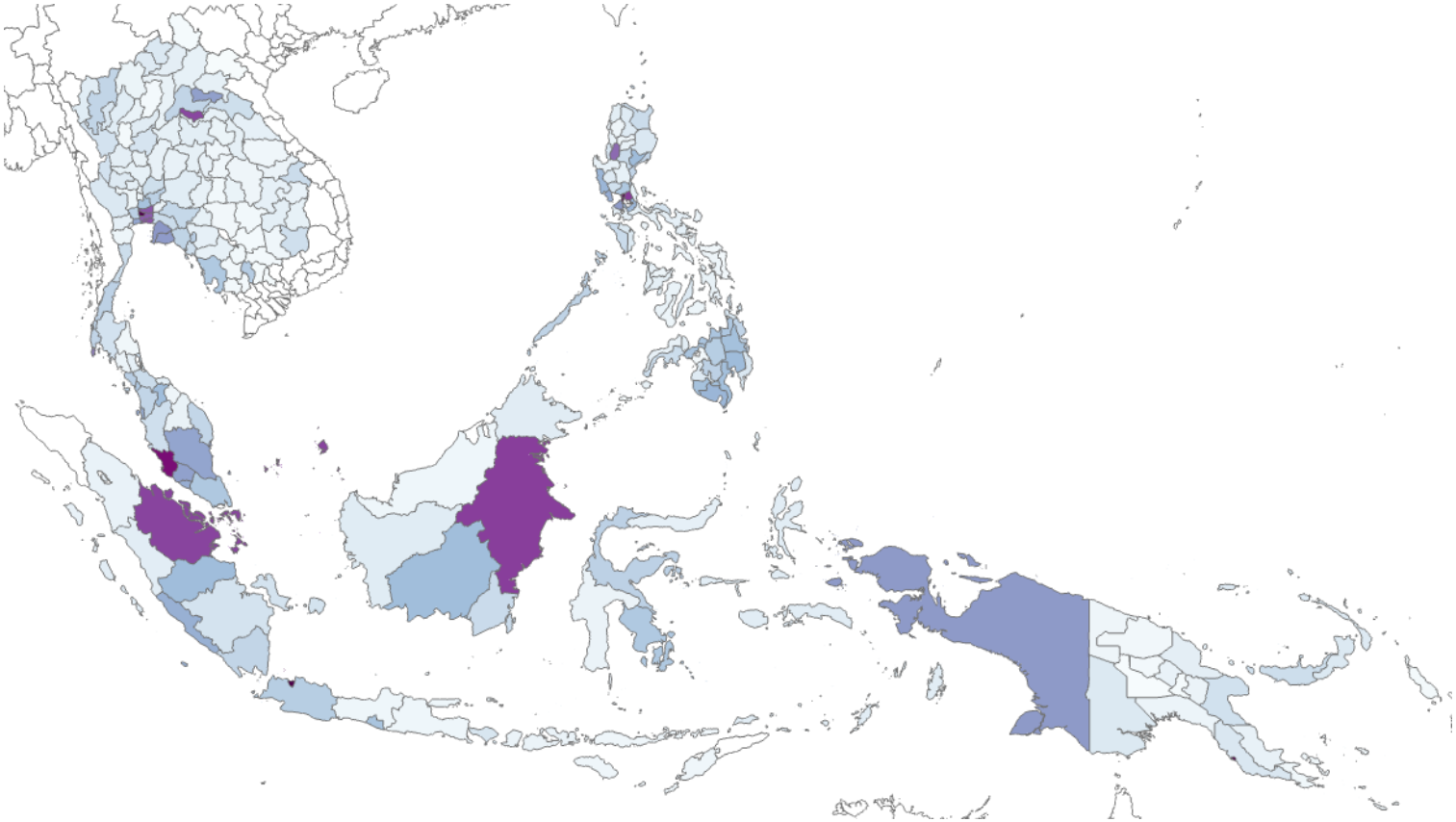
Relating the share of immigrants to economic development

Table: The share of migrants and economic development

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Share immigrants	0.183***	0.149***	0.170***	1.205***	0.943***	1.088***
	(0.033)	(0.030)	(0.034)	(0.280)	(0.254)	(0.309)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241
R-squared	0.73			0.61		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Differences in higher education, weighted for the share of immigrants



Notes: Purple regions indicate areas where immigrants have markedly higher shares of higher education than the sedentary population. The highest value is 0.43; the lowest value is 0.00; the mean is 0.07.

Differences in higher education and economic development

Table: Differences in higher education and economic development

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Lag	Error	OLS	Lag	Error
Differences higher education	0.525***	0.441***	0.511***	3.829***	3.363***	3.984***
	(0.069)	(0.068)	(0.071)	(0.530)	(0.492)	(0.527)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241
R-squared	0.76			0.67		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

▸ Results for Tertiary

Conclusion

1. Using micro-level data on 111 million individuals from 32 censuses, we establish a strong association between early educational attainment shares and sub-national economic development levels in ca. 2000
   - The roots can be traced back into the colonial era
   - Results support the argument that the causes of post-1970 take-off can be found further back in time

2. Especially strong effects of higher education
   - Explanation: 'late' industrializing countries have to catch up with more advanced technology frontier than 'early' industrializers such as Prussia

3. Contribution of interregional migration to (a) the human capital stock in most dynamic regions, and (b) economic development

Survivors' bias?

- What if educated persons live longer: are we systematically overestimating education levels?
- We exploit the fact that we can trace the same birth cohorts through consecutive decadal censuses

Country	Year of census	N	Share educated: 1925-cohort
Indonesia	1971	634,642	49.1%
	1976	281,170	
	1980	7,234,577	42.5%
	1985	605,858	
	1990	912,544	47.4%
	1995	718,837	
	2000	20,112,539	34.2%
	2005	1,090,892	
	2010	23,603,049	
Malaysia	1970	175,997	50.8%
	1980	182,601	48.1%
	1991	347,892	44.9%
	2000	435,300	41.4%
Thailand	1970	772,169	73.1%
	1980	388,141	70.9%
	1990	485,100	73.8%
	2000	604,519	69.8%

Back

Human capital and development: Results 1925 birth-cohort

Table: Share of the 1925 birth-cohort with tertiary education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Tertiary education, 1925	0.720***	0.570***	0.632***	7.953***	6.486***	7.859***
	(0.128)	(0.119)	(0.111)	(0.926)	(0.930)	(1.564)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	279	279	279	279	279	279
R-squared	0.71			0.61		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

▸ Back

Human capital and development: Results 1965 birth-cohort

Table: Share of the 1965 birth-cohort with tertiary education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Tertiary education, 1965	0.869***	0.785***	0.846***	4.941***	4.312***	4.569***
	(0.119)	(0.115)	(0.115)	(0.753)	(0.627)	(0.686)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	279	279	279	279	279	279
R-squared	0.76			0.60		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

▸ Back

Migrant Birth Regions

Country	Region	Birth Region	Share
Thailand	Phuket	Nakhon Si Thammarat	0,16
	Saraburi	Sa Kaeo, Prachin Buri	0,34
	Pathum Thani	Bangkok	0,17
	Nonthaburi	Bangkok	0,45
	Samut Prakan	Bangkok	0,21
Philippines	Manila Metro, 3rd District	Manila	0,15
	South Cotabato, Sarangani	Iloilo, Guimaras	0,22
	Quirino	Isabela	0,24
		Ifugao	0,17
	Davao (Davao del Norte)	Bohol	0,19
		Cebu	0,19
	Bukidnon	Cebu	0,24
		Bohol	0,15
Papua New Guinea	National Capital District	Central	0,18
Malaysia	Selangor, Kuala Lumpur Federal Territory	Perak	0,26
	Pahang	Kelantan	0,16
		Perak	0,16
Laos	Xaysomboun SR	Xiengkhuang	0,40
		Huaphanh	0,23
	Vientiane Capital	Vientiane	0,16
		Huaphanh	0,16
Indonesia	Kalimantan Timur	Jawa Timur	0,34
		Sulawesi Barat, Sulawesi Selatan	0,26
	DKI Jakarta	Jawa Tengah	0,36
		Banten, Jawa Barat	0,25
	Kepulauan Riau, Riau	Sumatera Utara	0,29
		Sumatera Barat	0,20
		Jawa Tengah	0,15
Cambodia	Pailin	Battambang	0,31
	Koh Kong, Preah Sihanouk	Kampot	0,28
		Takeo	0,16
		Phnom Penh, Kandal	0,15

Notes: Reporting regions with more than 40 percent migrants and focusing on birth regions with more than 15 percent.

Back

Human capital and development: Adjusting for migration

Table: Share of the 1925 birth-cohort with tertiary education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Tertiary education BP, 1925	0.841***	0.598***	0.685***	6.943***	5.132***	7.732***
	(0.155)	(0.136)	(0.139)	(1.191)	(1.201)	(2.122)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241
R-squared	0.72			0.62		

Table: Share of the 1965 birth-cohort with tertiary education

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1) OLS	(2) Lag	(3) Error	(4) OLS	(5) Lag	(6) Error
Tertiary education BP, 1965	0.808***	0.654***	0.771***	4.555***	3.720***	4.213***
	(0.113)	(0.108)	(0.110)	(0.830)	(0.771)	(0.848)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241
R-squared	0.76			0.62		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

▸ Back

Differences in tertiary education and economic development

Table: Differences in tertiary education and economic development

	Share non-agriculture, ca. 2000			Lights at night, 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Lag	Error	OLS	Lag	Error
Differences tertiary education	0.888***	0.713***	0.834***	7.537***	6.721***	8.089***
	(0.170)	(0.153)	(0.174)	(1.080)	(1.058)	(0.985)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241
R-squared	0.75			0.68		

Notes: Robust standard errors in parentheses. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Back