Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia

Samuel Bazzi
Boston University

Arya Gaduh
University of Arkansas

Alex Rothenberg
RAND Corporation

Maisy Wong
Wharton School

5 November 2015

Columbia University

Skill Transferability Is Important For Development

- Central role of **geographic mobility** in **development process**
- Labor **sorting** $\Longrightarrow$ productivity (Becker, 1962)
- We study **skill transferability** across locations within **agriculture**
- Natural policy experiment: large-scale **population resettlement**

Skill Transferability Is Important For Development

- Central role of **geographic mobility** in **development process**
- Labor **sorting** $\Longrightarrow$ productivity (Becker, 1962)
- We study **skill transferability** across locations within **agriculture**
- Natural policy experiment: large-scale **population resettlement**

we provide a causal argument that

location-specific human capital

$\Downarrow$

skill transferability $\Longrightarrow$ migration patterns

$\Downarrow$

spatial distribution of productivity

This Paper: How Transferable Are Skills Across Space?

- **Key Parameter**: elasticity of productivity w.r.t. skill transferability

Empirical Challenges

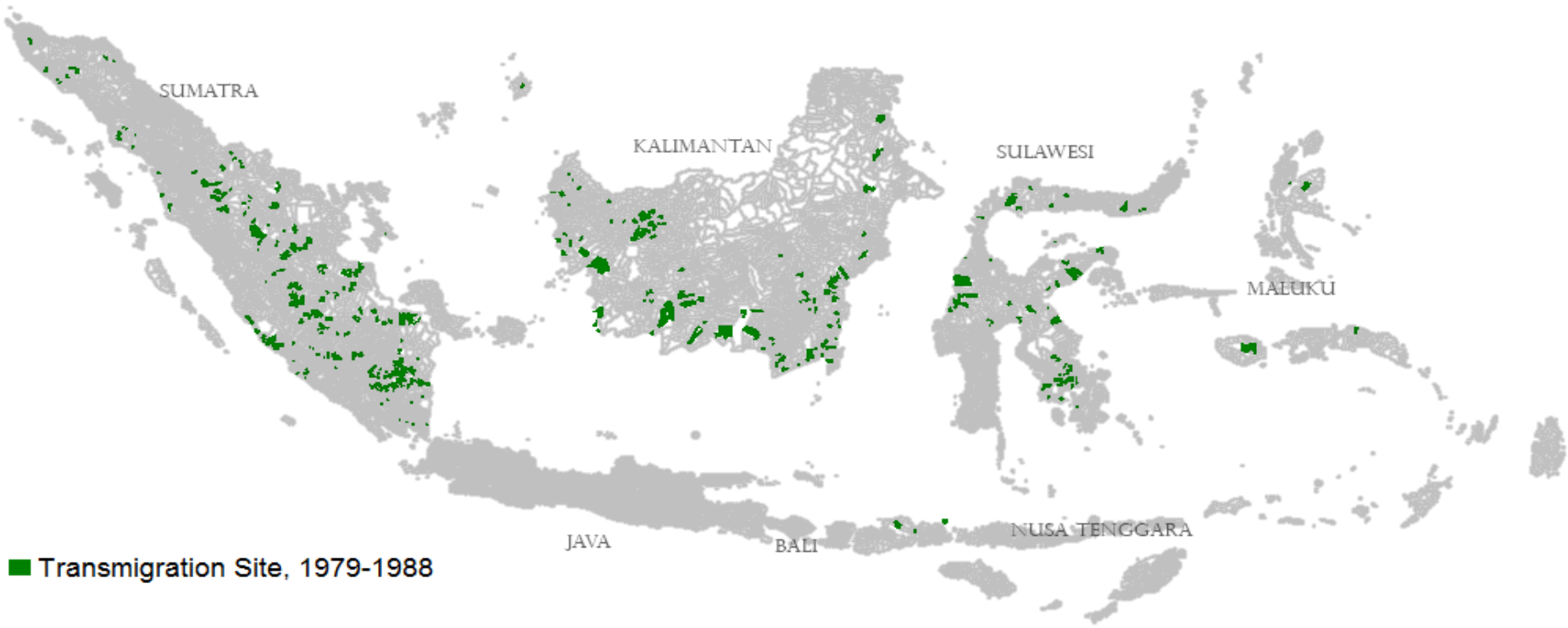
1. skill transferability **difficult to measure**
2. **endogenous sorting** on comparative advantage

Our Approach

1. We develop a **novel proxy** for skill transferability across locations
   - ▷ **agroclimatic similarity** ($\mathcal{A}$) between migrant origins and destinations
   - ▷ transferability $\Uparrow$ in similarity of endowments between two locations
     (akin to occupational similarity in labor, e.g., Gathmann & Schönberg, 2010)
   - ▷ precedent: "latitude-specific" farming skills, "east-west" mobility
     (Diamond, 1997; Steckel, 1983)
2. **Plausibly exogenous relocation of migrants** across rural Indonesia

A Natural Experiment in Spatial Labor Allocation

- **Transmigration**: rural-to-rural resettlement, 1979–1988
  - 2 million migrants from Java/Bali settled in **newly created villages**
  - goal: **population redistribution** with a focus on **rice production**

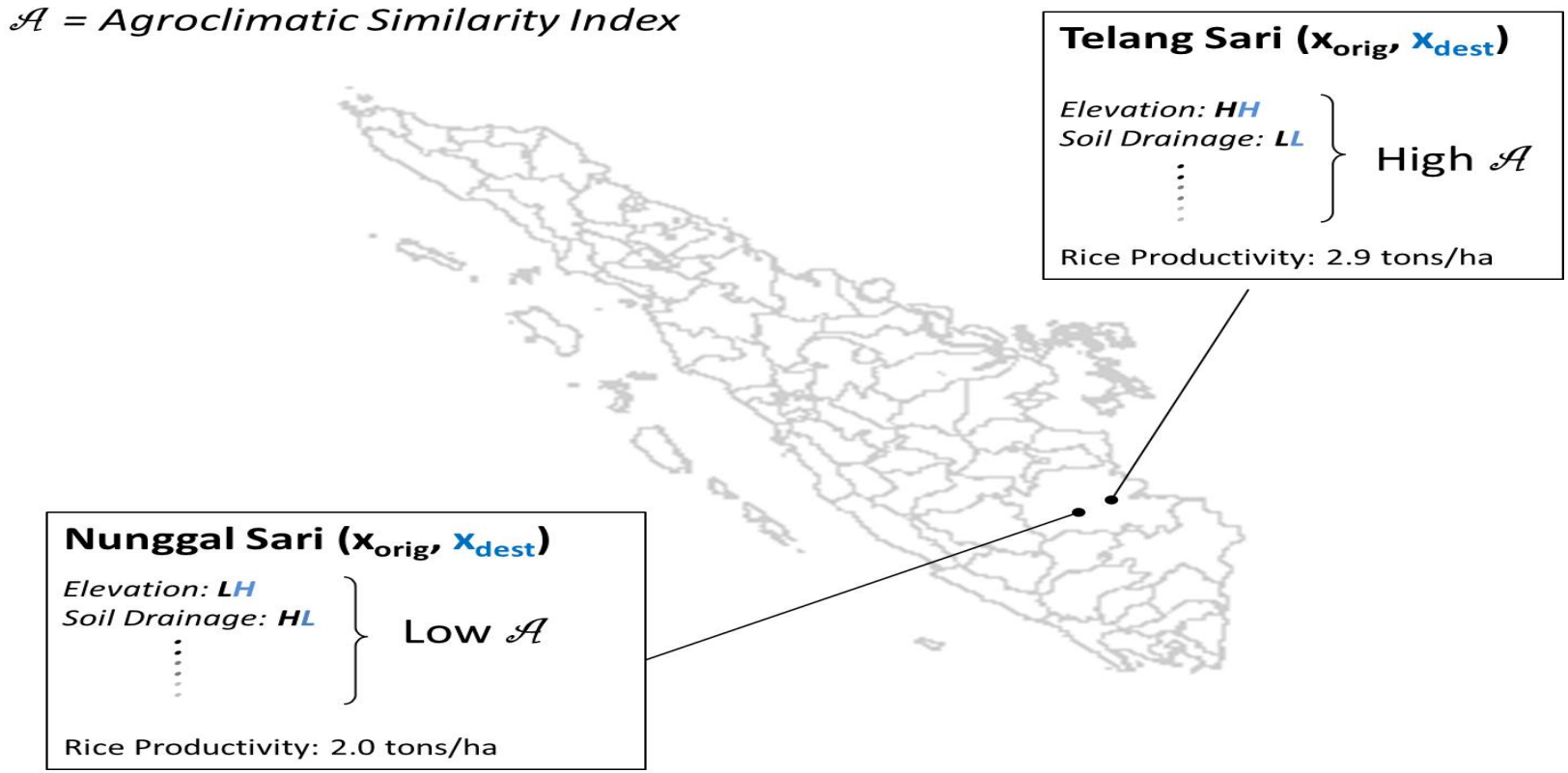


$\Longrightarrow$ rich **spatial variation** in agroclimatic conditions faced by migrants; **no systematic assignment** of agroclimatic origins to destinations

New Proxy for Skill Transferability

Identification: comparing rice productivity across observably identical villages with migrants from similar vs. dissimilar rice-growing origins

Stylized Case Study



Data: *many* agroclimatic *origins* and *destinations*, individual-level migration/demographics, village-level cross-section of productivity in 2001

Skill Transferability and Economic Development

Preview of Results

- **Large avg. elasticity**: 1 SD $\Uparrow$ similarity $\Longrightarrow$ 20%$\Uparrow$ **rice productivity**
  - similarly positive effect on other annual food crops
  - null effect on perennial cash crops (placebo)
- Several **adaptation mechanisms**
  - crop adjustments, occupational switching, interactions with natives
- **Costly, incomplete adjustment** over medium-run
  - large effects on **nighttime light intensity** in 2010
- **Policy** evaluation
  - simulated migrant reallocation $\Longrightarrow$ 27% $\Uparrow$ aggregate rice yield
  - average treatment effects: planned but unsettled villages as controls

Related Literature

1. **Barriers to mobility, spatial arbitrage, and labor (mis)allocation**

   (e.g., Bryan et al, 2014; Munshi & Rosenzweig, 2014; Young, 2013)

   here: skill specificity and barriers to transferability $\Longrightarrow$ gains from labor reallocation may be smaller than inferred from regional productivity differences

2. **Persistent consumption, occupation, and production choices**

   (e.g., Abramitzky et al, 2014; Atkin, 2013; Michalopolous, 2012)

   here: location-specific human capital has productivity implications

3. **Adaptation to (abrupt) climate change**

   (e.g., Costinot et al, 2014; Hornbeck, 2012; Olmstead & Rhode, 2011)

   here: skill specificity $\Longrightarrow$ added costs of climate change

4. **Human capital and long-run spatial diffusion of development**

   (e.g., Ashraf & Galor, 2013; Comin et al, 2012; Putterman & Weil, 2010)

   here: skill transferability $\Longrightarrow$ persistent effects on today's economic landscape

External Validity: Broader Relevance

1. **Resettlement increasingly recognized as crucial last resort policy**
   (de Sherbinin et al, 2011; IPCC, 2014)
   - ▷ growing displacement risk, e.g. 60 mn in S. Asia due to weather
   - ▷ skill mismatch: major challenge in relocation programs (World Bank OP)

2. **Annual food crops comprise 70% of global calories**
   - ▷ crops (esp. rice) expected to be most vulnerable to climate change
   - ▷ untilled, arable land being redistributed in Africa (World Bank, 2013)

3. **Rural mobility and agriculture**
   - ▷ rural-to-rural migration 1.5–2$\times$ rural-to-urban flows (Young, 2013)
   - ▷ agriculture employs 1.3 billion people
   - ▷ lack of convergence in agricultural productivity (Rodrik, 2013)
   - ▷ agricultural productivity gap $\Longrightarrow$ global inequality (Gollin et al, 2014)

Roadmap

Introduction

Conceptual Framework

Indonesia's Transmigration Program

Data

Empirical Strategy

Main Results

Conclusion

Future Work

Roadmap

Roy Model with Many Farms and Many Farmers

Farmer i Born in $b(i)$ Choosing Among J Destinations

- J potential outcomes, but only observe $j(i)^* = \arg\max_j V_{ij}$, where $V_{ij} = y_{ij} + \varepsilon_{ij}$ is the indirect utility of living in j.

- Determinants of **productivity** (abstracting from unobservables):

$$y_{ij} = \gamma \mathcal{A}_{ij} + \mathbf{x}_j' \boldsymbol{\beta}$$

  $\mathbf{x}_j$: natural advantages, local agroclimatic attributes

  $\mathcal{A}_{ij}$: **agroclimatic similarity** between i and j

- **location-specificity** in farming know-how (Griliches, 1957), especially salient in **diverse rice agriculture** (Munshi, 2004; Van Der Eng, 1994)

Identifying Skill Transferability Across Space

Concerns	**Our Natural Experiment**
endogenous location choice	plausibly exogenous relocation of migrants
endogenous occupational and crop choices	farming scheme with a focal crop
lack of variation in growing conditions	wide geographic scope of settlements

Identifying Skill Transferability Across Space

Concerns	Our Natural Experiment
endogenous location choice	plausibly exogenous relocation of migrants
endogenous occupational and crop choices	farming scheme with a focal crop
lack of variation in growing conditions	wide geographic scope of settlements

agroclimatic similarity: measurable, exogenous source of comparative advantage

- "no labor market data equivalent to agronomic data are available for estimating counterfactual task productivities..." (Autor, 2013)
- solves identification problems in multi-market Roy models (e.g., Bayer et al, 2011; Dahl, 2002; Heckman & Honore, 1990)

Roadmap

Big Resettlement Push from the Capital

Large scale resettlement proposed in late 1970s

- program began on small-scale in 1905 by Dutch colonial government
- target: 2.5 mn people in **1979–1983**, and 3.75 mn from **1984–1988**
- budget: $6.6 billion USD, funded by oil revenue windfall

Motivations for the program

1. **population redistribution**: Java/Bali 66% of pop., 7% of land
2. **food security**: increase national agricultural production (esp. **rice**)
3. **nation building**: integrating ethnic groups into "one nation"
   [other paper]

Program Details

Places (Ministry of Public Works)

- **new villages and farms** created on previously uncleared federal land

People (New Ministry of Transmigration)

- Voluntary participation: married, farmers, household head age 20-40
  - $>$95% of farmers in food crops (rice) in Java/Bali, late 1970s (Census)
- 1-2 hectare farm plots allocated by lottery, ownership after 5-10 years

  (also, free transport, new house, and initial provisions)
- Majority of participants: landless agricultural households
  - different from typical migrant; **similar to stayers in rural Java/Bali**
  - rural-to-urban migrants ($+3$ years of schooling) vs. transmigrants (-0.7 years)

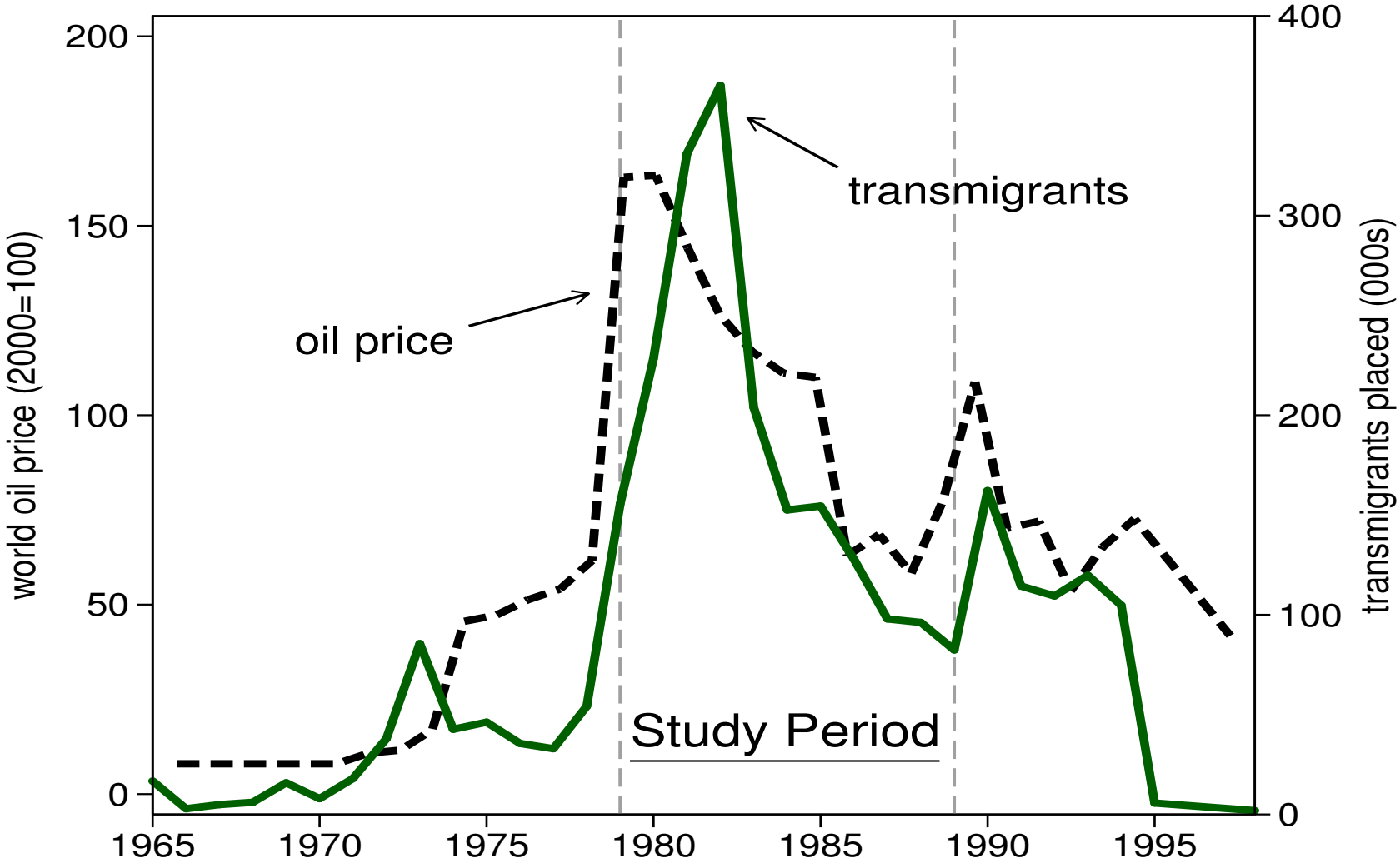
Advertising the Program

"A bright and vigorous future, together we move towards a joyous life"

Source: Donner (1987).

Rapid Scale Up and Sudden Contraction

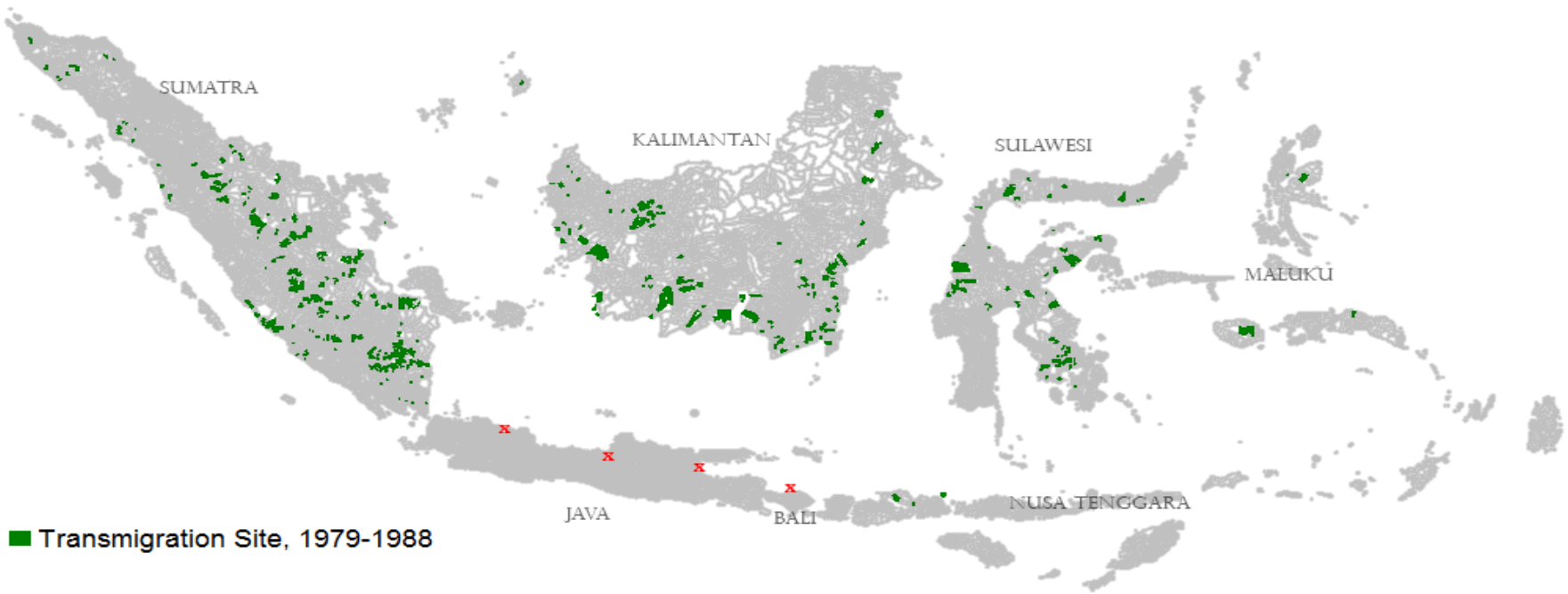
Driven by Oil Revenue



Notes: Totals calculated from the Transmigration Census of Villages prepared in 1999 by the Ministry of Transmigration. Oil price series from Bazzi and Blattman (forthcoming).

No Systematic Matching of People to Places

- Median Settlement (in 2000): 46 out of 119 origin districts, low Herfindahl=0.12
- Transmigrants sent from 4 transit camps (x) and could not choose destinations
  - knew very little pre-departure re destinations (Kebschull, 1986 camp survey)



- **plan-as-you-proceed**: "land use plans. . . abandoned"; "we would just ship out groups of transmigrants as they showed up in transit camps"
- Planners not concerned with matching on agroclimatic similarity
  - viewed Java/Bali rice farmers as superior to Outer Islanders
  - more concerned with mixing Java/Bali ethnic groups (for nation building)

Roadmap

Data Summary

Estimation Sample: 814 Transmigration villages, settled 1979-88

- ▷ coverage constraints limit our use of household surveys (e.g., IFLS)

$\Longrightarrow$ individual-level Census data on migration/demographics and ...

Village-Level Productivity and Development

1. **Rice productivity** in 2001/2: output per hectare
   - **focal crop**: primary staple, policy goal, *and* key crop in Java/Bali
2. Other agricultural productivity in 2001/2
3. **Nighttime light intensity** in 2010 (Henderson et al, 2012) ▸ maps

Key Regressor: village-level agroclimatic similarity, $\mathcal{A}_j \in [0, 1]$

Best Available Data

Many Administrative, Census, and GIS Sources

Source	Unit	Key variables
1998 Transmigration Census (newly digitized Ministry records)	Villages	year settled, # individuals settled
2000 Population Census	Individual	(birth) location, age, schooling, ethnicity, occupation
2003 Agricultural Census (*Podes*)	Villages	agricultural output, area planted
GIS/Maps	Various	light intensity, land attributes, rainfall, temperature
2004 Household Survey (*Susenas*)†	Individual	village, farm productivity, ethnicity, no origin data

† Only covers small random sample of 74 Transmigration villages.

Constructing an Agroclimatic Similarity Index ($\mathcal{A}_j$)

1. Distance metric: $d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{g=1}^{G} |x_{ig} - x_{jg}|$, origin i & destination j
   - ▷ **topography**: elevation, slope, ruggedness
   - ▷ **soil**: texture, sodicity, acidity, carbon content (1970s)
   - ▷ **water**: rainfall, distance to river, drainage, temperature

2. Individual agroclimatic similarity: $\mathcal{A}_{ij} = (-1) \times d(\mathbf{x}_i, \mathbf{x}_j)$

3. **Village-level (average) agroclimatic similarity**: $\boxed{\mathcal{A}_j = \pi_{ij}\mathcal{A}_{ij}}$

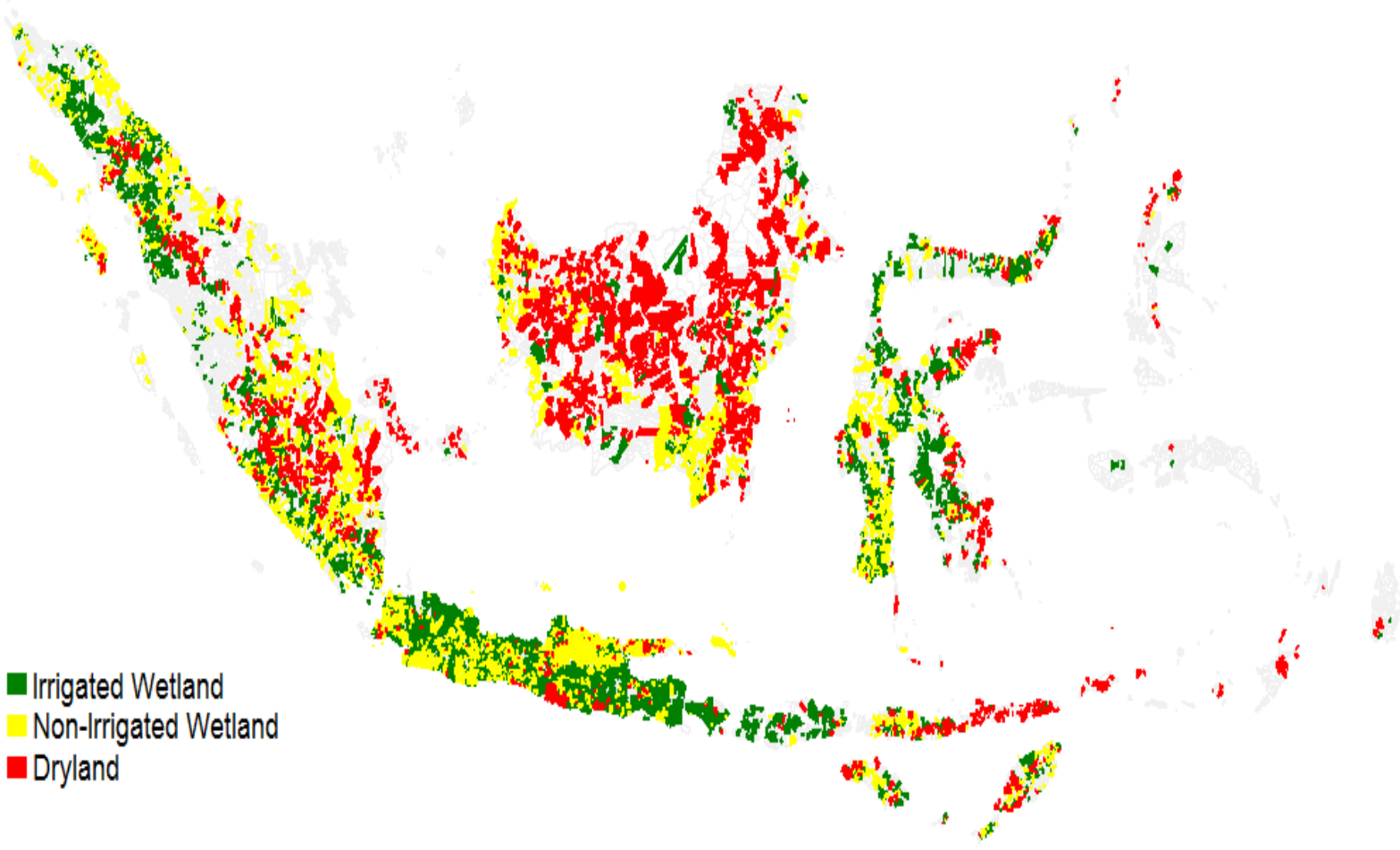
   π_{ij}: share of Java/Bali-born migrants in village j from district i

* mean, std. dev. of $\mathcal{A}_j$ indistinguishable from index based on random matches

* $\mathcal{A}_j$ uncorrelated with ethnic diversity (ELF) within Java/Bali-born population

Agroclimatic Diversity Within and Between Islands

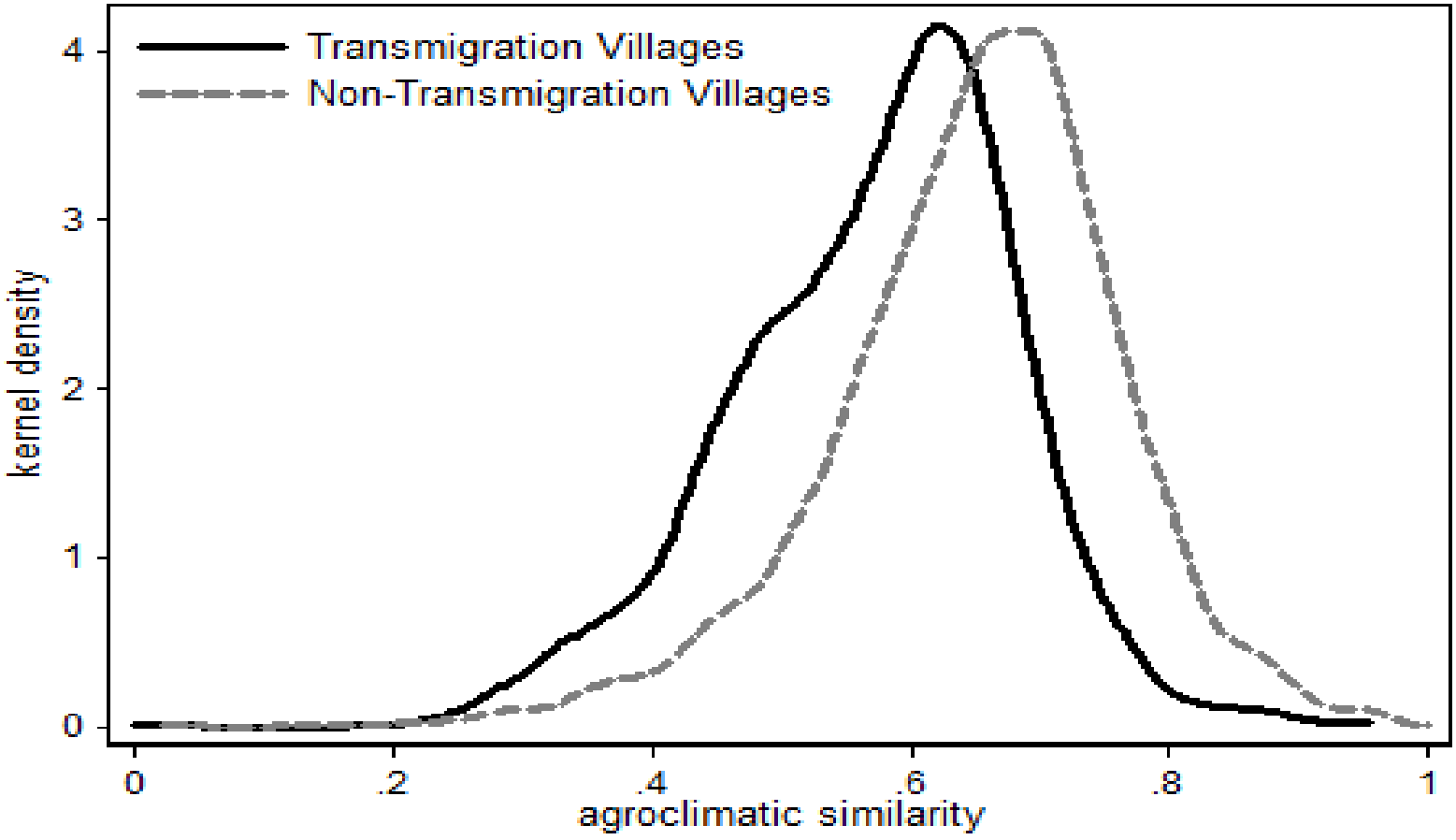
Three Rice Growing Systems



Notes: Data from *Podes* indicating the primary type of land on which rice is grown in the village in the 2001 growing season.

▸ agroclimatic diversity stats

The Natural Experiment Buys Us Relatively More Migrants from Dissimilar Origins



Agroclimatic Similarity (Immigrants) Across Villages

$\mathcal{A}_j = (-1) \times \sum_{i=1}^{I} \pi_{ij} d(\mathbf{x}_i, \mathbf{x}_j)$ where π: defined for **all immigrants** in j

Roadmap

Empirical Framework

Individual-level productivity: $y_{ij} = \gamma \mathcal{A}_{ij} + \mathbf{x}'_j \beta + \underbrace{\eta_i^u + \mu_j^u + \omega_{ij}}_{\textbf{unobservable}}$

Estimating equation at village-level:

$$\boxed{y_j = \gamma \mathcal{A}_j + \mathbf{x}'_j \beta + \sum_{i \in I_j} \eta_i^u + \mu_j^u + \omega_j}$$

defined over individuals I_j for whom j is optimal location $j(i)^* = j$

Identification of γ: high vs. low π share of migrants from similar origins in observably identical destinations

Identification Checks

Key Assumption: $\mathcal{A}_j \perp\!\!\!\perp \sum_{i \in I_j} \eta_i^u,\ \mu_j^u,\ \omega_j \mid \mathbf{x}_j$

Threat	Main Test
1. **Unobservable Natural Advantages**	balance on pre-1979 outcomes & potential yield
2. **Unobservable Demographics**	balance across schooling levels; bounding the selection out of rice farming
3. **Sorting**	gravity tests $\Longrightarrow$ no sorting on $\mathcal{A}_{ij}$; limited return/ex post migration
4. **Aggregation Bias**	robust to $\frac{native}{pop}$; individual-level regressions

1. (Lack of) Correlation of $\mathcal{A}_j$ with Potential Crop Yields

based on FAO-GAEZ

wetland rice potential yield (ton/Ha)	0.030
	(0.030)
dryland rice potential yield (ton/Ha)	0.046
	(0.049)
cocoa potential yield (ton/Ha)	-0.063
	(0.079)
coffee potential yield (ton/Ha)	-0.105
	(0.102)
palmoil potential yield (ton/Ha)	0.008
	(0.022)
cassava potential yield (ton/Ha)	-0.005
	(0.030)
maize potential yield (ton/Ha)	-0.070
	(0.051)

Notes: */**/*** significance at the 10/5/1 percent level. Correlations are conditional on island fixed effects and the predetermined village-level control variables $\mathbf{x}_j$. Conley (1999) standard errors with bandwidth of 150km.

further test ruling out unobservable rice-specific natural advantage confound ▸ test

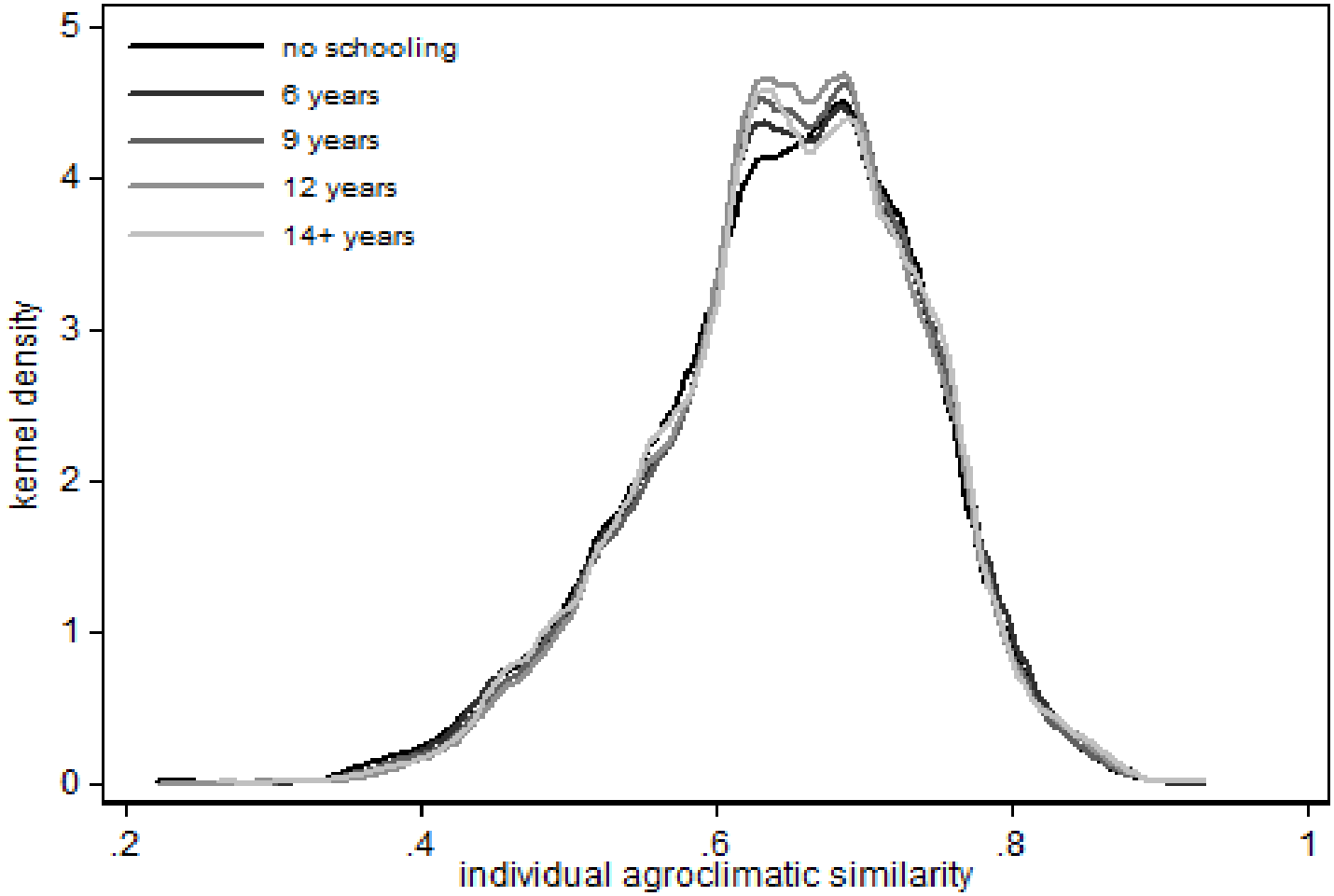
1. (Lack of) Correlation of $\mathcal{A}_j$ with Predetermined District-Level Outcomes based on 1978 Population Characteristics

log district population, 1978	-0.028
	(0.017)
own electricity (% district pop.)	-0.170
	(0.091)*
own piped water (% district pop.)	0.001
	(0.124)
own sewer (% district pop.)	-0.187
	(0.187)
use modern fuel source (% district pop.)	-1.366
	(1.419)
own modern roofing (% district pop.)	0.060
	(0.061)
own radio (% district pop.)	-0.027
	(0.196)
own TV (% district pop.)	-0.257
	(0.142)*
speak Indonesian at home (% district pop.)	-0.153
	(0.118)
literate (% district pop.)	-0.078
	(0.167)
average years of schooling in district	0.011
	(0.019)
agricultural sector (% district pop.)	0.125
	(0.079)
mining sector (% district pop.)	-0.202
	(0.505)
manufacturing sector (% district pop.)	-0.986
	(0.414)**
trading sector (% district pop.)	-0.393
	(0.265)
services sector (% district pop.)	-0.055
	(0.134)
wage worker (% district pop.)	-0.192
	(0.150)

Notes: */**/*** significance at the 10/5/1 percent level. Each variable in the row is based on data from the 1980 Population Census and restricted to the population in each district that did not arrive as immigrants in 1979 or earlier in 1980 (i.e., the still living population residing in the district in 1978). Correlations are conditional on island fixed effects and the predetermined village-level control variables $\mathbf{x}_j$. Standard errors clustered at the (1980) district level.

2. Agroclimatic Similarity: Comparable by Education

Predetermined schooling is uncorrelated with $\mathcal{A}_{ij}$



Notes: Agroclimatic similarity at the individual level for all Java/Bali-born migrants in Transmigration sites whose schooling was completed prior to the initial year of settlement. **Lack of correlation is robust to inclusion of individual-level Mincerian controls and also to scaling up to the village-level $\mathcal{A}_j$.**

3. No Sorting on $\mathcal{A}_{ij}$ into Similar Sites

Gravity Regression: All Possible Origin×Destination Pairs (2000 Census)

$$f(migrants_{ij}) = \alpha + \lambda_a \mathcal{A}_{ij} - \lambda_d \ln distance_{ij} + \mathbf{z}_j'\boldsymbol{\zeta} + \tau_i + \upsilon_{ij}$$

If endogenous sorting, then $\lambda_a,\ \lambda_d > 0$.

Dependent Variable:	$\mathbf{Pr}(\mathbf{migrants_{ij}} > \mathbf{0})$		$\ln(\mathbf{migrants_{ij}})$	
	(1)	**(2)**	**(3)**	**(4)**
agroclimatic similarity	0.0027	0.0015	-0.0004	0.0001
	(0.0066)	(0.0069)	(0.0200)	(0.0220)
$(-1)\times$ log distance	0.1262	0.1272	0.1287	0.2036
	(0.0192)***	(0.0238)***	(0.0597)**	(0.0753)***
Observations	96,866	96,866	37,446	37,446
Dep. Var. Mean (Levels)	.39	.39	16.8	16.8
Birth District (Java/Bali) Fixed Effects	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes
Year of Settlement Fixed Effects	Yes	Yes	Yes	Yes
Individuals Placed in Year of Settlement	Yes	Yes	Yes	Yes
Predetermined 1978 Controls, Destinations	No	Yes	No	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. The unit of observation is an origin district $_i$ (of which there are 119) by destination Transmigration village $_j$. All specifications include birth district fixed effects, destination island fixed effects, the log number of transmigrants placed in the initial year of settlement, and indicators for the year of settlement. Columns 2 and 4 additionally control for the predetermined district-level variables. Results unchanged with destination district or village FE. Standard errors are two-way clustered by birth district and destination district.

Roadmap

Higher Agroclimatic Similarity $\Longrightarrow$ Higher Rice Productivity

1 SD $\Uparrow \mathcal{A}_j \implies$ 20% $\Uparrow$ rice productivity (0.5 tons/Ha at mean of 2.5)

Specification	**Baseline**	**Drop $\mathbf{x}_j$**	**+ Origin Controls**	**+ Destination Controls**	**+ Both Controls**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
agroclimatic similarity	0.204	0.182	0.210	0.151	0.166
	(0.064)***	(0.045)***	(0.075)***	(0.057)***	(0.068)***
Number of Villages	600	600	600	600	600
R^2	0.252	0.149	0.277	0.367	0.400
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes
$\mathbf{x}_j$ Natural Advantage Controls	Yes	No	Yes	Yes	Yes
Origin Predetermined Controls	No	No	Yes	No	Yes
Destination Predetermined Controls	No	No	No	Yes	Yes

Notes: Agroclimatic similarity has a mean of 0.67 and standard deviation of 0.14, but is normalized to have mean zero and a standard deviation of one. Standard errors allow for unrestricted correlation between all villages within 150 km of each other (Conley, 1999). */**/*** significance at 10/5/1 %.

- selection on unobservables 'highly unlikely' (Altonji et al, 2005; Bellows/Miguel, 2009) (ratios from 4.9 in column 1 to 10.9 in column 5 vs. heuristic threshold of 3.6)
- individual selection out of rice farming would have to be implausibly large (order of magnitude larger than actual effect of similarity on occupational choice)

Interpretation: **Skill Specificity**

- effect size $\approx 2\times$ productivity gap between no school and junior sec.
  $\approx$ annual staple calories at subsistence

$\Longrightarrow$ large effects of location-specific human capital on rice productivity

- Null effects on cash crop productivity (baseline specification):

$$y_j = \underset{(0.049)}{\mathbf{0.024}}\ \mathcal{A}_j + \mathbf{x}_j'\boldsymbol{\beta} + \nu_j$$

  revenue-weighted across crops with mean of 1.0 (Jayachandran, 2006)
  (FAO national price, 28 cash crops, esp. palm oil, rubber, cocoa, coffee)

- Formally reject equality with 0.204 effect for rice (p-value$<$ 0.001)

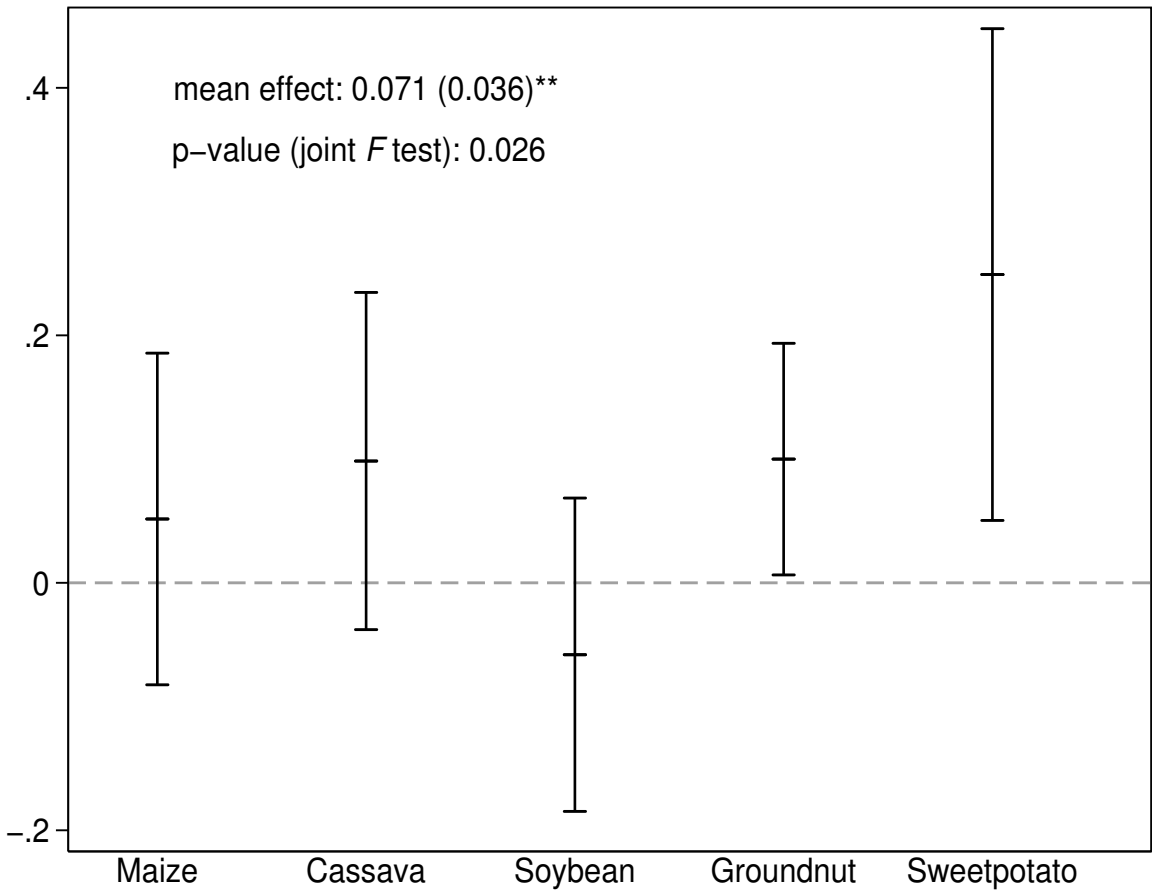
$\Longrightarrow$ $\mathcal{A}_j$ not proxying for unobservable general productivity

- Why the null? Cash crop require less complex, less labor-intensive, and fewer location-specific agroclimatic management practices

Interpretation: **Skill Specificity**

Relatively More Substitutable Food Crops

- secondary food crops (*palawija*) common across Indonesia
- *palawija* farmers in Java/Bali often switch to rice when rice prices ↑



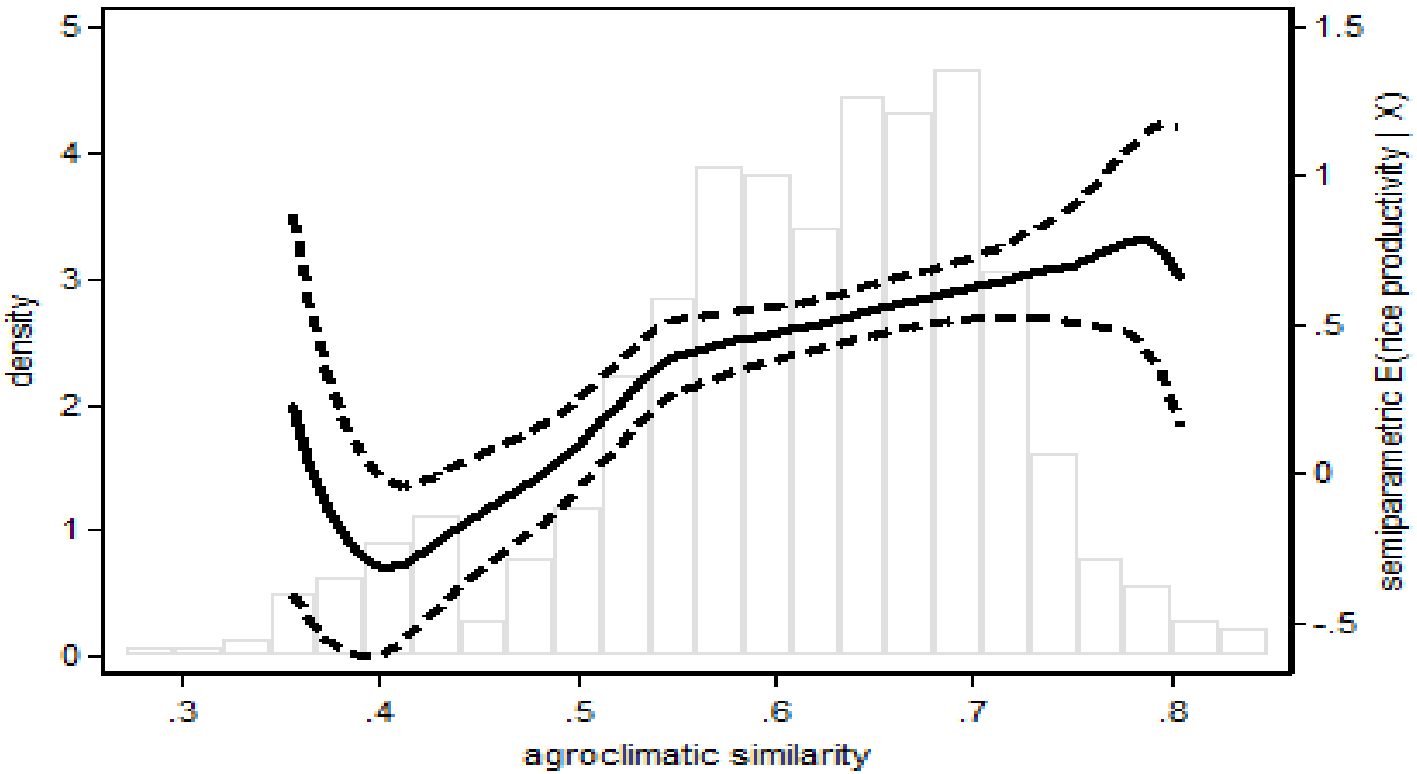
Notes: 90% confidence interval from baseline specification. Conley (1999) standard errors with 150km bandwidth. p-value based on hypothesis test of cross-equation restriction. Mean effect based on Katz et al (2007) approach.

Nonlinearity $\Longrightarrow$ Concave Adjustment Process

γ elasticity increasing in agroclimatic distance from origin

$$y_j = \alpha + g(\mathcal{A}_j) + \mathbf{x}_j'\boldsymbol{\beta} + \nu_j$$

where $g(\cdot)$ is estimated **semiparametrically** following Robinson (1988)



Notes: Semiparametric extensions of the main parametric specification for agroclimatic similarity. The dashed lines correspond to 90% confidence intervals. The estimates are based on local linear Robinson (1988) regressions with an Epanechnikov kernel and a bandwidth of 0.05. The histogram captures the distribution of standardized agroclimatic similarity. The top 5 and bottom 5 villages are trimmed for presentational purposes.

Robustness Checks

- **Program Features**
  - ▷ province × year of settlement fixed effects
  - ▷ number of transmigrants placed
  - ▷ number and concentration of origin districts
  - ▷ within-transmigrant ethnic fractionalization

- **Index Construction**
  - ▷ different population weights
  - ▷ different distance metrics

- **Aggregation Bias**
  - ▷ controlling for share of natives in village-level regression
  - ▷ household-level regression using auxiliary small sample survey

▸ tables

Heterogeneous Effects of Skill Transferability

1. **Average** — 1 SD $\Uparrow \mathcal{A}_j \implies$ 20% $\Uparrow$ rice productivity
2. **Heterogeneity** — Stronger effects in adverse growing conditions
   Soil-specific skills relatively less transferable
3. **Adaptation**
4. **Policy**

Similarity More Important in Adverse Locations

adverse growing conditions = low potential yield, dryland

Dep. Var.: rice productivity	(1)	(2)
agroclimatic similarity	0.424	
	(0.112)***	
$\cdots \times$ log potential rice yield	-0.536	
	(0.175)***	
$\cdots \times$ tercile 1 wetland share $\in [0, 0.16]$		0.355
		(0.079)***
$\cdots \times$ tercile 2 wetland share $\in (0.16, 0.66]$		0.141
		(0.059)**
$\cdots \times$ tercile 3 wetland share $\in (0.66, 1.0]$		0.059
		(0.120)
Number of Villages	599	600
R^2	0.327	0.340
Island Fixed Effects	Yes	Yes
$\mathbf{x}_j$ Natural Advantage Controls	Yes	Yes
Origin Predetermined Controls	Yes	Yes
Destination Predetermined Controls	Yes	Yes

Notes: Transmigration villages are split into terciles of the fraction of total farmland area that is wetland (*sawah*) as reported in 2003. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999). */**/*** denotes significance at the 10/5/1 percent level.

Which Skills Are Transferable?

Major Tasks: Land Preparation, Water and Soil Management

Decomposition of Component-Specific Skills

Dep. Var.: rice productivity	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
agroclimatic similarity	0.166				
	(0.068)***				
topographic similarity		0.070			0.033
		(0.071)			(0.078)
water condition similarity			0.041		0.001
			(0.071)		(0.089)
soil content similarity				0.188	0.172
				(0.079)**	(0.091)*
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes
$\mathbf{x}_j$ Natural Advantage Controls	Yes	Yes	Yes	Yes	Yes
Origin Predetermined Controls	Yes	Yes	Yes	Yes	Yes
Destination Predetermined Controls	Yes	Yes	Yes	Yes	Yes

Notes: Topography: elevation, ruggedness, slope. Water: drainage, rainfall, temperature, distance to river. Soil Nutrients: soil texture, distance to coast, carbon content, sodicity, topsoil pH. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999). */**/*** denotes significance at the 10/5/1 percent level.

similar patterns for substitutable *palawija* crops ▸ table

Skill Transferability and Adaptation Mechanisms

1.	**Average**	1 SD $\Uparrow \mathcal{A}_j \implies$ 20% $\Uparrow$ rice productivity
2.	**Heterogeneity**	Stronger effects in adverse growing conditions Soil-specific skills relatively less transferable
3.	**Adaptation**	Interacting with natives (linguistic similarity) Occupational sorting Crop choice/switching Migration: limited
4.	**Policy**	

Interacting with Natives

transmigrants' languages similar to nearby natives $\Longrightarrow \Uparrow$ rice productivity

Linguistic similarity, $\mathcal{L}_j \in [0,1]$: language map/tree, ethnic π shares ▸ details

Dep. Var.: rice productivity	**(1)**	**(2)**
agroclimatic similarity	0.166	0.150
	(0.068)***	(0.061)**
linguistic similarity		0.258
		(0.088)***
Number of Villages	600	600
R^2	0.400	0.410
Island Fixed Effects	Yes	Yes
$\mathbf{x}_j$ Natural Advantage Controls	Yes	Yes
Origin Predetermined Controls	Yes	Yes
Destination Predetermined Controls	Yes	Yes

Notes: Similarity measures are normalized to mean zero, standard deviation one. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999). */**/*** denotes significance at the 10/5/1 percent level. Agroclimatic and linguistic similarity have a very small correlation, $\text{corr}(\mathcal{A}_j, \mathcal{L}_j) = -0.03$.

Occupational Sorting

Testing a Simple 2×2 Setup with Micro Data

- **Two occupations**: farming and trading/services
- **Two skills**: agricultural and language
- Farming is relatively agricultural skill intensive
- Trading is relatively language skill intensive

Predictions: Sorting Based on Comparative Advantage

- ▷ high agroclimatic similarity $\Longrightarrow \Uparrow Pr(i = farmer)$
- ▷ high linguistic similarity $\Longrightarrow \Uparrow Pr(i = trader)$

Sorting Patterns Consistent with Comparative Advantage

Pr(Occupation = . . .)	**Farming**	**Trading/ Services**
	(1)	**(2)**
individual agroclimatic similarity	0.0090	-0.0037
	(0.0052)*	(0.0027)
individual linguistic similarity	-0.0139	0.0175
	(0.0161)	(0.0067)**
Number of Individuals	566,956	566,956
Dependent Variable Mean	0.622	0.099
Island FE	Yes	Yes
Year of Settlement FE	Yes	Yes
Individual-Level Controls	Yes	Yes
Village-Level Controls	Yes	Yes

Notes: Linear probability estimates for all Java/Bali-born individuals aged 15-65 in Transmigration villages in the 2000 Population Census. Both similarity measures are normalized to have mean zero and a standard deviation of one. All regressions include: (i) fixed effects for the year of settlement, (ii) predetermined village-level controls used in previous tables, and (iii) age interacted with other demographic characteristics. Standard errors clustered by district in parentheses. */**/*** denotes significance at the 10/5/1 percent level.

Crop Adjustments Matter

Dependent Variable:	**revenue weight on**		**share of cash**	**total agric.**
	cash crops	**rice**	**crop farmers**	**productivity**
	(1)	**(2)**	**(3)**	**(4)**
agroclimatic similarity	-0.043	0.047	0.001	0.014
	(0.021)**	(0.017)***	(0.022)	(0.079)
Number of Villages	770	770	770	770
R^2	0.156	0.161	0.229	0.086
Dep. Var. Mean (Levels)	0.572	0.273	0.348	0.996
Island Fixed Effects	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{x}_j$)	Yes	Yes	Yes	Yes
Origin Predetermined Controls	Yes	Yes	Yes	Yes
Destination Predetermined Controls	Yes	Yes	Yes	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. Agroclimatic similarity is normalized to have mean zero and a standard deviation of one. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

- **Strong caveats regarding data in cols. 2-4**
  - more ideal weights based on labor, local prices, profits
  - unusually high national price of cash crops relative to rice in 2001/2

Medium-Run Effects on Nighttime Light Intensity in 2010

Persistent Effects on Development: Costly, Incomplete Adjustment

1. **Average** — 1 SD $\Uparrow \mathcal{A}_j \implies$ 20% $\Uparrow$ rice productivity
2. **Heterogeneity** — Stronger effects in adverse growing conditions (drylands, low potential productivity)
3. **Adaptation** — Interacting with natives (linguistic similarity)
   Occupational sorting
   Crop choice/switching
   Migration: limited ▸ details
   Significant effects on light intensity (proxy for income)
4. **Policy**

Medium-Run Effects on Nighttime Light Intensity in 2010

Persistent Effects on Development: Costly, Incomplete Adjustment

	Dep. Var.: nighttime light ...			
	coverage		**intensity**	
	(1)	**(2)**	**(3)**	**(4)**
agroclimatic similarity	0.016	0.043	0.205	0.391
	(0.007)**	(0.008)***	(0.050)***	(0.044)***
Number of Villages	814	814	814	814
Dep. Var. Mean	0.08	0.08	0.75	0.75
Estimator	OLS		Poisson	
Island Fixed Effects	Yes	Yes	Yes	Yes
Full Set of Predetermined Controls	No	Yes	No	Yes
Year of Settlement FE	Yes	Yes	Yes	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

- 1 SD $\Uparrow$ agroclimatic similarity $\Longrightarrow$ 6-12% higher income

  (Henderson et al, 2012 applied to district GDP in Indonesia by Gibson & Olivia, 2015)

Policy Evaluation

Ex Post Optimal Assignment and Null Average Treatment Effects

1. **Average** — 1 SD $\Uparrow \mathcal{A}_j \implies$ 20% $\Uparrow$ rice productivity

2. **Heterogeneity** — Stronger effects in adverse growing conditions
   Soil-specific skills relatively less transferable

3. **Adaptation** — Interacting with natives (linguistic similarity)
   Occupational sorting
   Crop choice/switching
   Migration: limited ▸ details
   Significant effects on light intensity (proxy for income)

4. **Policy** — Reallocation to $\Uparrow$ agroclimatic similarity $\implies$ 27% $\Uparrow$ rice yields ▸ details
   Small average impact of program on local development ▸ details
   ▷ identification: planned but unsettled villages as counterfactuals

Skill Transferability Matters for Aggregate Productivity

Key Takeaways

1. **Location-specific human capital and labor (mis)allocation**
   - ▷ first quasi-experimental estimate of skill transfer elasticity in shaping successful expansion of agricultural settlement frontier
   - ▷ skill specificity $\Longrightarrow$ smaller gains from labor reallocation

2. **Implications for resettlement, increasingly important**
   - ▷ many governments have begun planning for climate change with huge numbers expected to be displaced (IPCC, 2014)
   - ▷ transmigrants: type of people most likely affected by climate change
   - ▷ **lessons**: avoid very bad matches; extension services; language skills

Ethnic Diversity and Nation Building

Ongoing Work on New Paper

How does ethnic diversity shape nation building and development?

- We exploit three plausibly exogenous sources of diversity
  1. mix of transmigrant ethnic groups within settlements
  2. share of natives in nearby settlements (conditional on $\mathbf{x}_v$, N_{v0})
  3. linguistic distance between transmigrants and nearby natives

- National language, *Bahasa Indonesia*, as technology with social and economic returns and associated adoption externalities
  - ▷ model of language diffusion $\Longrightarrow$ multiple adoption equilibria

- Social outcomes mediated by language: interethnic marriage, conflict
  - ▷ also, survey-based measures of trust and cooperation

- Still exploring economic outcomes, e.g., occupational diversity

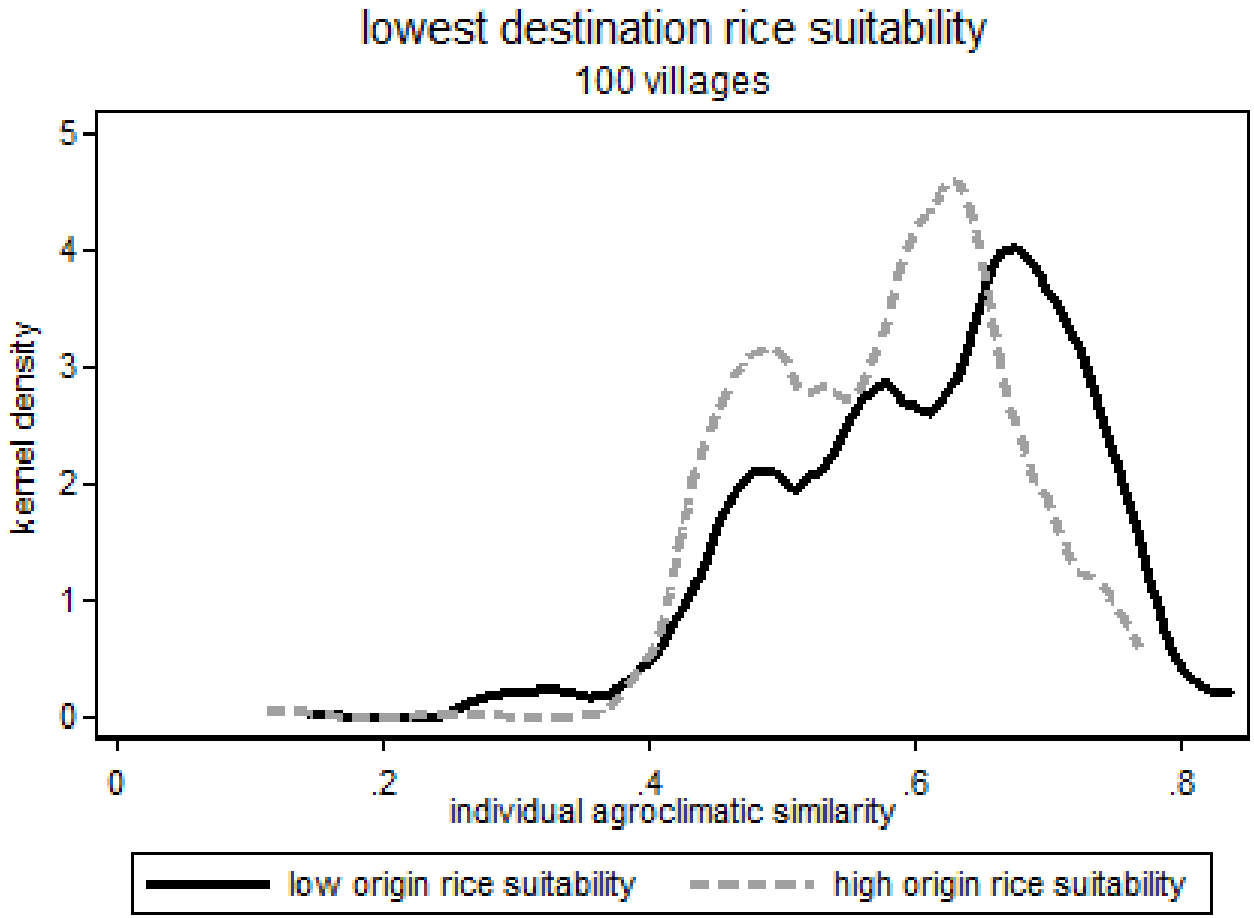
THANK YOU

sbazzi@bu.edu

APPENDIX

Evidence Against Rice-Specific Natural Advantages

Low potential productivity origins have higher agroclimatic similarity in low potential productivity destinations than high potential productivity origins



Notes: Individual-level agroclimatic similarity compared across migrants from the 20 out of 119 districts of Java/Bali with the lowest potential rice productivity versus those from the top 20 districts. Sample is restricted to the 100 Transmigration villages with the lowest potential rice productivity.

back

Which Skills Are Transferable?

Decomposition of Component-Specific (Management) Skills

	topography	**water**	**soil**
	(1)	**(2)**	**(3)**
Dep. Var.: ... productvity			
maize	-0.031	0.042	0.078
	(0.087)	(0.061)	(0.083)
cassava	0.058	0.003	0.131
	(0.075)	(0.066)	(0.099)
soybean	-0.002	-0.091	0.095
	(0.086)	(0.074)	(0.097)
groundnut	0.024	0.011	0.201
	(0.066)	(0.056)	(0.050)***
sweet potato	0.015	0.157	0.312
	(0.121)	(0.054)***	(0.140)**
joint F test maize...sweet potato=0	0.28	3.68	2.02
p-value	[0.92]	[0.01]***	[0.09]*
Mean Effect	0.009	0.019	0.137
	(0.025)	(0.024)	(0.053)***
Island Fixed Effects	Yes	Yes	Yes
$\mathbf{x}_j$ Natural Advantage Controls	Yes	Yes	Yes

Notes: Each cell is a separate regression. Topography: elevation, ruggedness, slope. Water: drainage, rainfall, temperature, distance to river. Soil Nutrients: soil texture, distance to coast, carbon content, sodicity, topsoil pH. Mean effect based on the Katz et al (2007) procedure. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999). */**/*** denotes significance at the 10/5/1 percent level.

back

Resettlement as an Optimal Assignment Problem

back

- **Counterfactual**: What is the productivity gain from an optimal allocation of migrants given the importance of agroclimatic similarity ($\widehat{\gamma}$ elasticity)?
- Generalized Assignment Problem is NP-hard (Fischer et al, 1986)
- Each transmigrant has g dimensional vector of (origin) attributes $\mathbf{x}_i$
- **Objective**: maximize total rice output

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \sum_{j=1}^{814} y_j$$

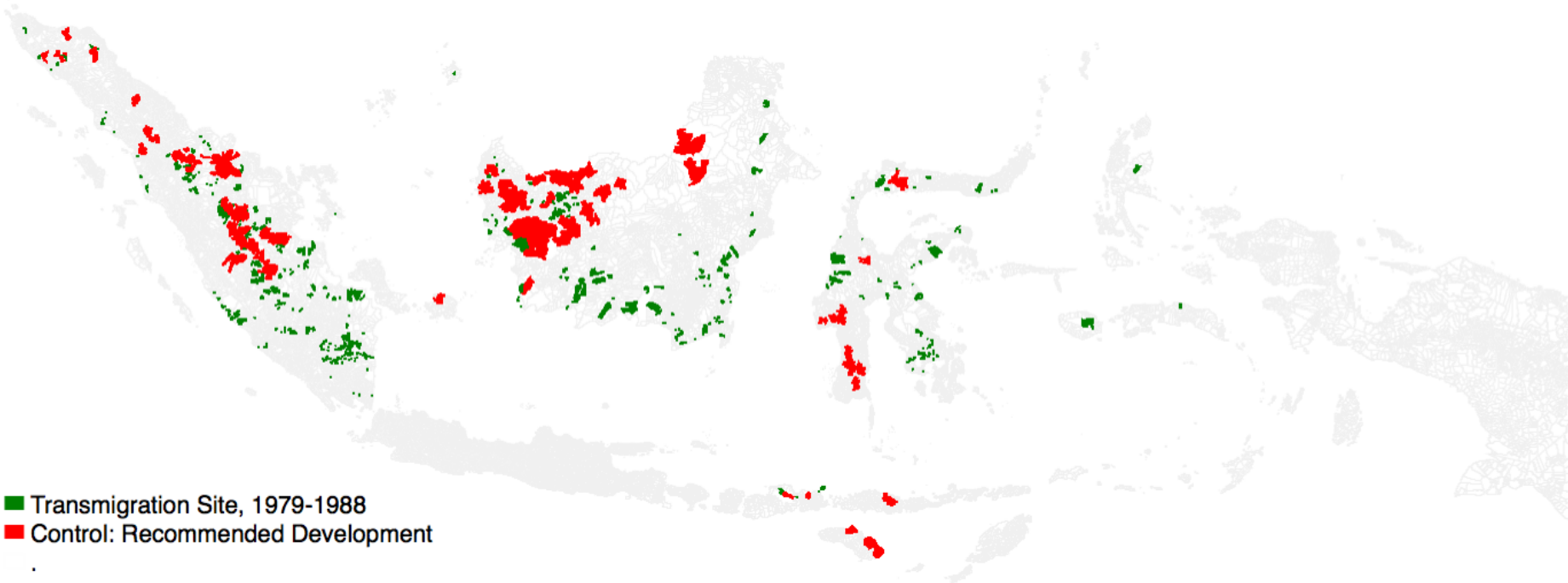
  where $\mathbf{W}$ is a matrix that assigns each i (transmigrants) to unique j (village).

- **Constraint**: $\sum_{i=1}^{N} W_{ij} = M_j$ for all $j = 1, \ldots, J$ where M_j is the number of slots (carrying capacity) in site j
- Solution concept (simplified): **"greedy" assignment** algorithm

$\Longrightarrow$ **total rice yields 27% higher than realized**

What Was the Impact of Transmigration on the Outer Islands?

Average Treatment Effects back



- Oil price $\downarrow \implies$ policy discontinuity $\implies$ counterfactual settlements

$$y_j = \alpha + \beta T_j + \mathbf{x}_j' \boldsymbol{\beta} + \widetilde{\nu}_j$$

where $T_j = 1$ if Transmigration (treated) village, $T_j = 0$ if control

  - 814 treated villages, 608 control villages ($>$ 10km from treated villages)
  - $\mathbf{x}_j$: predetermined site selection variables

Reweighting Planned but Untreated Villages

back

- Place-based evaluation: **reweighting** control villages
  (see Kline, 2011; Kline and Moretti, 2014; Busso et al, 2013)

- Reweighting control villages by $\widehat{\kappa} = \widehat{P}_j/(1 - \widehat{P}_j)$ where

$$\widehat{P}_j \equiv \mathbb{P}(T_j = 1) = \Lambda(\mathbf{x}_j'\widehat{\boldsymbol{\zeta}})$$

$\Longrightarrow$ balance along predetermined agroclimatic attributes table

Interpretation

Transmigration villages chosen randomly from eligible areas (conditional on observables)

Average Treatment Effects

back

Dependent Variable	(1)	(2)	(3)	(4)
log population density	-0.390	0.556	0.799	0.769
	(0.118)***	(0.132)***	(0.220)***	(0.170)***
any rice production	-0.041	-0.094	-0.027	-0.029
	(0.036)	(0.035)***	(0.059)	(0.060)
log rice productivity	-0.316	-0.241	-0.035	-0.166
	(0.099)***	(0.134)*	(0.175)	(0.218)
log revenue-weighted avg. agricultural productivity	-0.051	-0.193	0.023	0.134
	(0.083)	(0.136)	(0.172)	(0.142)
log revenue-weighted total agricultural output	0.641	0.170	0.410	0.472
	(0.134)***	(0.186)	(0.247)*	(0.258)*
percent any light coverage, 2010	-0.187	0.008	0.018	0.009
	(0.030)***	(0.017)	(0.033)	(0.025)
Treatment/Control Only	No	Yes	Yes	Yes
Geographic Controls	No	Yes	Yes	Yes
Reweighting	No	No	Yes	Yes
Blinder-Oaxaca	No	No	No	Yes

Notes: Each cell reports the ATE on the given dependent variable. Agricultural outcomes are as observed for the 2001 growing season. All specifications include island fixed effects. Standard errors clustered by district in parentheses. */**/*** denotes significance at the 10/5/1 percent level.

Return Migration Was Low

back

- Bounding Outer Island-to-Inner Island transmigrant returns:
  - 1985 intercensal survey: migration origin district to destination district
  - 5-year migrant: 30,000 households from district with Transmigration site
    $\Longrightarrow$ (very large) upper bound on 365,000 households placed thru 1984
  - similarly, not explained by gravity forces

- Also, $\mathcal{A}_j \perp\!\!\!\perp \Delta \ln(\text{settlers placed/resettled}), \ln(\text{population in 2000})$

- And, no systematic outmigration (on $\mathcal{A}_{ij}$) to nearby Outer Islands cities

- Why?
  - Not typical migrant
  - Land ownership (but had to wait for title and ability to sell)
  - 1984 survey finds 71% (11%) report higher (equal) income

Transmigrants Are Slightly Negatively Selected

This is the Relevant Population from a (Resettlement) Policy Perspective

	Years of Schooling Relative to Java/Bali-born Stayers in Transmigration-eligible Cohort			
	2000 Census		1985 Inter-Census	
	(1)	**(2)**	**(3)**	**(4)**
Migrant to Transmigration site	-0.650	-0.731	-1.179	-1.044
	(0.136)***	(0.088)***	(0.272)***	(0.229)***
Migrant to other Outer Islands rural area	3.267	2.407	3.272	2.600
	(0.122)***	(0.087)***	(0.256)***	(0.368)***
Migrant to other Outer Islands urban area	4.057	3.186	3.672	3.134
	(0.127)***	(0.111)***	(0.168)***	(0.216)***
Migrant to Java/Bali rural area	-0.212	-0.411	-1.014	-0.924
	(0.140)	(0.093)**	(0.187)***	(0.141)***
Migrant to Java/Bali urban area	3.762	2.652	2.709	2.138
	(0.177)***	(0.149)***	(0.278)***	(0.276)***
Number of Individuals	41,201,749	41,201,749	39,766,326	39,766,326
Age FE	No	Yes	No	Yes
Birth District FE	No	Yes	No	Yes

Regression of years of schooling on mutually exclusive dummy variables indicating type of migrant with non-migrants as the reference. Standard errors clustered at the district level. */**/*** denotes significance at the 10/5/1 percent level.

back

Agroclimatic Diversity in Origins and Destinations

	Villages in [...]			
	Java/Bali		Outer Islands	
	Mean	Std. Deviation	Mean	Std. Deviation
Topography				
ruggedness index	0.167	(0.169)	0.273	(0.159)
elevation (meters)	241.0	(316.8)	271.8	(376.9)
% land with slope between 0-2%	0.391	(0.358)	0.268	(0.296)
% land with slope between 2-8%	0.394	(0.270)	0.373	(0.245)
% land with slope between 8-30%	0.170	(0.237)	0.238	(0.238)
Soil Quality				
organic carbon (%)	0.021	(0.017)	0.033	(0.043)
topsoil sodicity (esp, %)	0.014	(0.003)	0.015	(0.005)
topsoil pH (-log(H+))	6.256	(0.686)	5.446	(0.748)
coarse texture soils (%)	0.045	(0.139)	0.060	(0.160)
medium texture soils (%)	0.528	(0.258)	0.699	(0.227)
poor or very poor drainage soils (%)	0.285	(0.315)	0.275	(0.335)
imperfect drainage soils (%)	0.076	(0.181)	0.135	(0.262)
Climate				
average annual rainfall (mm), 1948-1978	198.8	(56.1)	205.2	(49.3)
average annual temperature (Celsius), 1948-1978	24.8	(2.8)	25.3	(2.8)
Water Access				
distance to nearest sea coast (km)	27.3	(20.0)	37.2	(39.6)
distance to nearest river (km)	2.5	(5.6)	5.4	(12.0)

back

Measuring Linguistic Similarity

More than 700 Languages Across Indonesia

$$\mathcal{L}_j = \sum_{\ell=1}^{8} \pi_{\ell j} \left(\frac{branch_{\ell j}}{max\ branch} \right)^{\psi}$$

- $\pi_{\ell j}$: share of Java/Bali-born migrants in j from linguistic group ℓ
- $branch_{j\ell}$: sum of shared language tree branches (*Ethnologue*) between Java/Bali language ℓ and native language in j's region
- Caveat: $max\ branch = 7$ (Java/Bali languages close relatives)
- Functional form akin to Fearon (2003), $\psi = 0.5$

back

Robustness Checks: Rice Productivity

back

	agroclimatic similarity
1. Baseline Specification	0.204
	(0.064)***
2. Total Transmigrants Placed in Initial Year	0.205
	(0.064)***
3. Year of Settlement Fixed Effects	0.200
	(0.063)***
4. Province $\times$ Year of Settlement Fixed Effects	0.114
	(0.065)*
5. Controlling for Java/Bali-born Pop. Share and Overall Pop. Density	0.211
	(0.063)***
6. 3rd Degree Polynomial in Latitude/Longitude	0.193
	(0.077)**
7. Alternative Normalization of Agroclimatic Similarity Index	0.192
	(0.060)***
8. Euclidean Distance in Agroclimatic Similarity Index	0.161
	(0.086)*
9. Only pre-1995 Java/Bali Immigrants in Agroclimatic Similarity Index	0.206
	(0.067)***
10. Only Java/Bali-born age $>$30 in Agroclimatic Similarity Index	0.212
	(0.060)***

Robustness Checks: Rice Productivity

back

	(1)	(2)
agroclimatic similarity	0.166	0.156
	(0.068)**	(0.064)**
within-Java/Bali ethnic fractionalization		-0.032
		(0.053)
Herfindahl Index, Java/Bali origin district shares		0.039
		(0.061)
number of Java/Bali origin districts		-0.017
Number of Villages	600	600
R^2	0.318	0.320
Island Fixed Effects	Yes	Yes
Predetermined Village Controls ($\mathbf{x}_j$)	Yes	Yes
Predetermined Destination Controls	Yes	Yes

Main Results Using Individuals

back

Ruling Out Aggregation Bias

We can re-estimate our rice productivity regression at the individual level

$$y_{ej} = \alpha + \gamma_a \mathcal{A}_{ej} + \mathbf{x}_j' \boldsymbol{\beta} + \varepsilon_{ej}$$

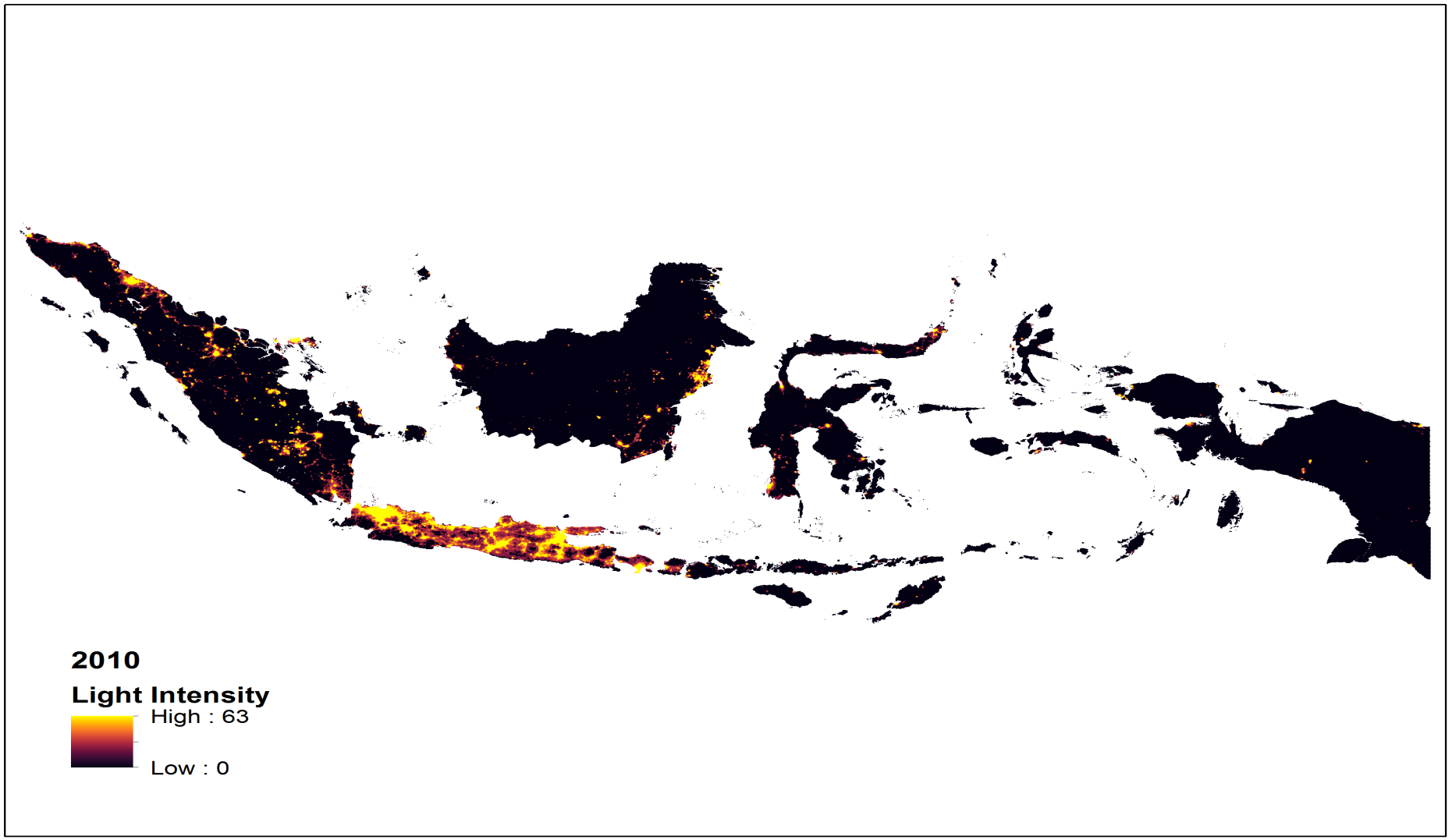
using individuals from a random sample of 74 Transmigration villages

Dependent Variable: **log rice productivity**

	(1)
agroclimatic similarity	0.150
	(0.073)**
Java/Bali-born household head	446
$\mathbf{x}_j$ natural advantage controls	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. Individual-level regressions of log rice output per hectare for individuals (household heads) living in a random sample of 74 Transmigration villages in a nationally representative household survey (*Susenas*) conducted in 2004. Agroclimatic similarity is defined at the individual-level based on an origin-weighted average of the ethnicity-specific agroclimatic similarity prevailing across individuals in the village as observed using the full 2000 Population Census.

Light Intensity and Growth Across Indonesia, 1992–2010



Notes: Data calculated from the Henderson et al (2012) satellite pixel data.

back

Reweighting Control Villages $\Longrightarrow$ Balance

Propensity Score Estimates

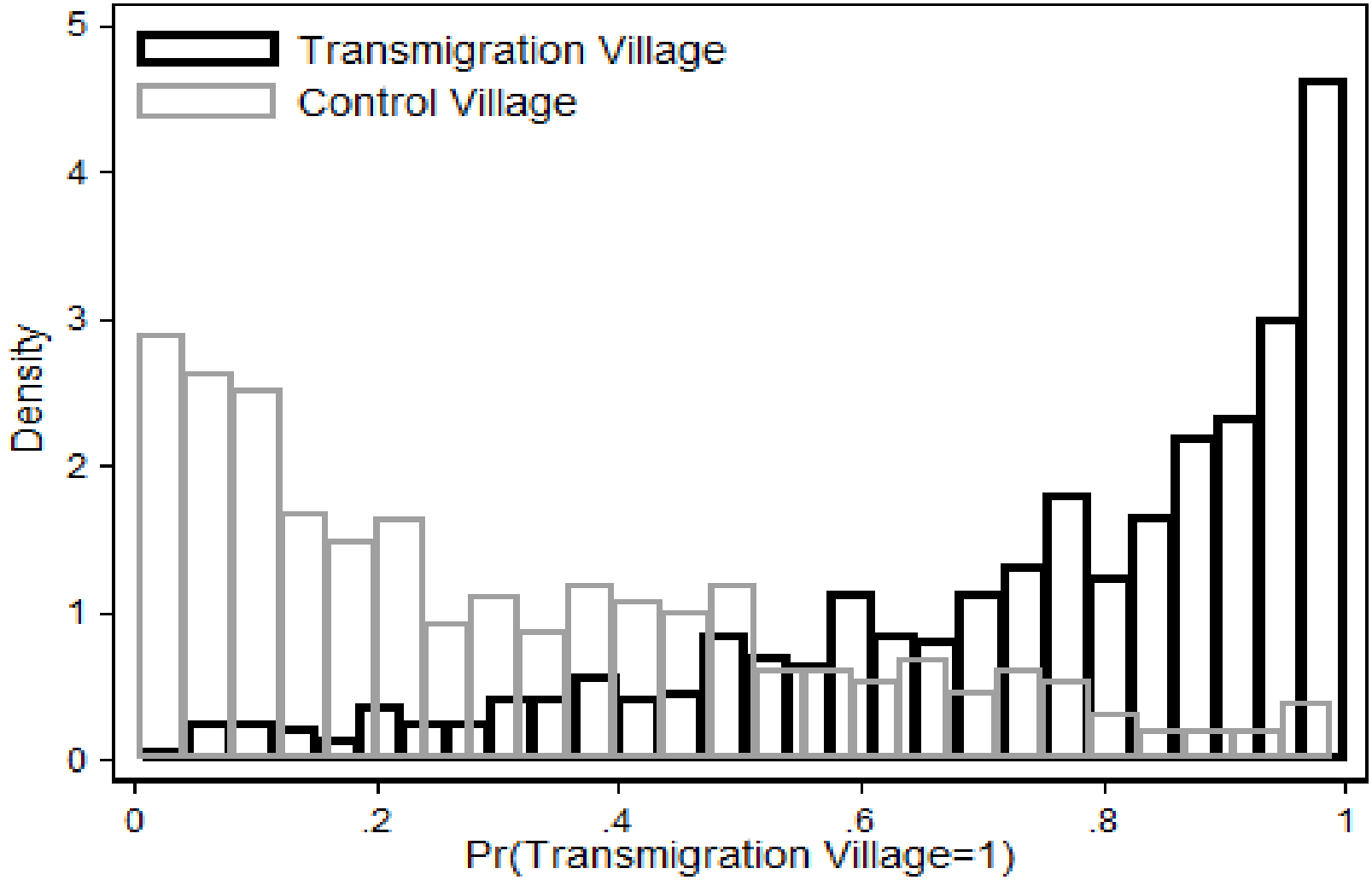
	(1)	(2)	(3)	(4)
Treated/Control Radius	0 km		10 km	
Treated Villages	Yes	Yes	Yes	Yes
Control Villages	Yes	Yes	Yes	Yes
Other Villages	Yes	No	Yes	No
% w/ slope between 0-2%	-0.000	0.006	0.000	0.002
	(0.000)	(0.002)***	(0.001)	(0.001)**
log altitude, m^2	0.000	-0.026	-0.002	-0.018
	(0.001)	(0.009)***	(0.004)	(0.008)**
Organic Carbon (%)	0.002	-0.020	0.006	-0.010
	(0.001)**	(0.006)***	(0.002)***	(0.007)
Topsoil pH (-log(H+))	-0.007	-0.145	-0.023	-0.155
	(0.008)	(0.051)***	(0.020)	(0.041)***
Coarse texture soils (%)	-0.005	-0.048	-0.062	0.108
	(0.024)	(0.223)	(0.066)	(0.214)
Imperfect drainage soils (%)	0.028	-0.219	0.084	-0.132
	(0.016)*	(0.134)	(0.036)**	(0.100)
Average. rainfall, 1948-1978	0.000	-0.002	0.001	-0.001
	(0.000)**	(0.001)	(0.000)*	(0.001)**
Avgerage Temp. (Celsius), 1948-1978	0.004	-0.024	0.016	0.002
	(0.002)**	(0.014)*	(0.005)***	(0.012)
Distance to Nearest Major Road	0.004	-0.265	-0.366	-0.255
	(0.036)	(0.166)	(0.113)***	(0.165)
Distance to Nearest Coast	0.018	-0.057	0.034	-0.065
	(0.005)***	(0.037)	(0.014)**	(0.029)**
Distance to Nearest River	0.004	-0.008	-0.009	-0.023
	(0.003)	(0.022)	(0.007)	(0.013)*
Distance to District Capital	0.016	0.029	0.034	0.014
	(0.004)***	(0.028)	(0.009)***	(0.017)
N	27119	1500	27119	5032
Pseudo R^2	0.124	0.359	0.130	0.284

Notes: This table reports average marginal effects. In columns 1 and 3, the dependent variable is a binary indicator equal to one if the village is located within 0 or 10 kilometers of either a Transmigration site or a control/RDA site. In columns 2 and 4, the dependent variable is a binary indicator equal to one if the village is located within 0 or 10 kilometers, respectively, of a control/RDA site. Standard errors clustered by district in parentheses. */**/*** denotes significant at the 10/5/1 percent significance levels.

▸ back

Reweighting Control Villages $\Longrightarrow$ Balance

Overlap



Notes: These figures plot the distribution of estimated probabilities of site selection.