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

Samuel Bazzi† *Boston Univ.*
Arya Gaduh‡ *Univ. of Arkansas*
Alexander Rothenberg§ *RAND Corporation*
Maisy Wong¶ *Univ. of Pennsylvania*

August 2014

Abstract

Between 1979 and 1988, the Transmigration Program in Indonesia relocated two million voluntary migrants from the Inner Islands of Java and Bali to newly created rural settlements in the Outer Islands. We use the plausibly exogenous assignment of migrants to destinations in this large-scale policy experiment to estimate the causal impact of skill transferability across locations on aggregate productivity and local development outcomes in receiving areas. We develop a novel proxy for skill transferability that measures how similar agroclimatic endowments are between migrants' origins and destinations. This empirical strategy provides measurable and exogenous variation in comparative advantage, offering a novel solution to well-known identification problems in multi-sector Roy models. We find that Transmigration villages exhibit significantly higher rice productivity and income growth (as proxied by nighttime light intensity) one to two decades later if they were assigned migrants from regions with more similar agroclimatic endowments. We explore several mechanisms for adapting to dissimilarity and find relatively more support for adaptation via learning and crop adjustments. Our results have important policy implications, suggesting that skill specificity may constrain labor reallocation across regions that differ in potential productivity and hence perceived spatial arbitrage opportunities.

JEL Classifications: J43, J61, O12, O13, O15, R12

Keywords: Spatial Labor Allocation, Internal Migration, Sorting, Comparative Advantage, Agricultural Development

*We thank Gilles Duranton, Gustavo Fajardo, Fernando Ferreira, Gordon Hanson, Seema Jayachandran, Rob Jensen, David Lam, Florian Mayneris, Mushfiq Mobarak, Dilip Mookherjee, Ben Olken, Ferdinand Rauch, Todd Sinai, Tavneet Suri, and seminar participants at Boston University, the Federal Reserve Board, University of Michigan, Haas, UCLA, USC, Wharton, the Economic Demography Workshop, the Urban Economics Workshop at the Barcelona Institute of Economics, the Barcelona Summer Forum on Migration, the 7th Migration and Development Conference, the NBER Summer Institute, and the Stanford Institute for Theoretical Economics for helpful suggestions. We thank numerous individuals for helpful data on and insights into agricultural production and the Transmigration Program in Indonesia. Maisy Wong is grateful for financial support from the Zell-Lurie Real Estate Center and the Wharton Global Initiatives program. Samuel Bazzi is grateful for financial support from the Center on Emerging and Pacific Economies at UC, San Diego. Lee Hye Jin and Chen Ying provided excellent research assistance. All errors remain ours. The Appendix for this paper, referenced below, can be found at http://j.mp/TransmigrationAppendix.

†Department of Economics. 270 Bay State Rd., Boston, MA 02215. Email: sbazzi@bu.edu.

‡Sam M. Walton College of Business. Department of Economics. Business Building 402, Fayetteville, AR 72701-1201. Email: agaduh@walton.uark.edu.

§1200 South Hayes St., Arlington, VA 22202-5050. Email: arothenb@rand.org.

¶Wharton Real Estate. 3620 Locust Walk, 1464 SHDH, Philadelphia, PA 19104-6302. Email: maisy@wharton.upenn.edu.

1 Introduction

The transferability of human capital has important implications for labor mobility and aggregate productivity (Becker, 1962). Yet, we have limited causal evidence of these relationships because skill transferability is "notoriously difficult to measure" (Bauer et al., 2013), and available proxies are often confounded by unobserved determinants of productivity. We use a natural experiment in Indonesia to estimate the elasticity of productivity with respect to skill transferability across locations. We develop a novel proxy to investigate how transferable skills are as farmers face different agroclimatic conditions. We find that agroclimatic similarity leads to sizeable productivity differentials. This echoes research on the striking east–west spatial diffusion pattern of migrants and technologies throughout world history (Diamond, 1997; Griliches, 1957; Steckel, 1983; Comin et al., 2012).

Our empirical setting is a large rural-to-rural resettlement program in Indonesia, known as Transmigration.[1] Between 1979 and 1988, the Government of Indonesia relocated two million voluntary migrants (transmigrants) to alleviate overpopulation concerns in the Inner Islands of Java and Bali and to develop the Outer Islands. The program provided free transport to *newly created* settlements in the Outer Islands, housing, and two hectare farm plots assigned by lottery.

Our research design leverages the wide spatial scope and exogenous relocation process of the program. First, the fact that migrants from many origins were settled across many Outer Island villages is useful because agroclimatic attributes change smoothly across space, necessitating geographic coverage beyond the scope of typical resettlement programs. Second, a large spike in global oil prices in the 1970s funded a massive increase in the scale of the program at a time when institutional capacity was quite limited. Given time, information, and logistical constraints detailed in Section 2, many activities were undertaken on an ad hoc "plan-as-you-proceed" basis (World Bank, 1988). This gave rise to plausibly exogenous variation in the arrival time of new transmigrants and opening of new settlements. As a result, migrants do not appear to have been systematically assigned to settlements where their skills would be most readily transferable.

We construct a novel proxy for skill transferability, *agroclimatic similarity*, that is higher when two locations have more similar agroclimatic endowments and hence growing conditions. Farming in different environments requires different production methods and associated technical know-how. We capture many of these differences using detailed geospatial data containing predetermined measures of topography, climate, and soil characteristics (from the *Harmonized World Soil Database*). Additionally, we use geospatial data on ethnolinguistic homelands (from the *Ethnologue* data), to measure *linguistic similarity*, which is higher when the indigenous language in nearby Outer Islands villages is more similar to migrants' languages. Both measures exhibit rich variation because Indonesia is home to over 230 million people from 700 distinct ethnolinguistic groups, living on more than 1,000 different islands. Importantly, agroclimatic similarity and linguistic similarity are uncorrelated, consistent with the ad hoc planning process giving rise to quasi-experimental variation in our similarity measures.[2]

[1] Relocation programs are found in many developing countries, including China, India, and Brazil (see Kinsey and Binswanger, 1993). Examples in developed countries include the Moving to Opportunity program in the United States (Kling et al., 2007) and various (refugee) resettlement programs in the United States and Europe (Edin et al., 2003; Sarvimäki et al., 2010; Beaman, 2012; Glitz, 2012; Bauer et al., 2013).

[2] By contrast, Michalopoulos (2012) finds spatial differences in land endowments gave rise to location-specific human capital, leading to the formation of localized ethnicities over the very long-run.

Using agroclimatic similarity as a novel and exogenous measure of comparative advantage, we show that the Transmigration program provides a natural experiment in spatial labor allocation that addresses an "extremely hard" identification problem in multi-sector Roy models (Combes et al., 2011). The problem of sorting based on unobservable comparative advantage was first highlighted by Heckman and Honore (1990) and spans multiple fields in economics.[3] Farmers can transfer their human capital more successfully if destinations are more similar to their birth locations. Hence, for a given destination, a simple Roy (1951) model suggests that migrants from similar origins have greater comparative advantage and are more productive relative to migrants from dissimilar origins. The exogenous relocation process described above assigns migrants to both high *and* low similarity destinations for reasons that are uncorrelated with comparative advantage. Indeed, we find considerably more low similarity assignments in Transmigration villages, compared to non-Transmigration villages, where individuals endogenously sort into locations based on comparative advantage.

Our empirical strategy compares Transmigration villages with a high share of Java/Bali migrants from origins with similar characteristics to observably identical Transmigration villages that have a high share of migrants from dissimilar origins. We show that our key agroclimatic similarity index is unrelated to predetermined development outcomes or potential agricultural productivity. Moreover, the spatial allocation of Java/Bali migrants does not follow the type of gravity patterns associated with endogenous sorting. Finally, our results cannot be fully explained by correlations with unobserved natural advantages, differences in demographic composition, or selection on observables such as schooling.

Our main village-level dataset combines the abovementioned geospatial data with a 1998 census of Transmigration villages that we digitized, data on individuals' birth districts and ethnicities from the 2000 Population Census, and village-level agricultural activity from a 2002 administrative census. Our key village-level outcomes are rice productivity, total agricultural productivity (for all crops), and growth in nighttime light intensity from 1992 to 2010 (a proxy for income growth, see Henderson et al., 2012).

We find that skill transferability has large effects on rice productivity. An increase in the agroclimatic similarity index by one standard deviation leads on average to a 20 percent increase in productivity.[4] This translates to an additional 0.5 tons per hectare—an effect size roughly equivalent to twice the productivity gap between farmers with no schooling versus junior secondary. These large productivity effects complement recent evidence of location-specific preferences for consuming and growing staple crops (Atkin, 2013; Michalopoulos, 2012). Further heterogeneity analysis shows that skill transferability is relatively more important in adverse growing conditions. Moreover, nonlinear semiparametric estimates reveal the largest elasticities in the bottom tercile of agroclimatic similarity where villages have a high share of Java/Bali migrants from dissimilar origins. These results are important given that rice is the primary staple crop in Indonesia, and one of the objectives of the program was to expand rice output.

Our findings provide new evidence of barriers to adaptation in response to agroclimatic change. The persistence of effects over two decades is consistent with research showing that farmers face difficulties adjusting to new agroclimatic conditions over the medium run. Olmstead and Rhode (2011) describe long periods of difficult adaptation by migrant farmers settling the Western frontier in the United States, and Hornbeck (2012) identifies limited adjustment in the first two decades following the 1930s Dust

[3]Recent studies can be found in labor (Bayer et al., 2011; Dahl, 2002; Gibbons et al., 2005), spatial and urban (Combes et al., 2008), development (Foster and Rosenzweig, 1996; Qian, 2008; Suri, 2011), and trade (Costinot et al., forthcoming).
[4]Our index is scaled between zero and one, with a relatively large standard deviation of 0.14.

Bowl. These barriers are particularly salient in developing countries, where an extensive literature documents slow adjustments in the agricultural context and the importance of agroclimatic conditions in this adjustment process (e.g., Ben Yishay and Mobarak, 2014; Conley and Udry, 2010; Huffman, 2001). We contribute to this literature by quantifying the importance of agroclimatic conditions and characterizing the adaptation responses to agroclimatic change, which has been emphasized by the most recent Intergovernmental Panel on Climate Change (IPCC, 2013).

We explore several adaptation mechanisms and find relatively more support for adaptation via learning and crop adjustments. First, linguistic similarity has significant positive effects on rice productivity, and appears to be more important in places with greater scope for learning from natives, in line with a large literature on learning in the agricultural context (see Foster and Rosenzweig, 2010). Turning to crop adjustments, in contrast to the large productivity effects on rice, agroclimatic similarity has precise zero effects on the productivity of cash crops. This is intuitive since most cash crops were not grown in Java/Bali during the time of the program, and hence skill transferability was less important for productivity. However, cash crops appear to be more important in low similarity villages, and a simple accounting exercise suggests that crop adjustment is an economically important adaptation response. This is consistent with findings in Costinot et al. (forthcoming) who use a trade model to highlight the welfare enhancing effects of crop adjustment in response to climate change.

By contrast, we find relatively less evidence of occupational adjustments and ex post migration. A one standard deviation increase in agroclimatic similarity leads to a 0.9% greater likelihood of Java/Bali migrants choosing farming as their primary occupation while a one standard deviation increase in linguistic similarity leads to a 1.7% greater likelihood of migrants working in trading and services occupations where language is important. These patterns suggest sorting into occupations based on comparative advantage but the magnitudes are small. Limited occupational adjustments are consistent with Abramitzky et al. (2014) who find little evidence that early twentieth century European migrants to the United States converged with natives by means of occupational switching. Likewise, we find limited evidence of selective ex post migration from settlement areas.

Finally, we find that agroclimatic similarity has positive effects on general economic performance, as captured by light intensity growth. This measure is increasingly used in studies exploiting highly localized identifying variation as we do here (e.g., Hodler and Raschky, forthcoming; Michalopoulos and Papaioannou, 2013). We estimate that a one standard deviation increase in agroclimatic similarity leads on average to an additional 0.15 percentage points of village-level income growth annually between 1992 and 2010 (based on implied income elasticities of light intensity growth, Olivia and Gibson, 2013). Coupled with the large effects on rice productivity and null effects on cash crop productivity, these results suggest that medium-run adjustments took place but remain incomplete.

Our study contributes to the literature on migration and spatial (mis)allocation of labor in developing countries. There has been a resurgence of research on barriers to mobility (e.g., Au and Henderson, 2006; Bryan et al., forthcoming; Dinkelman, 2012; Morten, 2013; Munshi and Rosenzweig, 2013). Using modern development accounting methods and survey data for 65 countries, Young (2013) argues that rural-urban wage gaps are explained by efficient geographic sorting rather than barriers to mobility. We focus on rural-to-rural migration, which has been understudied despite its importance in overall flows (see Lucas, 1997; Young, 2013). Our key innovation is to use a natural experiment to provide causal

evidence that complementarities between heterogeneous individuals and heterogeneous places can give rise to persistent spatial productivity gaps due to imperfect skill transferability across locations.[5] This has important policy implications. Our results suggest that skill specificity may constrain labor reallocation across regions that differ in (potential) productivity and hence perceived arbitrage opportunities.

Our findings also have implications for the optimal design of resettlement policies, which are growing in importance as millions of people are at risk of displacement (de Sherbinin et al., 2011) and various governments have begun to plan for resettlement in response to climate change (IPCC, 2013).[6] We estimate average treatment effects of the Transmigration program by comparing settlement villages against planned villages that were never assigned transmigrants. These counterfactual, almost treated villages exist because the program was abruptly halted due to budget cutbacks following the sharp drop in global oil prices in the mid-1980s. Using a place-based evaluation approach akin to Kline and Moretti (2014), we find null average impacts on agricultural productivity and light intensity growth. These null effects stem in part from the importance of agroclimatic similarity. We approximate an optimal reallocation of transmigrants across settlements on the basis of agroclimatic similarity (an NP-hard problem) and find that the program could have achieved 27 percent higher aggregate rice yields. This highlights the importance of matching people (skills) to places (production environment) when designing resettlement schemes. We show that transmigrants are negatively selected compared to typical migrants but are relatively comparable to stayers in rural villages. These are the types of individuals most likely to be adversely affected by agroclimatic changes and hence for whom such policy choices are most crucial.

Beyond the general implications for skill transferability, our results are important because the agricultural sector employs 1.3 billion people globally (World Bank, 2009) and is at the core of ongoing debates about world income inequality (Caselli, 2005). The large barriers to skill transferability that we find within the agricultural sector may point to larger barriers between (rural) agricultural and (urban) non-agricultural activities. Hence, imperfect skill transferability may provide additional insights into debates about structural transformation and the agricultural productivity gap (Gollin et al., 2014; Lagakos and Waugh, 2013; Michaels et al., 2012; Young, 2013).

The remainder of the paper proceeds as follows. Section 2 provides background on the Transmigration program. Section 3 presents our key measures of agroclimatic and linguistic similarity. Section 4 describes the sample construction and primary economic outcomes of interest. Section 5 develops our theoretical framework and empirical strategy in the context of a Roy model. Section 6 presents our main results. Section 7 concludes.

2 Indonesia's Transmigration Program

Like many countries, the spatial distribution of the population has historically been highly skewed in Indonesia, and certain areas were thought to suffer from overpopulation problems. For instance, the islands of Java and Bali were home to 66.1 percent of the population in 1971 (according to the Census),

[5] The reduced form skill transfer elasticity that we estimate is also related to work on labor mobility and skill-specificity (e.g., Gathmann and Schonberg, 2010) and the literature on the speed of economic assimilation of immigrants in Israel and the United States (e.g., Abramitzky et al., 2014; Friedberg, 2000; Lubotsky, 2007).

[6] Annually, 10 million are displaced due to dam construction and urban and transportation development (World Bank, 1999). According to the Stern Review on Climate Change, the number of people at risk of displacement or migration due to slow changes in climate or fast onset disasters is large, exceeding 60 million people in South Asia alone (Stern, 2007).

despite containing only 7.3 percent of the nation's total land area. The remaining population is found across the *Outer Islands*, consisting of the vast islands of Sumatra, Sulawesi, and Kalimantan, as well as Maluku, Nusa Tenggara, and Papua in Eastern Indonesia.

Indonesia's Transmigration program was designed primarily to alleviate these perceived population pressures. Over many decades, the program relocated millions of migrants (hereafter, transmigrants) from rural areas of Java and Bali to rural areas of the Outer Islands. Planners also hoped that the program would promote nation building by integrating ethnic groups from Java/Bali with Outer Islanders, and increase national agricultural output (especially rice production) by moving farmers to previously unsettled areas (Kebschull, 1986; MacAndrews, 1978).

Our study focuses on the most intensive waves of the program, taking place between 1979 and 1988, during President Suharto's third and fourth five-year plans (or *Pelita*).[7] At that time, the government chose to support rainfed foodcrops because Indonesia was the world's largest importer of rice (the primary staple), and annual crops were the quickest to establish, promoted early self-sufficiency, and cost less than other agricultural models. All Transmigration settlements were initially under the jurisdiction of the central government, and administration transferred to local governments after a few years, usually five (World Bank, 1988).

The government provided substantial support throughout the resettlement process. Participation in the program was almost entirely voluntary.[8] Transmigrants were given free transportation to their new settlements. Participating households would sell their assets and leave for transit camps located in each of the four provinces of Java/Bali. Here, transmigrants would wait for boats to transport groups of them at a time to the Outer Islands. Program officials identified land reserves that could be cleared and prepared for agricultural use, and connected these sites to the road network. They also built houses for transmigrants, and each household received a two hectare plot of agricultural land allocated by lottery upon arrival. Settlers were given provisions for the first few growing seasons, including seeds, tools, and food. In some cases, the government provided temporary agricultural extension services and support for the development of cooperatives and other social institutions. Additionally, some land at each site (about 10 percent) was reserved for members of the indigenous population, who would move from nearby areas to take advantage of access to new land. Planners hoped that transmigrants could teach the Outer Islanders as Javanese and Balinese farmers were thought to have superior farming techniques.

2.1 Selecting People, Places, and Assignments

In this subsection, we describe the different margins of selection in the program and discuss how they relate to identification in our empirical work.

[7] The Transmigration program began during the colonial period, but it received a major overhaul in *Pelita* III (1979-1983). Less than 600,000 people were resettled under the Dutch colonizers and post-independence under Sukarno and the early Suharto years (1945-1968) (Hardjono, 1988; Kebschull, 1986). In contrast, the program resettled 1.2 million people in *Pelita* III and initially planned to move 3.75 million people in *Pelita* IV. The total program budget during *Pelita* III and IV was approximately $6.6 billion (in 2000 USD) or roughly $3,330 per person moved, which is similar to other large-scale programs implemented in developing countries at the time (see World Bank, 1982, 1984).

[8] A very small part of the program involved involuntary resettlement of households displaced by disasters and infrastructure development (Kebschull, 1986). We exclude strategic settlements in Maluku and Papua associated with the Indonesian military as part of its territorial management system (Fearnside, 1997). We also omit Papua entirely from our study, due to concerns about data quality. Only 1.5% (3.7%) of Indonesia's (Outer Islands) population lived on Papua in 2010.

Individual Selection. To participate in the program, transmigrants had to be Indonesian citizens in good physical health. The program targeted entire families for resettlement, and according to official guidelines, couples had to be legally married, with the household head between 20 and 40 years of age. In practice, most participants were poor, landless agricultural laborers, with few assets, and very little schooling (Kebschull, 1986).[9] On average, Java/Bali born individuals who moved to Transmigration settlements had 0.5 fewer years of schooling, compared to stayers in Java/Bali. By contrast, Java/Bali born individuals who moved to urban areas in Java/Bali or to the Outer Islands have 3 to 4 more years of schooling compared to stayers.[10] Therefore, government-sponsored transmigrants are more comparable to stayers than to the typical non-sponsored or spontaneous migrant.

Site Selection. Planning documents describe a three-stage site selection process: (i) large-scale mapping of agricultural viability, (ii) aerial reconnaissance to identify "recommended development areas" (RDA), and (iii) local surveys to identify the placement and carrying capacity of settlements. However, numerous reports suggested that the process was not as detailed as planners had hoped. For example, Hardjono (1988) observes "(a)s a consequence of the focus on numbers, the land use plans developed during the 1970s were totally abandoned. Transmigrants were placed on whatever land was submitted by provincial governments for settlement purposes." And an advisory report found that sites were selected and cleared on an ad hoc "plan-as-you-proceed" basis (World Bank, 1988).

Individual Assignments. The key to our identification strategy is the assignment of transmigrants to Outer Island settlements. In our data detailed below, the average Transmigration village has Java/Bali migrants from 46 sending districts (out of 119) and three different ethnic groups (out of eight). The average origin district Herfindahl index in these villages is 0.16, suggesting that concentration is not high. We provide here several reasons why time, informational, and institutional constraints gave rise to plausibly exogenous variation in the allocation of migrants across settlements.

First, sharp changes in oil prices resulted in a rapid expansion and sudden contraction of the program when oil prices rose in the late 1970s and fell in the mid-1980s. Figure 1 shows large fluctuations in the annual number of transmigrants placed coinciding with large fluctuations in the world oil price. Due to the rapid expansion, a number of major activities were taken from the Directorate General of Transmigration (DGT), and delegated to separate agencies to speed up the settlement process. Inter-agency coordination posed challenges to the careful matching of transmigrants (whose information was collected by DGT) to their Outer Island settlements (developed under the Ministry of Public Works).

Second, our discussions with program officials suggest that many planners believed that Javanese and Balinese farmers had superior farming skills. The Green Revolution had begun to transform Javanese rice agriculture but had yet to reach much of the Outer Islands by the late 1970s. It was hoped that transmigrants would transfer some of this know-how to the Outer Islands. There was little concern about matching transmigrants on the basis of agroclimatic conditions. Moreover, assigning transmigrants to destination settlements in an optimal manner would have required a large amount of data on

[9] The author interviewed 348 transmigrant families in February/March 1982 across Java, Madura, and Bali. This is the largest, most systematic survey of transmigrants prior to their departure for the Outer Islands.

[10] These figures are based on universal microdata from the 2000 Population Census.

individual farming skills, growing conditions at destination settlements, and up-to-date information on which settlements had open slots for new transmigrants. And even with ideal data, we show in Section 6.3 that optimal assignment is a special case of a generalized assignment problem, a problem in combinatorial optimization which has been shown to be NP-hard (Fischer et al., 1986). Additionally, there were plans to provide agricultural extension services as well as irrigation. However, with the budget cutbacks in the mid-1980s, many of these planned services were never provided.[11]

Third, there were only four transit camps, one for each province of Java/Bali. This meant each transit camp would comprise transmigrants from many origin districts. If the camps were located at the district-level instead, the arrival of transmigrants in the Outer Islands might be clustered by origin locations (because the boats arriving from the transit camps would include migrants from one origin district only). Province-level camps ensure a richer mix of district origins in each arrival cohort as we observe in the data. In short, a key factor in determining the plausibly exogenous allocation of migrants appears to be the coincidental arrival time of transmigrants to the transit camps and the timing of when the sites were cleared in the Outer Islands, consistent with Hardjono's description above.

Fourth, participants exerted little, if any, control over their destination in the Outer Islands. Previous studies argue that even just prior to departure, transmigrants were ill-informed about the geographical location, native ethnic group, and agricultural systems in the areas where they were sent. For instance, in Kebschull's pre-departure survey, 82 percent knew nothing about the local agroclimatic conditions.

Finally, as we argue at greater length in Section 5.2, it appears that it was difficult for transmigrants to migrate *ex post*. This was perhaps due to the illiquidity of land markets (transaction taxes are high and Indonesia ranks 107th out of 177 countries in its difficulty to register private transactions, World Bank, 2008). Moreover, transmigrants did not receive their land titles immediately as the land was still under the jurisdiction of the DGT for the first 5-10 years. Evidence from Mexico suggests that landholdings without certification tend to reduce outmigration (De Janvry et al., 2012). There are also several reasons why the relatively poor government-sponsored transmigrants may not be as mobile as typical migrants.

3 Skill Transferability and Origin-by-Destination Match Quality

We construct a novel measure of skill transferability, *agroclimatic similarity*, which captures how similar agroclimatic endowments are between migrant origins and destinations. This proxy is similar in spirit to an index developed by Gathmann and Schonberg (2010) to measure the transferability of task-specific human capital across occupations. Other studies that examine the relationship between notions of economic proximity and productivity outcomes include Ellison et al. (2010), Greenstone et al. (2010), Hsieh et al. (2013), and Moretti (2004). Additionally, we also capture linguistic similarity between origins and destinations, building on related measures in Fearon (2003), Desmet et al. (2009), and Esteban et al. (2012). We focus on these two salient dimensions of origin-by-destination match quality that were alluded to in case studies of Transmigration settlements by anthropologists throughout the 1980s. These studies mention familiarity with local agroclimatic conditions and learning from native farmers as likely key ingredients for successful economic development in resettlement areas.

[11] Davis and Garrison (1989) note that extension officers were to be stationed in transmigration sites beginning in *Pelita* III. However, as borne out in numerous field studies, extension services were rarely provided in practice, and when extension workers were present, they often lacked local expertise (Clauss et al., 1988; Donner, 1987; Kebschull, 1986; Otten, 1986).

3.1 Agroclimatic Similarity

We use data from the *Harmonized World Soil Database* (HWSD) and other sources to measure many agroclimatic characteristics, including (i) *topography* (elevation, slope, ruggedness, altitude, and distance to rivers and the sea coast), (ii) *soil* (texture, drainage, sodicity, acidity, and carbon content), and (iii) *climate* (rainfall and temperature). These characteristics, which we measure at a high spatial resolution, are fundamental components of agricultural and especially rice productivity (Moormann, 1978).[12]

There is remarkable agroecological variation across transmigrants' (potential) origins and destinations. Table 1 presents the mean and standard deviation for each of the variables separately for the villages of Java/Bali and the Outer Islands. Note that the time-varying characteristics are measured pre-program and hence not subject to concerns about reverse causality. For instance, all of the soil type information is based on data from 1971–1981. The differences between Inner and Outer Islands are particularly prominent in rice farming with more irrigated, wetland rice farming and a higher propensity to have multiple cropping seasons each year in the Inner Islands.[13]

Given location j's G-dimensional vector of agroclimatic characteristics, $\mathbf{x}_j$, the agroclimatic similarity of an individual's origin location i and her destination location j can be defined as:

$$agroclimatic\ similarity_{ij} \equiv \mathcal{A}_{ij} = (-1) \times \mathrm{d}\left(\mathbf{x}_i, \mathbf{x}_j\right)$$

where $\mathrm{d}\left(\mathbf{x}_i, \mathbf{x}_j\right)$ is the agroclimatic distance between district i and location j, using a metric defined on the space of agroclimatic characteristics.[14] We use the sum of absolute deviations as this distance metric: first, we calculate the absolute differences in each characteristic between the origins and destinations where each characteristic has been converted to z-scores. Then, $\mathrm{d}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sum_g |x_{ig} - x_{jg}|$ projects these differences in G dimensions onto the real line (by taking a mean across the normalized differences). We multiply by (-1) so that larger differences correspond to lower values of agroclimatic similarity.

We use $\mathcal{A}_{ij}$ to construct an agroclimatic similarity index for location j by aggregating using population weights:

$$agroclimatic\ similarity_j \equiv \mathcal{A}_j = (-1) \times \sum_{i=1}^{I} \pi_{ij}\, \mathrm{d}\left(\mathbf{x}_i, \mathbf{x}_j\right), \tag{1}$$

where $\pi_{ij} \in [0, 1]$ is the share of migrants residing in transmigration site j who were born in district i. Our preferred index uses all individuals born in Java/Bali to calculate these weights. To construct π_{ij}, we use the universe of microdata from the 2000 Population Census, which identifies each individual's district of birth and his or her village of current residence (see Appendix A). As a baseline, we view all individuals born in Java/Bali and living in settlements in the Outer Islands as potential transmigrants. We explore

[12] These fixed factors explain around 25 percent of within-island variation in rice productivity in the Outer Islands of the country. Details on the data sources can be found in Online Appendix A. The HWSD are available at a 1 km resolution (30 arc seconds by 30 arc seconds). These data are more detailed than other similar datasets used in the literature, such as the *Atlas of the Biosphere* data (used by Michalopoulos, 2012, among others), which is available at a 55 km resolution (0.5 degree by 0.5 degree), or the FAO's Global Agro-Ecological Zones (GAEZ) dataset (used by Costinot et al., forthcoming, among others), which is available at a 10 km resolution (5 arc minute by 5 arc minute).

[13] According to the 1983 *Podes*, 69 percent of villages in Java/Bali report some type of wetland utilized for two or more harvests compared to only 24 percent in the Outer Islands.

[14] We observe origins i at the district-level and hence construct the index based on measures of $\mathbf{x}$ in the destinations at that same spatial frequency.

other weights and distance metrics in robustness checks. Overall, the index captures a weighted average of the agroclimatic differences between village j and the origin districts of transmigrants, with weights equal to migrant shares. We refer to $\mathcal{A}_{ij}$ ($\mathcal{A}_j$) as individual- (village-) level agroclimatic similarity.

Appendix Figure B.1 helps to illustrate how the index is constructed. We highlight several agroclimatic characteristics in two nearby destination villages on the island of Sumatra and the district that sent the largest share of transmigrants to each of them. Consider the village of *Telang Sari*, which is at elevation of 3m, has an average topsoil pH of 4.73, and 25% of its soil is fine textured, and 23 percent of its residents were born in Kebumen in Central Java. This primary sending district is quite similar, where average elevation (94m), topsoil pH (5.41) and share of fine textured soil (29.5%) are all well within one standard deviation of the characteristics of *Telang Sari*.

To arrive at the agroclimatic similarity index of 0.5 for *Telang Sari*, we first convert these characteristics into z-scores, and then take the sum of the absolute differences in these characteristics, $\mathrm{d}(\mathbf{x}_i, \mathbf{x}_j)$, for each sending district i represented in *Telang Sari*. This gives us $\mathcal{A}_{ij}$. Then, to construct $\mathcal{A}_j$, we take averages by applying migrant weights for each sending district (e.g., 0.23 for Kebumen). Finally, we multiply by negative one and scale the index to lie on the unit interval. By contrast, *Nunggal Sari*, which is 36 kilometers away from *Telang Sari*, has similar agroclimatic characteristics but a primary sending district with different growing conditions, resulting in an index value more than half a standard deviation lower (its agroclimatic similarity index is 0.4 while the standard deviation is 0.14).

3.2 Linguistic Similarity

To capture linguistic similarity, we use both the *Ethnologue* data on language structure and the *World Language Mapping System* (WLMS) data on linguistic homelands to construct a measure of the distance between each of the eight ethnolinguistic groups ℓ indigenous to Java/Bali and each of the nearly 700 ethnolinguistic groups prevailing across the Outer Islands.[15] *Linguistic similarity* for village j can then be represented as

$$linguistic\ similarity_j \equiv \mathcal{L}_j = \sum_{\ell=1}^{8} \pi_{\ell j} \left(\frac{branch_{\ell j}}{max\ branch} \right)^{\psi}, \tag{2}$$

where $\pi_{\ell j} \in [0, 1]$ is the share of Java/Bali-born immigrants in j from ethnolinguistic group ℓ in Java/Bali, $branch_{j\ell}$ is the sum of shared language tree branches between ℓ and the language indigenous to village j, $max\ branch = 7$ is the maximum number of shared branches between any Java/Bali language and any native Outer Islands language, and ψ is a parameter, set to 0.5 as a baseline following Fearon (2003). As with others using these types of measures in the economics literature (e.g., Desmet et al., 2009; Esteban et al., 2012), we view linguistic proximity as reflecting not only ease of communication but also cultural proximity, shared preferences, and hence the fluidity of potential interactions between groups.

The wide geographic scope in destination sites exposed transmigrants to many different types of agroclimatic and linguistic settings. Figure 2 demonstrates the considerable spatial variation—within and between islands—in our two measures of similarity, $\mathcal{A}_j$ and $\mathcal{L}_j$, across Transmigration villages in the Outer Islands. Both measures are based on π weights constructed from the 2000 Population Census.

[15] The indigenous Java/Bali ethnicities include, in descending order of population shares in the Outer Islands: Javanese, Sundanese, Balinese, Madurese, Betawi, Tengger, Badui, and Osing. For each village j, we deem the native language to be the linguistic homeland polygon with maximum coverage of village area. See Appendix A for details.

Table 2 reports summary statistics showing that the average Transmigration village has around 2,000 people (or 140 people per square kilometer), 40 percent of whom were born in Java/Bali, and 69 percent of whom identify with ethnic groups from Java/Bali (most of whom are second generation transmigrants born in the Outer Islands).[16]

4 Administrative and Census Data Sources

Our main analysis includes 814 Transmigration villages established between 1979 and 1988 that we identify by digitizing the 1998 Transmigration Census, produced by the Ministry of Transmigration. These villages received on average 1,885 migrants in the initial year of settlement. More than half of the Transmigration villages, seen in Figure 2, are on the island of Sumatra (482 out of 814), but many are also found on Kalimantan (192) and Sulawesi (128), with smaller numbers on Maluku and Nusa Tenggara.

Our empirical work primarily relies on measures of economic development at the *village* level. Ideally, we would estimate the impact of similarity on individual-level outcomes. While a major benefit of the large spatial scope of the program is the rich variation in agroclimatic and linguistic attributes, it is also a major data constraint. Since only a few of the more than 60,000 villages in Indonesia are Transmigration villages, coverage is a major limitation. This is why all six of our main data sources are either maps or censuses. Datasets with useful individual outcome data either do not contain enough Transmigration sites (e.g., the *Indonesia Family Life Survey* covers only 23 settlement villages) or do not contain information on migration flows to villages (e.g., annual *Susenas* household surveys). Our best individual-level dataset, the 2000 Population Census covers all settlement areas and contains useful information on migration, occupation, and schooling. However, productivity outcomes, such as wages or agricultural yields, are not recorded. We therefore draw on several sources to construct key village-level outcomes (and rule out concerns with aggregation bias in Section 6.1).

First, we measure agricultural productivity using the triennial administrative census known as *Podes* (or Village Potential). The August 2002 round provides detailed information on agricultural activities including area planted and total yield for over one hundred crops in the 2001-2 growing season(s). We focus on rice, which is Indonesia's most important staple food that is consumed by nearly all households and is grown across the country.[17] As rice is the primary crop grown on Java/Bali, most transmigrants expected to be able to grow rice in their new villages, and policymakers envisaged the program as a means of increasing national rice output (see Kebschull, 1986). From Table 2, we see that on average, 70 percent of residents were employed in farming. Nearly 75 percent of Transmigration villages were growing some rice, and the average village produced 2.5 tons per hectare. We winsorize rice yields at 20 tons/ha to account for possible misreporting, but our results are robust to alternative cutoffs.

While rice is by far the most important crop in Indonesia, we also consider other crops using a measure of total agricultural productivity based on the revenue-weighted average yield across all crops

[16] This is extremely high compared to other (non-Transmigration) villages in the Outer Islands where Java/Bali-born migrants comprise 4 percent of the population, and 13 percent identify with a native Java/Bali ethnicity.

[17] According to *Podes*, rice is grown in 88% of villages in Java/Bali, 77% in Sumatra, 84% in Kalimantan, and 63% in Sulawesi. Also, 88% of villages report rice as their primary staple.

grown in the village. We follow Jayachandran (2006) and define

$$\ln \widehat{yield}_j = \sum_{c \in C_j} \left(\frac{p_c y_{cj}}{\sum_{c \in C_j} p_c y_{cj}} \right) \ln \left(\frac{y_{cj}}{a_{cj}} \right)$$

where $C_j \subseteq C$ is the subset of crops grown in village j among all $C = 37$ crops for which we have price data, p is the average national unit producer price in 2001 and 2002 from FAO/PriceSTAT, y_c is total (harvested) output, and a_c is area planted as reported in *Podes* 2002. We normalize tons per hectare for each crop to mean one for comparability across crops. Although local prices are not available, the FAO national price data cover all major crops grown in Indonesia and serve primarily to render productivity comparable across crops in potential revenue terms. The median village grows 8 out of the 37 crops. Rice is the top crop in terms of total revenue in around 30 percent of villages and among the top 3 crops in over 60 percent. Other food crops such as maize and cassava as well as cash crops such as palm oil, rubber, and cocoa are among the most important in revenue terms besides rice (see Appendix Table B.1). However, less than five percent of farmers in Java/Bali in the late 1970s were growing cash crops (according to the 1976 and 1980 (inter-)Census), and even in the early 2000s, coffee was the only major cash crop grown in more than two percent of villages across Java/Bali (according to *Podes*).

Broader economic development, beyond agricultural productivity, is captured using nighttime light intensity from the National Oceanic and Atmospheric Administration (see Henderson et al., 2012, for details). Light intensity has been identified as a strong proxy for local income within Indonesia over a period of rapid electrification beginning in the late 1980s (Olivia and Gibson, 2013). Changes in light intensity from 1992 to 2010 serve as our main proxy for overall economic growth at the village level. In 1992, only 6.5 percent of Transmigration villages recorded any nighttime lights. By 2010, that figure jumped to 24 percent. By comparison, 12 percent of the 12.5 km^2 geographic grids across sub-Saharan Africa studied in (Michalopoulos and Papaioannou, 2013) have no recorded nighttime lights in 2007 and 2008. As in other studies, we account for the prevalence of zeros by using two measures of the 18 year growth in lights: (i) $\Delta \ln(c + light\ intensity_t)$ for some small constant c where $light\ intensity_t \in [0, 57.6]$,[18] and (ii) the change in the fraction of the village covered by any lights.

In summary, we have six main data sources. These include maps to capture nighttime light intensity, agroclimatic attributes (HWSD), linguistic homelands (WLMS), as well as the 2000 Population Census, the 2002 Village Census *Podes*, and the 1998 Transmigration Census. We also merged in five auxiliary datasets, including the FAO-GAEZ data to measure potential agricultural yields by crop, the 2004 *Susenas* which includes household-level rice productivity (but no migration data), the 1980 Population Census (to calculate pre-1979 variables), the 1985 inter-censal survey (to calculate district-to-district migration flows in the 1980s) as well as planning maps known as Reppprot (Regional Physical Planning Program for Transmigration) published in the 1980s to identify planned but untreated settlements (discussed later). We provide further details on the data sources in Appendix A.

[18] The raw light intensity measures are top coded at 63, but we do not face this censoring issue in our rural context.

5 Theoretical and Empirical Framework

This section lays out our conceptual framework. We first explain how agrolimatic similarity proxies for skill transferability across locations and serves as a measurable source of comparative advantage that shapes the spatial distribution of aggregate productivity. We then derive our key estimating equation and discuss identification.

5.1 Theoretical Framework

Following Dahl (2002), we adapt the classic Roy (1951) model with two sectors to a multi-location choice model where heterogeneous farmers sort across heterogeneous locations. There is a discrete set of J locations, indexed by $j = 1, ..., J$. The farming methods (production functions) are different in each location. Locations are differentiated by a bundle of characteristics which we denote using a fixed $(G \times 1)$ vector, $\mathbf{x}_j$, as discussed in Section 3. Individual farmers, indexed by i, are born into a birth location, $b(i) \in \{1, ..., J\}$.

Farmers acquire farming skills that are specific to local growing conditions at their birth locations, captured by $\mathbf{x}_{b(i)}$. This location-specificity, which captures the notions of "latitude-specific" farming skills (Steckel, 1983) and "location-specific amenities" (Huffman and Feridhanusetyawan, 2007), is consistent with local learning models in development economics that show how heterogeneous growing conditions can hamper the spatial diffusion of farming knowledge (see Foster and Rosenzweig, 2010, for a review). Agroclimatic differences are particularly salient in the case of rice for which location-specific knowledge includes, among others, knowledge of what types of varieties are best suited to local growing conditions.[19] Hereafter, we denote $\mathbf{x}_{b(i)}$ with $\mathbf{x}_i$ to simplify notation.

We assume that farmers can only own one unit of land in their location of choice (where they both live and work), and we normalize the output price to one. For now, we abstract from unobservables to highlight the observable determinants of productivity central to our hypotheses. The value of output per unit of land owned by farmer i in location j is given by:

$$y_{ij} = \gamma \mathcal{A}_{ij} + \mathbf{x}_j' \boldsymbol{\beta}, \tag{3}$$

where $\mathbf{x}_j' \boldsymbol{\beta}$ maps observable agroclimatic characteristics of location j into productivity, $\mathcal{A}_{ij} = (-1) \times \mathrm{d}(\mathbf{x}_i, \mathbf{x}_j)$ is our measure of agroclimatic similarity between locations. If skills were *perfectly* transferable, a migrant's origin does not matter and $\gamma = 0$. Conditional on $\mathbf{x}_j$, a positive γ means $\mathcal{A}_{ij}$ is an important predictor of productivity at the destination, above and beyond the effects of $\mathbf{x}_j$ on output. $\mathcal{A}_{ij}$ reflects complementarity between a farmer's location-specific knowledge and local farm characteristics. For a given destination, farmers migrating from more similar origins are more productive because it is easier to transfer their farming skills, compared to farmers from dissimilar origins.

When we aggregate across individuals, our model sheds light on the role of comparative advantage in shaping the spatial distribution of *aggregate* productivity. Since higher similarity reflects a better match

[19] Van Der Eng (1994, p. 26) notes, for example, that "(t)here was a wide range of natural cross-bred varieties from which farm households could choose... many (farmers) had managed to select varieties that suited their circumstances best through a century-old process of trial and error... The quality of irrigation systems, altitude, soil conditions, planting time, crop rotation schemes, the use of fertilizer, the preferences of local consumers or rice mills, rainfall, and the availability of labor could all differ substantially from one area in Java to another."

quality (or greater complementarity) between migrants' skills and local growing conditions, villages assigned a higher share of migrants from agroclimatically similar origins experienced a "shock" of higher quality-adjusted labor endowments.[20] Such villages therefore have greater comparative advantage at farming than villages assigned a high share of migrants from dissimilar origins.

The ability to measure skill transferability is an important innovation of our research design. We are able to do so because a wealth of agronomic research has identified and collected data on (predetermined) agroclimatic characteristics that are vital to farm output.[21] Moreover, land and local climate characteristics change slowly, so that agroclimatic characteristics measured in the 1970s are still highly predictive of productivity in 2000. All land attributes and ethnolinguistic homelands are predetermined and hence unaffected by settler farming activities and any corresponding inflow of capital or labor. We can therefore abstract from reverse causality concerns.

Having described the determinants of productivity, we are ready to characterize the location choice process. As in Dahl (2002), we assume that productivity and taste differences determine how farmers sort across locations. That is, the indirect utility of farmer i in location j is

$$V_{ij} = y_{ij} + \varepsilon_{ij}, \tag{4}$$

where y_{ij} is as above, and ε_{ij} is her individual-specific taste for living in location j.

In equilibrium, farmers choose locations to maximize V_{ij}, and we denote farmer i's optimal location by $j(i)^*$. In this setting, each farmer i has J potential outcomes for agricultural productivity, which we can write as $y_{i1}, y_{i2}, \ldots, y_{iJ}$. As shown by Heckman and Honore (1990), it is difficult to identify the importance of comparative advantage because when people sort based on their comparative advantage, we do not observe all J potential outcomes for each farmer (we observe $y_{ij(i)^*}$ only). A common solution is to identify instruments that affect location choice but are excluded from the determination of productivity. The problem is that many variables jointly affect location choice (V_{ij}) as well as productivity (y_{ij}).[22] Moreover, in our setting, each individual has J potential locations, instead of a binary choice (such as whether or not to work). Hence, a good instrument would need to have a strong "first stage" for *each* of the J potential locations in addition to satisfying the exclusion restriction.

A stylized two-sector Roy (1951) assignment model helps clarify the identification problem. Suppose there are two types of farms (e.g., *L*owlands and *H*ighlands) and two types of farmers (born in L and H, respectively). There are four potential outcomes: Y_{LL}, Y_{HH}, Y_{LH}, Y_{HL} where Y_{ij} is the agricultural productivity of a farmer born in location i and farming in location j. Here, similarity is high for *LL* and *HH* pairs. If farmers born in lowlands have a comparative advantage at growing rice in lowlands (relative to farmers born in highlands) and vice versa for farmers born in highlands, and if farmers sort

[20] Gibbons et al. (2005) relate comparative advantage with a matching process (as in Jovanovic, 1979) between heterogeneous people and heterogeneous places.

[21] In a recent survey of Roy assignment models where workers sort into tasks based on comparative advantage, Autor (2013) notes that there is a "difficulty of identifying credible counterfactuals" and "no labor market data equivalent to agronomic data are available for estimating counterfactual task productivities."

[22] For example, Dahl (2002) argues that family status affects migration probabilities but not earnings, and Bayer et al. (2011) argue that birth location affects the nonpecuniary components of utility (and hence, location choice) but does not affect productivity. For us, birth location fixed effects are not excludable from the productivity equation because comparative advantage is a function of the proximity between origins and destinations. Moreover, a long line of research in development economics has shown that family ties are a source of informal insurance that jointly determines both income and migration choices (see Morten, 2013; Munshi and Rosenzweig, 2013, for recent examples).

into locations based on comparative advantage, then the econometrician would only observe two of the four outcomes, namely those associated with high similarity only: Y_{LL}, Y_{HH}. In this case of perfect sorting, there is no observed variation in agroclimatic similarity.

The Transmigration program provides quasi-experimental variation in spatial labor allocation, allowing us to observe migrants assigned to both high and low similarity destinations for exogenous reasons. Indeed, as discussed below, we observe more low similarity realizations in Transmigration villages, compared to non-Transmigration villages, where spontaneous migrants are to free to sort. With this in mind, we derive our key estimating equation and discuss threats to identification.

5.2 Empirical Strategy

Our goal is to estimate the elasticity of aggregate productivity with respect to agroclimatic similarity between migrant farmers' origins and destinations. Our key regression is at the village level, but it is instructive to begin at the individual level. We augment the model above to allow for unobservable determinants of productivity:

$$y_{ij} = \underbrace{\gamma \mathcal{A}_{ij} + \mathbf{x}_j' \boldsymbol{\beta}}_{\text{observable}} + \underbrace{\eta_i^u + \mu_j^u + \omega_{ij}}_{\text{unobservable}}, \tag{5}$$

where η_i^u is unobserved individual characteristics, μ_j^u is unobserved natural advantages and ω_{ij} is an idiosyncratic error term.

Our village-level estimating equation aggregates (5) over a non-random set of individuals, I_j, whose optimal location is j:

$$y_j = \gamma \mathcal{A}_j + \mathbf{x}_j' \boldsymbol{\beta} + \underbrace{\sum_{i \in I_j} \eta_i^u + \mu_j^u + \omega_j}_{\text{unobservable}}, \tag{6}$$

where the key regressor, $\mathcal{A}_j$, is aggregated to the village level by averaging $\mathcal{A}_{ij}$ over all Java/Bali migrants living in j (using the π_{ij} migrant weights in equation (1)), $\sum_{i \in I_j} \eta_i^u$ is unobserved demographic composition differences due to endogenous sorting, and ω_j is an idiosyncratic error term.[23]

The key parameter of interest, γ, measures the elasticity of aggregate agricultural productivity for the village with respect to average agroclimatic similarity. The ideal experiment to estimate this population elasticity is to take farmers from many origins and randomly assign them to many destinations.[24] In order to interpret this elasticity as a measure of the importance of skill transferability, it is necessary that the migrants remain farmers and grow the same crop(s) as they did prior to moving. This allows the econometrician to abstract from thorny selection issues due to crop, occupation, and location choices. In practice, numerous constraints make it infeasible to implement this type of randomized experiment.

However, our research design approximates this ideal experiment. The broad spatial scope and exogenous relocation process from the Transmigration program generates uniquely rich and plausibly exogenous variation in origin-by-destination matches of migrants. Moreover, the previously landless transmigrants embarked on the program with the goal of farming, and their newly acquired land would

[23] The fact that y_j is a village-level outcome and our key, plausibly exogenous similarity measure, $\mathcal{A}_j$, is potentially defined over a subset of the total village-level population raises concerns about aggregation bias that we address below.

[24] Having farmers from many origins assigned to many destinations is useful to estimate the average elasticity for the population by spanning the full support of the key regressor, $\mathcal{A}_j$. In the stylized two-by-two example, spanning the full support is akin to observing the full set of counterfactuals.

serve to tie the first generation movers to farming. We focus on rice, which is the ideal crop because it is the primary staple, expanding rice output was a program objective, and most importantly, nearly all farmers in Java/Bali during the Transmigration period were growing rice, as discussed in Section 4. Hence, farming skills were mostly specific to rice. We therefore view other non-rice crops—and, in particular, major cash crops such as palm oil, rubber, and cocoa—as suitable for placebo tests because skill transferability as proxied by agroclimatic similarity should only have productivity effects on crops in which transmigrants have farming experience. We can then characterize ex post adaptation responses and argue that selection via crop, occupation, and ex post migration adjustments cannot fully explain the main effects of agroclimatic similarity on rice productivity.

Our regression conditions on observably identical destination villages and compares villages that have a high share of Java/Bali migrants from similar origins against villages that have a high share of Java/Bali migrants from dissimilar origins. The key sources of plausibly exogenous variation in our village-level index $\mathcal{A}_j$ include: (i) variation in the absolute differences between predetermined agroclimatic characteristics in destinations versus origins, and (ii) variation in the share of Java/Bali migrants in destination village j who are from origin district i, π_{ij}. Our regression identifies the *added* productivity effect of agroclimatic similarity (after conditioning on $\mathbf{x}_j$) and exploits variation in π's from origins with similar versus dissimilar $\mathbf{x}$.

Validating the Identification Strategy. We take several steps to show that the exogenous variation in agroclimatic similarity is due to the Transmigration policy and that our estimate of γ in equation (6) is not driven by spurious correlations with origin and destination attributes.[25]

We first show that the distribution of agroclimatic similarity is different among Transmigration villages compared to other villages in the Outer Islands. Panel A of Figure 3 plots kernel densities of village-level agroclimatic similarity, aggregated over all individuals (migrants and natives). There is a mass at 1 because many natives are stayers ($\mathcal{A}_{ij} = 1$ for stayers). Panel B uses π weights that include migrants only (both Java/Bali migrants and migrants born in other districts in the Outer Islands). These plots show two things. First, absent the policy, individuals appear to sort in a way that increases the agroclimatic similarity between origins and destinations. The distribution for non-Transmigration villages is shifted to the right. Second, there is greater dispersion in realized similarity in Transmigration villages. This is consistent with the discussion in Heckman and Honore (1990) that the Roy model has "no empirical content."[26]

Using a simulation exercise, we compare the actual distribution of agroclimatic similarity across villages with the distribution that would have resulted from purely random assignment. Over 10,000 simulated distributions, we never reject that the means and standard deviations of the random and actual distributions are equal.

Our main identification threat is that high and low similarity villages are not comparable because

[25]Since our key source of variation for agroclimatic similarity is at the origin-by-destination level, ideally, we would include both origin and destination fixed effects. However, we are constrained because our main estimation sample is a single cross-section of 814 villages, and there are 119 origin districts and 70 destination districts.

[26]This is because workers sort into occupations based on comparative advantage, so that the realized distribution of wages is endogenous in two ways. First, it is shifted to the right because workers select the occupation with the highest wage, and hence we do not see the worker's other potential outcomes. Second, the variance in log earnings is lower in a Roy economy relative to an economy where workers are randomly assigned to jobs.

of unobserved differences in demographic compositions (due to endogenous sorting) and differences due to unobserved natural advantages that are correlated with our outcomes. The latter is potentially worrisome because destinations that are agroclimatically similar to Java/Bali may have unobservable natural advantages because Java/Bali is known to be naturally advantaged for rice production.

We first show that pre-program determinants of productivity are not correlated with agroclimatic similarity in Table 3. The table reports estimates from separate regressions of agroclimatic similarity on island fixed effects, natural advantages $\mathbf{x}_j$, and each of 24 variables capturing potential agricultural productivity, as well as district population size, quality of housing and utilities, schooling, literacy, language skills, and sector of work for those living in villages near the Transmigration settlement in 1978.[27] It is important to note that these Transmigration villages are *new* settlements, and hence there are no pre-1979 outcome measures for the given village j.

Importantly, agroclimatic similarity is not correlated with potential crop yields obtained from auxiliary agronomic data. This rules out first-order concerns about unobserved natural advantages. Related to this, we show later that our estimates of γ are relatively stable when we add or drop observed natural advantage controls at the origins and destinations, implying that agroclimatic similarity is not merely proxying for agroclimatic quality. The subsequent rows in the table show that agroclimatic similarity is also uncorrelated with other pre-determined measures of development. Only one variable is significant at the 5 percent level, and the difference is negative, which works against our findings. We also find that agroclimatic and linguistic similarity are uncorrelated ($\rho < 0.04$), which is consistent with the plausibly exogenous assignment of heterogeneous people to heterogeneous places.

We tackle a related concern that agroclimatically similar destinations are initially assigned or subsequently attract different settlers along unobserved dimensions that are correlated with productivity. Although transmigrants are (weakly) negatively selected on average, agroclimatic similarity does not have an economically or statistically significant relationship with pre-program schooling acquired by eligible individuals born in Java/Bali. This can be seen most clearly in Figure 4, which plots the nonparametric densities of individual-level agroclimatic similarity for all Java/Bali-born migrants in Transmigration villages by schooling. The distributions are effectively indistinguishable across schooling levels.[28]

Using a quasi-gravity specification, we show that transmigrants do not appear to have endogenously sorted in (out) of those sites in which they (do not) have observable comparative advantage. The results help rule out concerns that farmers are sorting based on unobservable sources of comparative advantage that are spuriously positively correlated with similarity. In particular, we examine whether the stock of Java/Bali migrants from origin district i residing in Transmigration village j in 2000 is increasing in agroclimatic similarity ($\mathcal{A}_{ij}$) between i and j. In Table 4, we use OLS to estimate variants of the following equation

$$f(migrants_{ij}) = \alpha + \lambda_a \mathcal{A}_{ij} - \lambda_d \ln distance_{ij} + \mathbf{z}_j'\boldsymbol{\zeta} + \tau_i + \upsilon_{ij}, \tag{7}$$

[27] We use the FAO's GAEZ data (see footnote 12) to construct potential yield measures. We use the 1980 Population Census to construct measures of economic activity and well-being across Outer Island districts before the major onset of the program. In particular, we estimate district-level characteristics using the population that had been living in each district prior to 1979 when the transmigrant influx began. This ensures the exclusion of all potential transmigrants and the population of non-transmigrant immigrants that may have arrived in response to the program.

[28] Taking a more parametric approach, Appendix Table B.2 shows that village-level agroclimatic similarity is uncorrelated with schooling among Java/Bali immigrants after conditioning on the natural advantage controls x_j and other predetermined demographic characteristics of both Java/Bali-born and other village residents.

where $\mathbf{z}_j$ includes island fixed effects, the year of initial settlement, and the log number of individuals placed in j. Columns 2 and 4 additionally include all of the predetermined variables in Table 3. We estimate equations for the extensive margin, $f(migrants_{ij}) := Pr(migrants_{ij} > 0)$, and intensive margin, $f(migrants_{ij}) := \ln(migrants_{ij})$, of migration flows. In all cases, we define transmigrants in j as individuals born in Java/Bali and two-way cluster standard errors (Cameron et al., 2011) by i and j (with similar results using less conservative clustering approaches).

In all specifications of equation (7), we cannot reject the null hypothesis that $\lambda_a = 0$. Moreover, the estimated λ_a are very small relative to the mean of the given dependent variables. This provides strong suggestive evidence that 12 to 20 years after the initial wave of resettlement, migrants from Java/Bali did not endogenously sort in (out) of more (dis)similar sites. Migrant stocks tend to be somewhat higher in physically closer sites ($-\lambda_d > 0$, perhaps due to transport costs, see Section 2.1). However, agroclimatic "distance" does not exhibit the same hypothesized gravity forces along either the extensive or intensive margin. We find similar precise zeros for linguistic similarity when estimating an augmented equation (7) at the $ij\ell$ level for the ethnolinguistic groups ℓ indigenous to Java/Bali.

We also address concerns that improved planning between *Repelita*s III and IV led to the establishment of more agroclimatically similar sites towards the end of our study period. In ordered logit regressions for the year of settlement $t \in \{1979, \ldots, 1988\}$, agroclimatic similarity ($\mathcal{A}_j$) has a negligible, statistically insignificant relationship with t, conditional on $\mathbf{x}_j$. Moreover, agroclimatic similarity is uncorrelated with the number of individuals placed in t.

Overall, the evidence suggests that agroclimatic similarity is quasi-randomly distributed across Transmigration villages, even as observed up to two decades after resettlement. Endogenous sorting and selection on comparative advantage matter in most migration contexts. However, the near-random initial assignment of transmigrants across sites had persistent effects on location choices over subsequent decades. We now exploit this persistence to identify the medium-run effects of agroclimatic similarity on productivity.

6 Empirical Results

In this section, we present our main empirical results. We begin by reporting large average elasticities of skill transferability for rice productivity. We relate this to recent work on location-specificity of crop and staple consumption preferences of migrant farmers. We turn next to heterogeneity analysis to identify where similarity matters the most. We explore possible mechanisms of adaptation, including learning, switching crops, occupation and migration. Next, we report smaller effects for broader economic development outcomes, suggesting barriers to adaptation are most significant for rice farming. Finally, as a policy exercise, we estimate the average treatment effects of the program using planned but unsettled villages as controls. Ultimately, we argue that the persistent effects of similarity may explain the limited average impact of the program on local economic development in the Transmigration villages. We show this using simulations that reallocate migrants on the basis of agroclimatic similarity.

6.1 Average Effects on Rice Productivity

Table 5 reports our main estimates of γ, the coefficients on agroclimatic similarity in the following regression based on equation (6):

$$y_j = \alpha + \gamma_a \mathcal{A}_j + \mathbf{x}_j' \boldsymbol{\beta} + \nu_j, \tag{8}$$

where village-level agroclimatic similarity ($\mathcal{A}_j$) is based on the Java/Bali migrant weights; $\mathbf{x}_j$ includes island fixed effects, the full set of predetermined agroclimatic endowments elaborated in Section 3;[29] and ν_j is a composite error term capturing all unobservables in equation (6). As a baseline, we cluster standard errors using the Conley (1999) GMM approach allowing for arbitrary correlation in unobservables across all villages within 150 kilometers of village j.[30] Column 1 reports our preferred specification and is the basis of the foregoing analysis. In all regressions, we rescale the independent variables so that we can read a one standard deviation impact directly from the tables.

Our first result implies that a one standard deviation (0.14) increase in the agroclimatic similarity index leads to a 20 percent increase in rice productivity (column 1). This suggests agroclimatic similarity is an important predictor of cross-sectional differences in aggregate rice productivity, translating into a level effect of an additional 0.5 tons per hectare for the average village (with 2.5 tons per hectare, see Table 2). This productivity effect is large, equivalent to twice the productivity gap between having no education and junior secondary.[31] The magnitude is plausible, especially since our village-level productivity measure aggregates across multiple cropping seasons, and rice farmers in Indonesia report up to three harvest cycles per year. These productivity effects for rice are important because rice is, by far, the most important crop and staple in Indonesia, and expanding rice production was one of the program's main goals.

These quasi-experimental estimates of productivity losses due to agroclimatic dissimilarity complement other work on the effects of location-specificity in migrant farmers' crop and staple consumption choices. Michalopoulos (2012) finds that ethnic groups living outside their indigenous homeland tend to grow staple crops more similar to those grown in the homeland relative to those grown in their non-coethnic region. This could be driven by preferences to grow staples for consumption, which is consistent with recent work on the geographic variation in staple consumption preferences. Atkin (2013) estimates sizable caloric losses incurred by migrants in India who move to places with different "food cultures."

This key result is robust to several important concerns about identification. Column 2 shows that the results are stable when we drop natural advantage controls. This addresses the concern that agroclimatic similarity is only picking up spurious correlation with unobserved natural advantage. If this is the case, then the coefficient should change if we drop observed natural advantage controls, but the effect is stable. Column 3 adds controls at the origins: four province-level aggregates of the (119) origin district i-specific π_{ij} terms used to construct $\mathcal{A}_j$, a π_{ij}–weighted average of distance to the origins, and a π_{ij}–weighted average of predetermined controls at the origins including potential yields (i.e., the variables reported in Table 3). This addresses concerns that the π_{ij} used to construct the agroclimatic similarity index is

[29] The controls also include the log of the great circle distance to the closest point in Java/Bali, log total land area, log distance to the subdistrict and district capital, and log distance to the nearest pre-1979 major road.

[30] Inference is robust to varying the bandwidth from 50 to 500 kilometers or clustering by administrative district boundaries.

[31] This figure is based on household-level data on rice productivity and education of the household head from 2004 *Susenas*.

correlated with unobserved determinants of productivity at the origins. Column 4 adds predetermined controls at the destinations (from Table 3) as well as controls for demographic composition, including the gender, age, and schooling shares of Java/Bali- and Outer Islands-born residents in each village. Column 5 is our most saturated regression that includes origin and destination controls. Given that we only have 600 villages, this is a very demanding specification that includes 87 control variables. The effects change slightly but are not statistically significantly different from column 1.

Still, endogeneity concerns remain because not all villages and not all individuals produce rice. To deal with the former, we run OLS and Tobit regressions with rice productivity in levels instead of logs—villages that do not produce rice have zero productivity—and find similarly large productivity effects. The latter concern is that the lower rice productivity in low similarity villages is driven by selection of unobservably different individuals out of rice farming (because relatively more high ability farmers have selected out of rice farming in low similarity villages). Using a simple selection model, we can show that the degree of selection needed to explain the productivity effects is too large, when compared to the estimated effects of agroclimatic similarity on crop choices. We also address aggregation bias by using *Susenas* data for a small sample of Transmigration villages that includes household-level rice productivity measures and find similar results (see Appendix Table B.3).

We provide additional evidence of robustness in Appendix Table B.4. Each row introduces a single change to the baseline specification, which is reproduced in row 1 for reference. We address concerns related to confounding program features by controlling for the scale, the timing of the initial transmigrant influx, or destination province $\times$ year of settlement fixed effects. We also consider alternative definitions of our agroclimatic similarity index based on different distance metrics and migrant weights. We further address aggregation bias by controlling for the share of natives and overall population density. None of these changes affects our key finding of a statistically and economically significant effect of agroclimatic similarity on rice productivity.

Heterogeneity: Where Does Similarity Matter Most? Columns 6 and 7 of Table 5 show that agroclimatic similarity is more important in places with adverse growing conditions. In column 6, we interact agroclimatic similarity with the FAO-GAEZ measure of potential rice productivity.[32] The negative and significant coefficient on the interaction term implies agroclimatic similarity is less important in villages with high potential productivity. The magnitude suggests the productivity losses of one standard deviation of dissimilarity can be offset by an increase of 1.34 tons/ha in potential productivity of the land.

Column 7 interacts agroclimatic similarity with indicators for three groups of Transmigration villages, with low, medium and high shares of wetland (the most prevalent type of land used to farm rice in Java/Bali). This reduces an otherwise high-dimensional vector of agroclimatic attributes into a single land quality measure that is particularly informative about variation in cultivation methods (and underlying productivity). The coefficients are largest in villages with mostly dryland and decrease monotonically as the share of wetland increases. One potential explanation is that farmers from Java/Bali accustomed to wetland agriculture found it difficult to adapt to the dryland approaches in the settlement area.[33] Second, adaptation to wetland production is relatively easy even for farmers accustomed

[32] We take a weighted average of potential dryland and wetland yields with weights based on the share of farmland that is wetland. Results are similar using a simple average. We also control for potential productivity separately.

[33] Donner (1987) succinctly captures this possibility: "The Javanese transmigrants, mostly experienced in growing wet-rice,

to dryland methods in Java/Bali. In this context, agroclimatic differences can be easily overcome given the strong natural advantages of wetland production systems. This is in line with results in Appendix Table B.5 where we decompose the agroclimatic similarity index into three sub-indices of topographic, soil, and climatic attribute similarity (see Section 3). We find that topographic similarity is a key determinant of rice productivity.

Next, we show that the large positive average effects of agroclimatic similarity are driven by villages in the lower tail of the similarity distribution. In particular, we estimate a semiparametric version of equation (8):

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

where $g(\cdot)$ is a partially linear function that relates agroclimatic similarity to the outcome y_j using the approach in Robinson (1988). Figure 5 shows the shape of $g(\cdot)$ for our key rice productivity outcome.

The semiparametric estimate reveals nonlinear effects that are consistent with a concave adjustment process where adjustments are increasingly costly the greater the agroclimatic distance to the origins. The steepest effect size is found in the bottom tercile of the index ($\mathcal{A}_j \leq 0.55$) after which the effects of similarity kink and then level off with little difference in productivity between villages at the median $\mathcal{A}_j = 0.63$ and the 95th percentile $\mathcal{A}_j = 0.75$. For these villages in the bottom tercile, a back-of-the-envelope calculation suggests the calories implied by their low annual rice output is right around the subsistence threshold. This is consistent with findings from Bryan et al. (forthcoming) that subsistence farmers may underinvest in adaptation because losses from risky experimentation (e.g., with new technologies) is particularly costly near subsistence.

There are two ways policymakers could have perhaps mitigated the productivity losses seen in the lower tail of Figure 5. First, more careful matching of transmigrants' skills to destination growing conditions may have pushed all villages into the portion of the figure where agroclimatic similarity has relatively less impact. The semiparametric results suggest that it is most important to avoid very bad matches rather than achieving the best match. Second, even in the absence of better matching, greater investments (targeted to low similarity villages) in agricultural extension, retraining programs, and complementary capital inputs may have facilitated greater adaptation and ultimately limited the persistent effects of initial match quality. We revisit the policy question in Section 6.3.

Summary. In summary, we show that agroclimatic similarity has important productivity effects on rice farming, on average. Further heterogeneity analysis show that the effects are mostly concentrated in the bottom tercile of agroclimatic similarity (consistent with a concave adjustment process) and appear to be more important in places with adverse growing conditions (especially drylands or places with low potential productivity for rice). Next, we explore several adaptation mechanisms that might mitigate losses due to dissimilarity.

6.2 Adaptation

We investigate four adaptation mechanisms, including learning, switching occupations, crop choice, and ex post migration. We find relatively more support for learning and crop adjustments, especially

were promised irrigated land in the new settlements, but found only dryland and had to change to rain-fed cultivation."

switching to cash crops. Finally, while we find some adaptation, we show that there remain sizable differentials between high and low similarity villages in light intensity growth, a proxy for local income growth used in several recent studies.

Learning. There is an extensive literature on the importance of learning in the agricultural context (see Foster and Rosenzweig, 2010, for a review). This rich literature provides evidence of several types of learning models, including social learning (from peers) and local learning under heterogeneous growing conditions. Our results on productivity effects due to agroclimatic similarity are consistent with models of local learning under heterogeneous growing conditions. For example, Munshi (2004) documents stronger evidence of learning from neighbors in the case of wheat relative to rice because rice varieties are more sensitive to local growing conditions. Hence, information on production methods extracted from neighboring regions' rice varieties is less useful if growing conditions are heterogeneous. Similarly, Ben Yishay and Mobarak (2014) find that farmers are most persuaded by information provided by other farmers who face comparable agricultural conditions.

We also find evidence consistent with social learning from natives. Column 8 of Table 5 shows that a one standard deviation increase in linguistic similarity increases rice productivity by 13 percent. As discussed in Section 3, our linguistic similarity index in equation (2) measures the structural proximity between languages native to Java/Bali and languages native to the Outer Islands.[34] This effect is robust to controlling for all the pre-determined variables in Table 3 as well as demographic compositions (the specification in column 4).

Column 9 shows that linguistic similarity is more important in places with a greater scope for learning from natives. We split the villages into two groups, based on whether they were assigned above- or below-median number of transmigrants, and separately re-estimate equation (8). We assume a larger scope for learning from natives in places with a smaller transmigrant stock in the initial year.[35] We find that linguistic similarity is indeed more important in villages with a smaller number of initial transmigrants. Overall, these results echo case studies of Transmigration settlements that discuss the importance of learning from natives.[36]

Occupational Choice. Another way in which transmigrants may have dealt with dissimilarity is by switching occupations (i.e., out of farming). Consider a simple Roy model with two skills, agricultural

[34] Unfortunately, there was insufficient local variation in the data to examine whether linguistic similarity could mitigate the adverse effects of agroclimatic dissimilarity using an interaction term, $\mathcal{A}_j \times \mathcal{L}_j$. In attempts to do so, we are unable to rule out large positive or large negative effects.

[35] Although we do not observe the initial native population size, the size of the initial transmigrant population is a good proxy for relative group sizes. Given that program planners accounted for the surrounding native population size when they calculated the carrying capacity, conditional on agroclimatic endowments $\mathbf{x}_j$, a large (small) initial transmigrant population is indicative of a small (large) initial native population. It is also important to recall that agroclimatic similarity is uncorrelated with the size of the initial transmigrant population, conditional on $\mathbf{x}_j$.

[36] As Donner (1987) notes: "The majority of land allocated to new settlers will, however, consist of rain-fed upland. Only a few of the transmigrants are familiar with the cultivation of such land since they are either sawah [wetland] farmers used to irrigation facilities, or dryland farmers of comparatively fertile volcanic soils." In surveying a few settlement areas, Kebschull notes that "It seems that most of them also could overcome the difficulties resulting from the lack of agricultural extension. They gained their own experience and tried to learn as much as possible from their successful neighbors and friends." Donner (1987) continues, "A further problem arising from the relations between the original, autochthonous population and the new transmigrants was that the former did not learn a more advanced agriculture from the latter as had been hoped; rather, it was often the case that the new settlers had to learn from the original inhabitants how to produce under local conditions."

and language, and two occupations, farming and trading/services. Farming is relatively more intensive in agricultural skills while trading/services is relatively more intensive in language skills (given the need to communicate with non-coethnics in the local marketplace). The theory of comparative advantage predicts that individuals assigned to agroclimatically similar villages are more likely to remain as farmers (as they were in Java/Bali) and those assigned to linguistically similar villages are more likely to switch into trading and services.

We test these predictions in Table 6 using the universe of individual-level Population Census data for Transmigration villages. We model binary occupational choices as a linear probability function of individual-level demographic controls, village-level controls, year of settlement fixed effects, and individual agroclimatic ($\mathcal{A}_{ij}$) and linguistic similarity, which is the term after $\pi_{\ell j}$ in equation (2). The flexible set of individual- and village-level controls ensures that we are comparing the effects of agroclimatic and linguistic similarity on occupational choices across otherwise observably identical individuals in observably identical villages.[37] Columns 1-3 in Panel A report estimates for the probability of being a farmer working in either food or cash crop production, while columns 4-6 report the probability of being involved in trading or services. The sample in columns 1 and 4 include the Java/Bali-born population between the working ages of 15 to 65. Columns 2 and 5 (3 and 6) restrict to young (old) individuals who were less (older) than 10 years old in the year of initial settlement.

We find some adjustment in occupation choices, consistent with the theory of comparative advantage, but the magnitudes are small. Agroclimatic similarity increases the likelihood of farming and decreases the likelihood of trading. A one standard deviation increase in individual agroclimatic similarity leads to a 0.9 percentage point (pp) higher probability of an individual reporting farming as their primary occupation. Meanwhile, a one standard deviation increase in linguistic similarity is associated 1.7 pp higher probability of trading/services. However, the effects are quantitatively small. For example, the 0.9% agroclimatic similarity effect in column 1 implies that only 5,100 more individuals chose farming (relative to 350,000 individuals in the sample who are farmers). The effects of linguistic similarity are relatively larger but still limited. Comparing across columns, we find no significant differences in the patterns of occupational choices across the young and old generation of transmigrants. This points to intergenerational persistence in occupational choices (not unlike Abramitzky et al., 2014).

Crop Choice. Although many low similarity transmigrants remained farmers, crop switching may have been another potentially important margin of adjustment. We examine this possibility in Table 7, where we retain the baseline specification in column 1 of Table 5 and consider the effects of agroclimatic similarity on total agricultural productivity for cash crops (column 1), food crops (column 2) and all crops (column 3).

Interestingly, we find that agroclimatic similarity has significant productivity effects on food crops (column 2) but precise zero effects on productivity for cash crops (column 1). The 95 percent confidence interval for the effect of a one standard deviation change in agroclimatic similarity ranges from -0.04 tons per hectare to 0.09 tons per hectare (a narrow range relative to a mean of one ton per hectare and a stan-

[37] The individual-level controls include gender, married, years of schooling, residence five years ago (Java/Bali, other Outer Islands province or district) and an indicator for belonging to a native Java/Bali ethnic group, and eight indicators for religious affiliation. All but the last are interacted with age. The village-level controls are the same as those used in column 1 of Table 5.

dard deviation of 2.7). Moreover, repeating the semiparametric estimation for cash crop productivity, we find a flat line, instead of the concave shape for rice productivity (see Appendix Figure B.2).

The null effects for cash crops serve as a placebo test, lending further support to the view that the large rice productivity effects reflect transferability of skills acquired in Java and Bali.[38] This finding is plausible given that several of the cash crops were not grown in Java and Bali during the 1970s and 1980s, as discussed in Section 4. Hence, our proxy for skill transferability should has no effect on cash crop productivity given that transmigrants had not acquired these crop-specific skills prior to moving.

When we average across all crops using revenue weights, we find that agroclimatic similarity has null effects on total agricultural productivity (column 3). This is not surprising given the higher revenue weights for cash crops (60%) compared to food crops (40%). We also repeat the same regression using total revenues for all crops and find that agroclimatic similarity has a null effect, implying a one standard deviation increase in similarity only increases annual revenue per farm household by around 10,000 *Rupiah* (or roughly 1 USD).

As an accounting exercise, we decompose the effect on total productivity into the effect on rice productivity and the top non-rice crop in revenue terms. Our calculations suggest that the productivity effect of agroclimatic similarity on the top non-rice crop offsets the productivity effect on rice. We estimate that a one standard deviation increase in agroclimatic similarity translates to (i) a 3.7 percent (s.e. of 2.4 percent) increase in the revenue weight for rice and a 20 percent increase in rice productivity, but (ii) a 5.2 percent (s.e. of 1.9 percent) decrease in the revenue weight for the top non-rice crop and a statistically insignificant 5.8 percent decrease in its productivity (s.e. of 5.3 percent). For the average village, this suggests the rice effects led to a 7 percent increase in total productivity, but this is offset by a 7 percent decrease in total productivity due to the effect on the top non-rice crop.[39]

Although instructive, this accounting exercise should be interpreted with caution. The crop-specific weights are based on national prices, which are averages across regions that vary in the market prices fetched by farmers. For rice in particular, national prices may understate the overall importance of the crop in local agricultural income in the Outer Islands.[40]

This decomposition exercise suggests that crop adjustment is an important adaptation mechanism for low similarity villages. We examine this further by studying crop choices. In Table 8, we examine the

[38] Incidentally, the null effects for cash crop productivity also suggest that agroclimatic similarity is not driven by unobserved land quality only. If this was the case, then agroclimatic similarity should lead to productivity effects for all crops.

[39] Nearly all top revenue-generating crops besides rice are cash crops, and we focus on rice and the (next) top crop because together they comprise 82 percent of total agricultural revenue on average across villages. Rice is among the top five revenue-generating crops in 69% of Transmigration villages and 94% of the 600 villages growing rice. The decomposition follows from the product rule. Total productivity is calculated as $\omega_R \ln y_R + \omega_{NR} \ln y_{NR} + \mathcal{O}$ where ω_R and $\ln y_R$ are the revenue weights and log productivity for rice, ω_{NR} and $\ln y_{NR}$ are the analogues for the top non-rice crop in the village, and $\mathcal{O}$ is the weighted average for all other crops. The effect of a one standard deviation increase in agroclimatic similarity ($\mathcal{A}$) on average productivity is then the sum of the effects for each crop:

$$\frac{\mathrm{d}\omega_R}{\mathrm{d}\mathcal{A}} \ln y_R + \omega_R \frac{\mathrm{d} \ln y_R}{\mathrm{d}\mathcal{A}} + \frac{\mathrm{d}\omega_{NR}}{\mathrm{d}\mathcal{A}} \ln y_{NR} + \omega_{NR} \frac{\mathrm{d} \ln y_{NR}}{\mathrm{d}\mathcal{A}} + \frac{\mathrm{d}\mathcal{O}}{\mathrm{d}\mathcal{A}}.$$

The offsetting 7 percent figures are derived by estimating each term using the baseline specification in Table 5 and evaluating this equation using revenue weights (27 percent for rice and 55 percent for the top non-rice crop) and log productivity (0.43 for rice and 0.73 for the top non-rice crop) for the average village.

[40] Using the most comprehensive local producer price data (from *Susenas* 2004), we find that farmgate rice prices in early 2003 in the Outer Islands range from Rp 900/kg at the 10th percentile to Rp 1,500/kg at the 90th percentile. By comparison, the 10–90 range in Java/Bali is Rp 900–1,300. Average farmgate prices were around Rp 100/kg higher in the Outer Islands during this time period. See Bazzi (2014) for further evidence of spatial variation in rice prices across Indonesia.

extent to which the Java/Bali migrants bring their preferences for growing rice with them to Transmigration sites. Following Michalopoulos (2012), we estimate the following regression for Transmigration villages,

$$\frac{rice_j}{staples_j} = \alpha + \rho_1 \left(\frac{rice_{-j}}{staples_{-j}} \right) + \rho_2 \left(\frac{rice_{j(i)}}{staples_{j(i)}} \right) + \mathbf{x}_j' \boldsymbol{\phi} + \nu_j,$$

where $rice_j/staples_j$ is the fraction of rice paddy in total staples (rice, maize, cassava) planted in 2001; $rice_{-j}/staples_{-j}$ is the corresponding measure in neighboring villages (measured as the average share in the district, excluding Transmigration villages); and $rice_{j(i)}/staples_{j(i)}$ is the corresponding measure for Java/Bali-born migrants' origin districts weighted by the usual π_{ij} term capturing the share of migrants from different origins represented in j. After conditioning on the usual $\mathbf{x}_j$ vector, ρ_1 captures the correlation in cropping patterns across nearby villages subject to the same unobservable ecological constraints (as reflected in the cropland allocation of longstanding native farming communities in surrounding villages), and ρ_2 captures the persistence of migrants' growing preferences beyond these constraints. If $\rho_2 = 0$, then transmigrants fully adapted their cropping patterns to such constraints.

Although $\rho_1 > 0$ across all specifications in Table 8, columns 2 and 4 show that origin region cropping patterns explain about 15-20 percent of the patterns accounted for by spatial autocorrelation across nearby villages. Consistent with Michalopoulos (2012), these results indicate that Java/Bali migrants appear to have preferences for growing (and consuming) rice and replicating the basket of goods grown in their origin regions. While the estimates are not directly comparable, the relative magnitudes of ρ_1 and ρ_2 are larger in our context, with relatively less weight on origin cropping patterns and more weight on destination patterns, suggesting some crop adjustments by individual farmers. In particular, Panel B of Table 6 shows that a one standard deviation increase in individual agroclimatic similarity leads to a 1.5 percentage point higher likelihood of young Java/Bali migrants farming cash crops, but we do not find this effect for older Java/Bali migrants. This is a meaningful effect relative to the mean (17.5 percent) but still limited in magnitude, implying 2,600 more cash crop farmers for a one SD differential.

Appendix Table B.6 provides additional evidence on crop selection. Maintaining the baseline specification in equation (8), we find that greater agroclimatic similarity stimulated entry into rice production. A one standard deviation increase in agroclimatic similarity increases the likelihood that farmers in the village grow any rice by 8.1 percentage points relative to a mean of 74 percent. Hence, some of the productivity effects for rice farming that we observe in Table 5 are due to an increase in the extensive margin of rice production in settlement areas. However, agroclimatic similarity also induced some entry into producing cash crops such as coffee, which is grown in Java/Bali (in 12 percent of villages in 2001). This is why we view crop choice at the village-level as a measure of adaptation and use the revenue-weighted agricultural productivity measure in Table 7 to compare productivity outcomes across villages with potentially different patterns of comparative advantage across crops.

(Non-)Selective Migration Patterns. Another way in which farmers may adapt to initial low quality matches is by moving out of the village and perhaps returning to Java and Bali. While bias from return migration has been shown to be important in the literature (e.g., Abramitzky et al., 2014), we argue that this margin of adjustment is less important in our context.[41] First, transmigrants are not as mobile as

[41] If return migration was greater in dissimilar villages and return migrants were more unproductive (which is why they re-

the typical "spontaneous" migrants. Transmigrants volunteered to a program that would assign them to an unfamiliar place because they were unable to migrate on their own due to credit, information, or other constraints. Second, these transmigrants, previously (mostly) landless agricultural laborers, were given land, and the eventual property rights may play a role in tying them to the Transmigration villages. Those property rights and the strong government oversight in these villages may have acted as a deterrent to subsequent large scale in-migration to the settlements given constraints on land availability, which informed the planners' initial determination of the size of the transmigrant cohort in each location. Finally, aggregate statistics from a 1984 Income Survey of transmigrants show that 71 percent (11 percent) report higher (equal) income compared to income levels at their origins.

We confirm that selective out-migration is indeed low. First, the 1998 Transmigration census reports the number of individuals initially placed as well as the population size when a Transmigration village was deemed independent enough that it no longer required official supervision (typically within 5-10 years of placement). We regress the log ratio of these two population sizes on agroclimatic similarity and find small, statistically insignificant effects (with or without the origin π weights). If there were selective out-migration from dissimilar villages, these coefficients would be positive and significant. Second, we use the 1985 inter-censal survey (*Supas*) to estimate an upper bound on the number of return migrants to Java/Bali. Our calculations suggest there were 1,800,264 individuals in Java/Bali in 1985 who reported being in a different district in 1980. Of these migrants, only 2 percent (36,000) reported living in an Outer Island district that had transmigration villages. This upper bound indicates a low rate of short-term return migration, compared to the roughly 1,298,000 transmigrants who were moved in this period. Finally, recall that the gravity results in Table 4 showed that longer-term *ex post* sorting patterns are not positively correlated with agroclimatic similarity. We can also show that agroclimatic similarity is uncorrelated with population size and the Java/Bali-born migrant share in Transmigration villages in 2000 (results available upon request).

Light Intensity as a Proxy for Economic Growth. The preceding discussion shows that learning and crop adjustments appear to be the more important adaptation mechanisms, but there is less evidence of occupational adjustments and ex-post migration. Having identified strong effects of agroclimatic similarity on rice productivity and null effects on total agricultural productivity, it is important to ask whether these adaptation responses can undo the effects of dissimilarity over time. We use the best available measure of local income growth to investigate whether agroclimatic similarity has persistent effects on overall economic development. Our main proxy is growth in nighttime light intensity between 1992 and 2010.

We find statistically and economically significant positive effects of agroclimatic similarity on growth in village-level income as captured by light intensity. This supports the conclusion that skill transferability has persistent effects on economic development 2–3 decades after the initial settlement. Although transmigrant farmers adapted in several ways to low quality matches, the results in columns 4 and 5 of Table 7 imply that such adaptation was relatively incomplete. Column 4 shows that a one standard deviation increase in agroclimatic similarity raises light intensity growth by 5.4 percent relative to a mean

turned), the correlation between the probability of no return migration (so that we observe them in our data) and similarity would be positive and the correlation between no return migration and productivity would be positive, so the bias is positive (the true effect is weaker).

of around 16 percent.[42] Using an approach similar to Henderson et al. (2012), Olivia and Gibson (2013) estimate that a one percent increase in annual light intensity growth is associated with a 0.5 percent increase in district-level gross GDP. Assuming that this elasticity holds at lower levels of administration, this implies that one SD increase in agroclimatic similarity increases village-level income by around 0.15 percentage points annually over the 18 year period.

In column 5 of Table 7, we show that these results are not driven by nonlinearities stemming from the high prevalence of villages with no recorded light coverage. A one SD increase in agroclimatic similarity leads to a 2.2 percentage point increase in the share of the village that has any nighttime light coverage relative to a base of a 5.4 percent increase.[43] This is important given that only 6 percent of Transmigration villages had any light coverage in 1992. Hence, villages with a larger share of migrants from similar agroclimatic origins exhibit greater long-run growth in lights along both an extensive margin of any light activity and an intensive margin of light intensity. We next ask whether this variation in economic performance across Transmigration settlements has implications for the average impact of the program.

6.3 Average Treatment Effects of the Transmigration Program

We use a place-based evaluation approach that exploits a policy discontinuity to provide the first causal estimates of the average impact of the Transmigration program on local economic development. The ATE estimates are important because the Transmigration program is a large-scale policy that affected two millions migrants and cost 6.6 billion US dollars. Moreover, resettlement policies are also growing in importance, as discussed in the Introduction. Even while various governments have begun to plan for and implement resettlements in response to climate change, there remains limited causal evidence of the medium to long run impacts of these resettlement programs, especially in developing countries (IPCC, 2013). We use the ATE results and an optimal migrant assignment exercise to demonstrate the aggregate implications of origin-by-destination match quality for program effectiveness.

Exploiting the Policy Discontinuity. As mentioned in Section 2.1, global oil prices collapsed in the mid-1980s, and declining government revenues forced dramatic cutbacks in the MOT budget, leading to a significant reduction in the number of sponsored households over the coming years.[44] As a result, numerous selected sites never received any transmigrants. We use this set of planned but unsettled villages as counterfactual settlements.

We identify control villages using the MOT's maps of recommended development areas (RDAs) constructed during the site-selection process. There were a total of 969 RDAs identified by the maps, though many were adjacent to one another. We digitally traced these RDAs using GIS software and overlaid the results onto maps of village boundaries in 2000. We define as controls those 907 villages that shared any area with the RDA polygons (see Appendix Figure B.3).

[42] Although only significant at the 10% level, this estimate turns out to be relatively conservative. When introducing many of the robustness checks considered for rice productivity, we find larger and/or more precisely estimated effects. Results available upon request.

[43] This effect size is similar to Michalopoulos and Papaioannou (2013) who show that a one standard deviation increase in an institutional quality index increases the probability that a local area in sub-Saharan Africa has any nighttime lights by around half the average probability.

[44] The budget fell from Rp 578 billion in FY 1985-86 to Rp 325 billion in FY 1986-87. In response, the MOT reduced its FY 86/87 targets for settlement on sites already under preparation from 100,000 to 36,000 sponsored households.

We use these "almost treated" villages as controls in the following equation:

$$y_j = \alpha + \theta T_j + \mathbf{x}_j'\boldsymbol{\beta} + \nu_j, \tag{9}$$

where T_j is a treatment indicator equal to one for Transmigration villages and zero for planned but unsettled RDAs, and $\mathbf{x}_j$ is the usual vector of predetermined controls from equation (8). The key parameter of interest is the ATE, θ, which measures the causal impact of being a Transmigration village.

A key concern with assigning θ a causal interpretation is that there are omitted place variables correlated with treatment assignment that both influenced site selection and outcomes. Spatial policies like the Transmigration program often target underdeveloped or distressed areas, which can lead to downward bias in θ. We rule out first-order concerns with program placement bias by restricting the sample to treated and planned but untreated villages and use a reweighting procedure akin to recent evaluations of place-based policies (Busso et al., 2013; Kline and Moretti, 2014).

ATE: Development. In Table 9, column 1 reports estimates based on comparing Transmigration villages to all other Outer Island villages while columns 2-4 restrict to the set of treated and control villages. Column 2 controls for the predetermined site selection (and agroclimatic) characteristics in $\mathbf{x}_j$. Column 3 implements a double robust approach (Robins et al., 1995) that additionally reweights control villages according to their odds of treatment based on propensity scores estimated using site selection variables. Column 4 employs the Oaxaca-Blinder reweighting estimator developed in Kline (2011). All specifications include island fixed effects. Standard errors are clustered at the district level. Sample sizes vary across outcomes (depending on data availability) and columns but include as many as 31,185 villages in column 1, and 832 treated villages and 668 controls in columns 2-4.[45] As detailed in Appendix B.1, reweighting effectively rebalances the sample as if planners in 1979 randomly chose treated villages among the initial potential settlements.

Panel A reveals the large, long-run demographic change caused by the Transmigration program. Focusing on the preferred Kline reweighting estimator in column 4, treated villages have substantially higher population density (0.74 log points) than almost treated villages. Not accounting for endogenous program placement in column 1 delivers the opposite conclusion. This is intuitive because planners targeted underdeveloped areas—as is common in other place-based programs. This population shock is driven largely by the influx of transmigrants a few decades prior. The Java/Bali-born population increased from a base of 2 percent of the population in control villages to around 37 percent in treated villages. The influx of migrants also caused a large increase in ethnic diversity in the Outer Islands. In the average treated village, nearly 60 percent of individuals identify with ethnicities native to Java/Bali, relative to a base of 6 percent in control villages.

Panel B shows that, on average, the Transmigration program had weak effects on local agricultural development and income growth. First, treated villages exhibit no difference in rice productivity along the intensive margin of tons/ha. The same holds for total output, output/worker, and output/capita

[45] We exclude treated villages on the islands of Maluku and Nusa Tenggara because there is not sufficient within-island variation (see Appendix Figure B.2). While Nusa Tenggara has five treated villages (and 57 controls), we cannot reliably map outcomes and $\mathbf{x}_j$ to these villages. We also exclude control villages that are within 10 km of Transmigration settlements to minimize bias from spillovers.

(column 4). This is not due to differential selection into rice production. Rice is grown in 80 percent of villages, and the program did not lead to any changes between treated and control areas. Nor are the null effects due to low power. Moreover, these null productivity effects are not specific to rice or to measures of agricultural productivity. The same holds for both revenue-weighted average yields across all crops grown in the village as well as light intensity growth. Controlling for predetermined schooling levels of transmigrants and natives in treated and control villages leaves these results unchanged, suggesting that general human capital differentials do not explain the weak program impacts.

Discussion: Optimal Assignment. We argue that our results on the persistent consequences of origin-by-destination mismatch and limited adaptation in the Transmigration villages can partly explain these weak average productivity effects. In particular, complementarities between heterogeneous individuals and heterogeneous places can give rise to persistent spatial productivity gaps. If labor is unable to sort optimally across locations, then the potential gains from labor reallocation may go unrealized. In our setting, if these frictions are strong enough and are not binding in control villages, then the positive productivity gains of land clearing and other non-labor inputs to production in Transmigration villages could have been undone after two decades. In other words, the low quality matches and limited adaptation could have pulled down average productivity in treated villages, leading to the null results we find in Table 9.

To approximate the aggregate output losses that resulted from poor allocations of transmigrants to destinations, we use the rice productivity results of column 1 in Table 5 and attempt to reassign transmigrants to destinations to maximize similarity, and hence, rice output. As we discuss in Appendix C, this assignment problem is a special case of the generalized assignment problem, a problem in combinatorial optimization that has been shown to be NP-hard. However, we can approximate the optimal solution using a greedy assignment algorithm, in which similarity is sequentially maximized, village-by-village; although such a solution may not be a global optimum, it is computationally feasible and represents an approach to the problem that could be carried out by future resettlement planners. Using this greedy assignment algorithm, we find that aggregate rice production could have been 27.5% higher if individuals had been assigned in a more optimal manner.

7 Conclusion

This paper used plausibly exogenous variation from a large-scale rural-to-rural resettlement program in Indonesia to identify the importance of skill transferability in determining the persistent impact of spatial labor reallocation. We show that villages that were assigned a higher share of migrants from agroclimatically similar origins in Java/Bali (migrants with greater comparative advantage) exhibit greater food crop and especially rice productivity compared to villages that were assigned migrants from less similar origins. We then characterize adaptation responses to agroclimatic dissimilarity and find that learning and crop adjustments appear to be the more important adaptation strategies compared to occupation adjustment or ex post migration. We find null effects on total agricultural productivity, but this is perhaps due to imperfectly measured revenue weights assigned to crop-specific productivities. Nevertheless, the results are consistent with skill transferability being most important for crops in which migrants have

prior experience. Moreover, the positive effects of agroclimatic similarity on growth in nighttime light intensity point to incomplete adjustment over the medium-run period in this study.

Our findings shed new light on the importance of comparative advantage in shaping the spatial distribution of productivity. A growing literature argues that labor is spatially misallocated across locations (e.g., Young, 2013), sectors (e.g., Gollin et al., 2014), and occupations (e.g., Hsieh et al., 2013). Our natural experiment suggests that some of this misallocation may be explained by barriers to transferring skills and ultimately adjusting to new economic environments. Quantifying the welfare costs of these barriers is an important task for future research.

Our results also have important implications for the design of future resettlement programs. When comparing Transmigration villages against planned but unsettled villages, we find small average treatment effects on economic outcomes (in spite of large effects on population density). This can be explained in part by the poor matching of migrants' farming skills to local growing conditions. We provide evidence from a simulation exercise suggesting sizable aggregate rice productivity gains from optimally allocating migrants on the basis of agroclimatic similarity. Our results also indicate that complementary government inputs are crucial to help farmers mitigate the effects of dissimilarity.[46] Although negatively selected from their rural cohort at the time, the participants in the Transmigration program are precisely the types of individuals most likely to be adversely affected by climate change and hence for whom such policy choices are most crucial.

Finally, our paper focuses on economic outcomes only. In future work, it would be interesting to study the effects of the program and agroclimatic and linguistic similarity on social outcomes including language diffusion, interethnic marriage, and political preferences. Nation-building was an important non-economic goal of the program which our quasi-experimental design is well suited to evaluate.

[46] As noted by Donner (1987), "... even sensitive soils can be used economically viabl(y) if the proper techniques are applied. In many cases however, such techniques are either unknown to the transmigrants or require too high an investment to be feasible."

References

Abramitzky, R., L.P. Boustan, and K. Eriksson, "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 2014, *122* (3), 467–717.

Atkin, David, "The Caloric Costs of Culture: Evidence from Indian Migrants," *NBER Working Paper No. w19196*, 2013.

Au, C. and J. V. Henderson, "Are Chinese Cities too Small?," *The Review of Economic Studies*, 2006, *73*, 549–576.

Autor, David, "The "Task Approach" to Labor Markets: An Overview," *Journal of Labour Market Research*, 2013, *46* (3), 185–199.

Bauer, Thomas K., Sebastian Braun, and Michael Kvasnicka, "The Economic Integration of Forced Migrants: Evidence for Post-War Germany," *The Economic Journal*, 2013, *123* (571), 998–1024.

Bayer, Patrick, Shakeeb Khan, and Christopher Timmins, "Nonparametric Identification and Estimation in a Roy Model with Common Nonpecuniary Returns," *Journal of Business & Economic Statistics*, 2011, *29*, 201–215.

Bazzi, S., "Wealth Heterogeneity and the Income Elasticity of Migration," *Unpublished manuscript*, 2014.

Bazzi, Samuel and Christopher Blattman, "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, forthcoming.

Beaman, L. A., "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the US," *The Review of Economic Studies*, 2012, *79*, 128–161.

Becker, Gary S, "Investment in human capital: A theoretical analysis," *The Journal of Political Economy*, 1962, *70* (5), 9–49.

Ben Yishay, Ariel and A. Mushfiq Mobarak, "Social Learning and Communication," *Unpublished Manuscript*, 2014.

Bryan, G., S. Chowdhury, and A. M. Mobarak, "Under-investment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, forthcoming.

Busso, M., J. Gregory, and P. Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *The American Economic Review*, 2013, *103*, 897–947.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, "Robust Inference with Multiway Clustering," *Journal of Business & Economic Statistics*, 2011, *29* (2).

Caselli, Francesco, "Accounting for cross-country income differences," *Handbook of economic growth*, 2005, *1*, 679–741.

Clauss, Wolfgang, Hans-Dieter Evers, and Solvay Gerke, "The Formation of A Peasant Society: Javanese Transmigrants in East Kalimantan," *Indonesia*, 1988, (46), 79–90.

Combes, P.-P., G. Duranton, and L. Gobillon, "The Identification of Agglomeration Economies," *Journal of Economic Geography*, 2011, *11*, 253–266.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, 2008, *63* (2), 723–742.

Comin, Diego A., Mikhail Dmitriev, and Esteban Rossi-Hansberg, "The Spatial Diffusion of Technology," *NBER Working Paper No. w18534*, 2012.

Conley, T.G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92*, 1–45.

Conley, Timothy G. and Christopher R. Udry, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

Costinot, Arnaud, Dave Donaldson, and Cory Smith, "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields Around the World," *Journal of Political Economy*, forthcoming.

Dahl, G. B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 2002, *70*, 2367–2420.

Davis, Gloria and Helen Garrison, "Indonesia: The Transmigration Program in Perspective," *Sojourn: Journal of Social Issues in Southest Asia*, 1989, *4* (2), 298–304.

de Sherbinin, A., M. Castro, F. Gemenne, M.M. Cernea, S. Adamo, P.M. Fearnside, G. Krieger, S. Lahmani, A. Oliver-Smith, A. Pankhurst, T. Scudder, B. Singer, Y. Tan, G. Wannier, P. Boncour, C. Erhart, G. Hugo, B. Pandey, and G. Shi, "Preparing for Resettlement Associated with Climate Change," *Science*, 2011, *344*, 456–457.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber, "Linguistic Diversity and Redistribution," *Journal of the European Economic Association*, 2009, *7* (6), 1291–1318.

Diamond, Jared M, *Guns, Germs, and Steel*, W. W. Norton & Company, 1997.

Dinkelman, T., "Can Spatial Mobility Insure Families Against Long-Term Impacts of Economic Shocks? Evidence from Drought and Disability in South Africa," 2012.

Donner, Wolf, *Land Use and Environment in Indonesia*, C. Hurst and Company: London, UK, 1987.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund, "Ethnic Enclaves and the Economic Success of Immigrants–Evidence from a Natural Experiment," *Quarterly Journal of Economics*, 2003, *118* (1), 329–357.

Ellison, Glenn, Edward L Glaeser, and William R Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 2010, pp. 1195–1213.

Esteban, J., L. Mayoral, and D. Ray, "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 2012, *102* (4), 1210–1342.

Fearnside, Philip M., "Transmigration in Indonesia: Lessons from Its Environmental and Social Impacts," *Environmental Management*, 1997, *21* (4), 553–570.

Fearon, James D., "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 2003, *8* (2), 195–222.

Fischer, Marshall L., R. Jaikumar, and Luk N. van Wassenhove, "A Multiplier Adjustment Method for the Generalized Assignment Problem," *Journal of Management Science*, 1986, *32* (9), 1095–1103.

Foster, Andrew D and Mark R Rosenzweig, "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market," *Review of Economic Studies*, 1996, *63* (3), 347–374.

Foster, Andrew D. and Mark R. Rosenzweig, "Microeconomics of Technology Adoption," *Annual Review of Econonomics*, 2010, *2* (1), 395–424.

Friedberg, Rachel M., "You Can't Take It With You? Immigrant Assimilation and the Portability of Human Capital," *Journal of Labor Economics*, 2000, *18*, 221–251.

Gathmann, Christina and Uta Schonberg, "How General Is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 2010, *28* (1), 1–49.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent, "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 2005, *23* (4), 681–724.

Glitz, Albrecht, "The Labor Market Impact of Immigration: A Quasi-Experiment Exploiting Immigrant Location Rules in Germany," *Journal of Labor Economics*, 2012, *30* (1), 175–213.

Gollin, Douglas, David Lagakos, and Michael E Waugh, "The Agricultural Productivity Gap in Developing Countries," *Quarterly Journal of Economics*, 2014, *129* (2), 939–993.

Greenstone, Michael, Rick Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118*, 536–598.

Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 1957, *25* (4), 501–522.

Hardjono, J., "The Indonesian Transmigration Program in Historical Perspective," *International Migration*, 1988, *26* (4), 427–439.

Heckman, James J. and Bo E. Honore, "The Empirical Content of the Roy Model," *Econometrica*, 1990, *58*, 1121–49.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 994–1028.

Hodler, Roland and Paul A Raschky, "Regional Favoritism," *The Quarterly Journal of Economics*, forthcoming.

Hornbeck, Richard, "The Enduring Impact of the American Dust Bowl: Short-and Long-Run Adjustments to Environmental Catastrophe," *The American Economic Review*, 2012, *102* (4), 1477–1507.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow, "The Allocation of Talent and U.S. Economic Growth," *NBER*

Working Paper No. w18693, 2013.

Huffman, Wallace E, "Human Capital: Education and Agriculture," in B. L. Gardner and G. C. Rausser, eds., *Handbook of Agricultural Economics, Volume 1, Chapter 7*, 2001, pp. 333–381.

Huffman, Wallace E. and Tubagus Feridhanusetyawan, "Migration, Fixed Costs, and Location-Specific Amenities: A Hazard Analysis for a Panel of Males," *American Journal of Agricultural Economics*, 2007, *89* (2), 368–382.

IPCC, *Climate Change 2013: The physical Science Basis. Contribution of Woking Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, United Kingdom and New York, NY, USA: Combridge University Press, 2013.

Janvry, Alain De, Kyle Emerick, Marco Gonzalez-Navarro, and Elisabeth Sadoulet, "Delinking Land Rights from Land Use: Certification and Migration in Mexico," 2012.

Jayachandran, Seema, "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries," 2006, *114* (3), 538–575.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *The Journal of Political Economy*, 1979, *87*, 972–990.

Kebschull, Dietrich, *Transmigration in Indonesia: An Empirical Analysis of Motivation, Expectations and Experiences*, Hamburg, Germany: Transaction Publishers, 1986.

Kinsey, Bill H. and Hans P. Binswanger, "Characteristics and Performance of Resettlement Programs: A Review," *World Development*, 1993, *21*, 1477–1494.

Kline, P. and E. Moretti, "Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 2014, *129*, 275–331.

Kline, Patrick, "Oaxaca-Blinder as a Reweighting Estimator," *American Economic Review: Papers & Proceedings*, 2011, *101* (3), 532–537.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental analysis of neighborhood effects," *Econometrica*, 2007, *75* (1), 83–119.

Lagakos, D. and M. E. Waugh, "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, 2013, *103*, 948–980.

Lubotsky, Darren, "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings," *Journal of Political Economy*, 2007, *115* (5), 820–867.

Lucas, R. E. B., "Internal Migration in Developing Countries," in M. R. Rosenzweig and Ol Stark, eds., *Handbook of Population and Family economics*, Vol. 1, Amsterdam, Netherlands: North Holland Publishing Company, 1997, pp. 721–798.

MacAndrews, Colin, "Transmigration in Indonesia: Prospects and Problems," *Asian Survey*, 1978, *18*, 458–472.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding, "Urbanization and Structural Transformation," *The Quarterly Journal of Economics*, 2012, *127* (2), 535–586.

Michalopoulos, S., "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 2012, *102*, 1508–1539.

Michalopoulos, Stelios and Elias Papaioannou, "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, 2013, *129* (1), 151–213.

Moormann, Frans R., *Rice: Soil, Water, Land*, International Rice Research Institute, 1978.

Moretti, Enrico, "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data," *Journal of Econometrics*, 2004, *121* (1), 175–212.

Morten, M., "Temporary Migration and Endogenous Risk Sharing in Village India," *Unpublished Manuscript*, 2013.

Munshi, Kaivan, "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

__ **and Mark Rosenzweig**, "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *Unpublished Manuscript*, 2013.

Olivia, Susan and John Gibson, "Economic Rise and Decline in Indonesia–As Seen from Space," 2013.

Olmstead, Alan L. and Paul W. Rhode, "Responding to Climatic Challenges: Lessons from US Agricultural Development," in "The Economics of Climate Change: Adaptations Past and Present," National Bureau of Economic Research, Inc, 2011, pp. 169–194.

Otten, Mariël, "Transmigrasi: Myths and Realities. Indonesian Resettlement Policy, 1965-1985.," *IWGIA document*, 1986, (57), 1–254.

Qian, Nancy, "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance," *Quarterly Journal of Economics*, 2008, *123* (3), 1251–1285.

Robins, James M, Andrea Rotnitzky, and Lue Ping Zhao, "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, 1995, *90*, 106–121.

Robinson, Peter M, "Root-N-consistent Semiparametric Regression," *Econometrica*, 1988, *56* (4), 931–954.

Roy, Andrew Donald, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, *3*, 135–146.

Sarvimäki, M., R. Uusitalo, and M. Jantti, "Long-term Effects of Forced Migration," 2010.

Steckel, Richard H, "The Economic Foundations of East-west Migration during the 19th Century," *Explorations in Economic History*, 1983, *20* (1), 14–36.

Stern, Nicholas Herbert, *The Economics Of Climate Change: The Stern Review*, Cambridge, UK: Cambridge University Press, 2007.

Suri, Tavneet, "Selection and Comparative Advantage in Technology Adoption," *Econometrica*, 2011, *79* (1), 159–209.

Van Der Eng, Pierre, "Development of seed-fertilizer technology in Indonesian rice agriculture," *Agricultural History*, 1994, pp. 20–53.

World Bank, *World Development Report, 1982: Agriculture and Economic Development*, International Bank for Reconstruction and Development, 1982.

_ , *World Development Report, 1984: Population Change and Development*, International Bank for Reconstruction and Development, 1984.

_ , "Indonesia: The Transmigration Program in Perspective," A World Bank Country Study, World Bank 1988.

_ , *Resettlement and Development: The Bankwide Review of Projects Involving Involuntary Resettlement, 1986–1993*, International Bank for Reconstruction and Development, 1999.

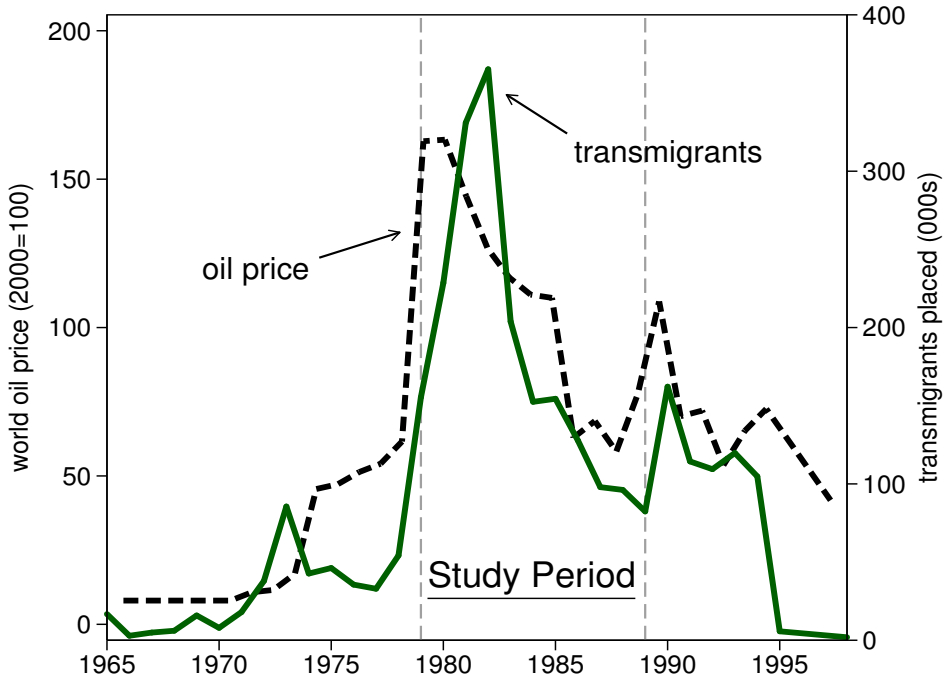
_ , *Doing Business 2009: Country Profile for Indonesia*, Washington, DC: World Bank, 2008.

_ , *Indonesia Country Profile*, World Bank, Washington, DC, 2009.

Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.

Figures

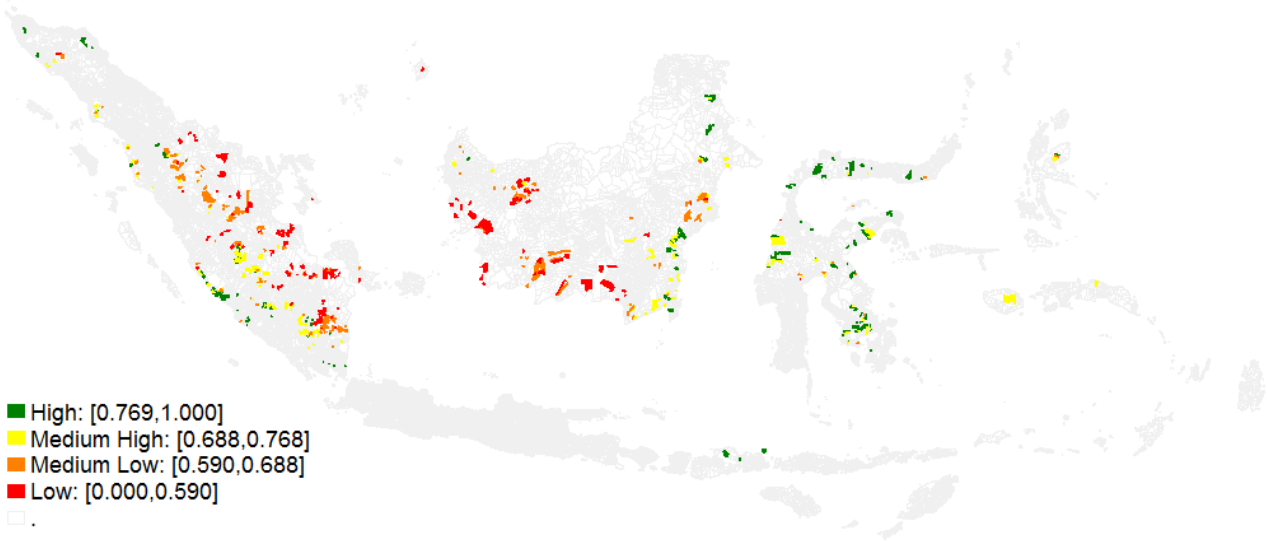
Figure 1: Transmigration Flows and Oil Prices



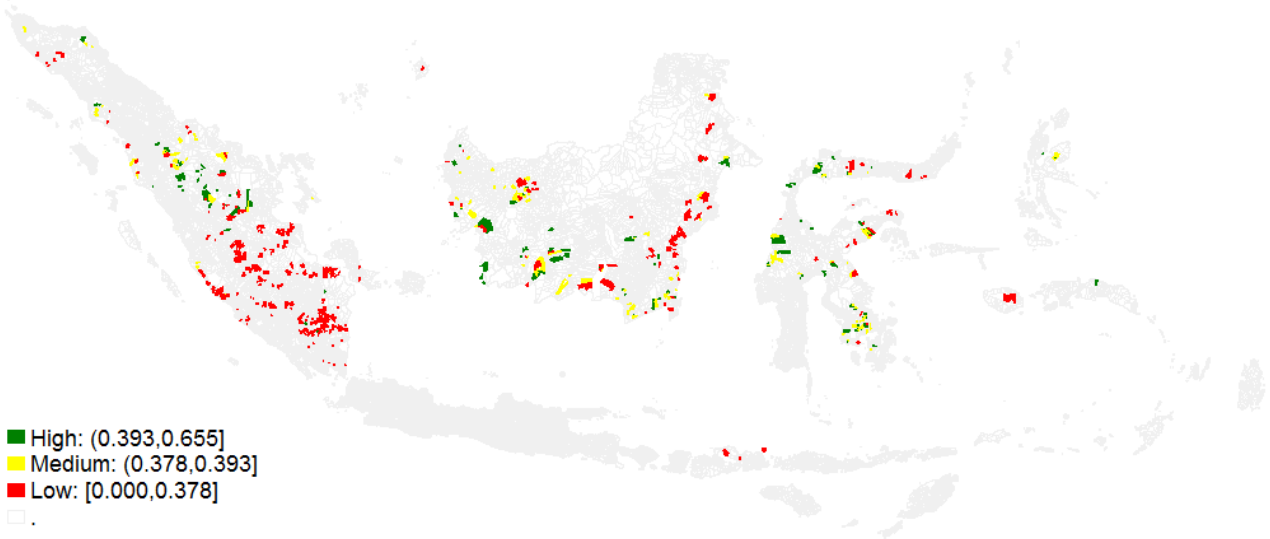
Notes: Authors' calculations from Transmigration Census data. The oil price index is from Bazzi and Blattman (forthcoming). The dark gray vertical lines correspond to our study period.

Figure 2: Maps of Similarity Indices Across Transmigration Villages

(a) Agroclimatic Similarity



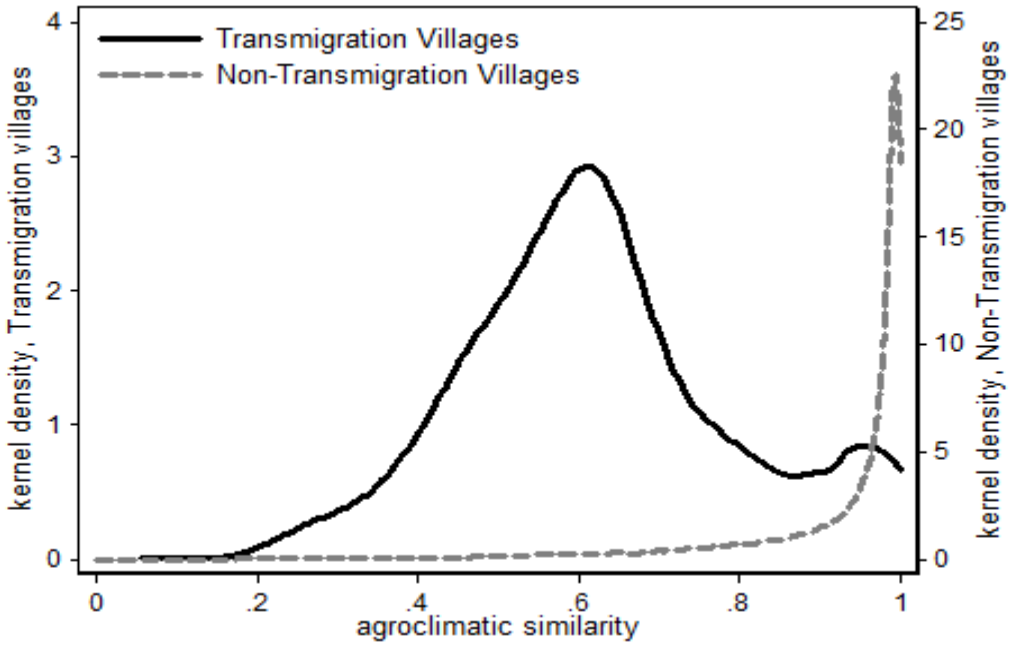
(b) Linguistic Similarity



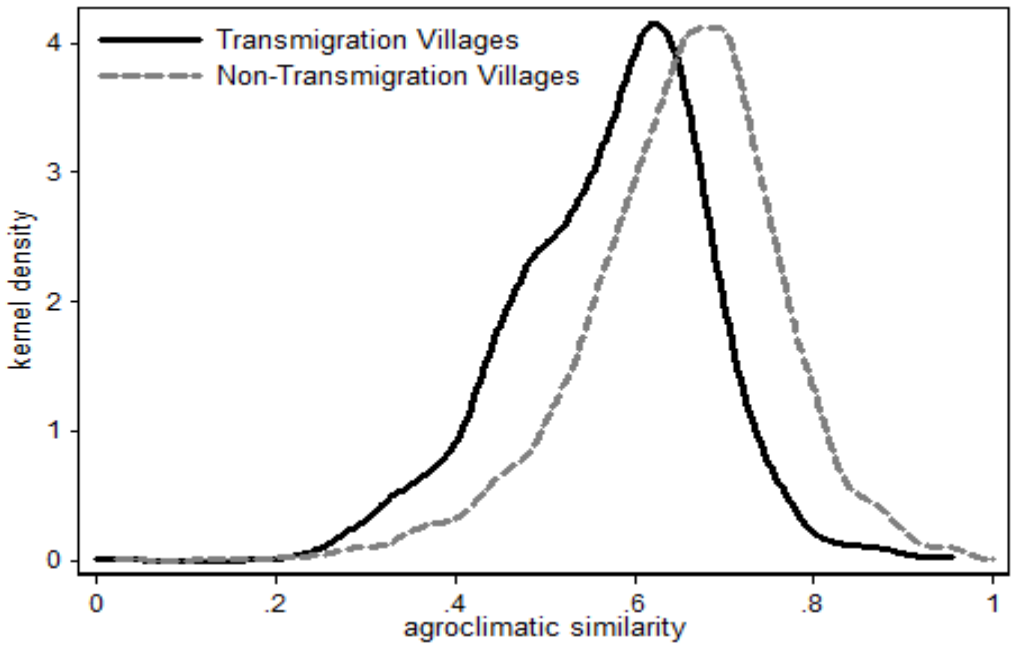
Notes: Each colored location on the map corresponds to a Transmigration village. The agroclimatic similarity index for village j, $\mathcal{A}_j$, is constructed according to equation (1) and is standardized to lie on the unit interval.The linguistic similarity index for village j, $\mathcal{L}_j$, is constructed according to equation (2) and is standardized to lie on the unit interval.

Figure 3: Agroclimatic Similarity: Transmigration vs. Other Outer Islands Villages

(a) All Individuals (Natives and Immigrants)

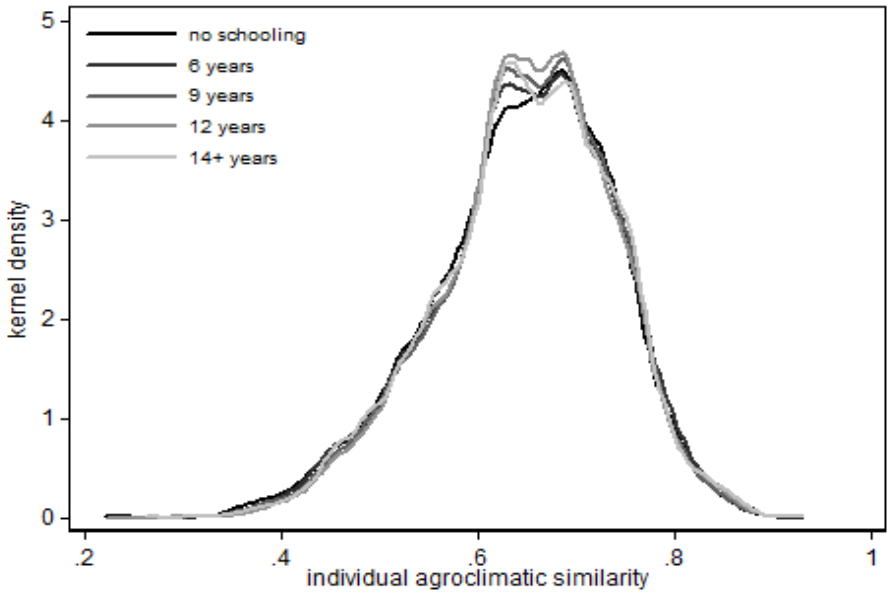


(b) All Individuals (Immigrants Only)



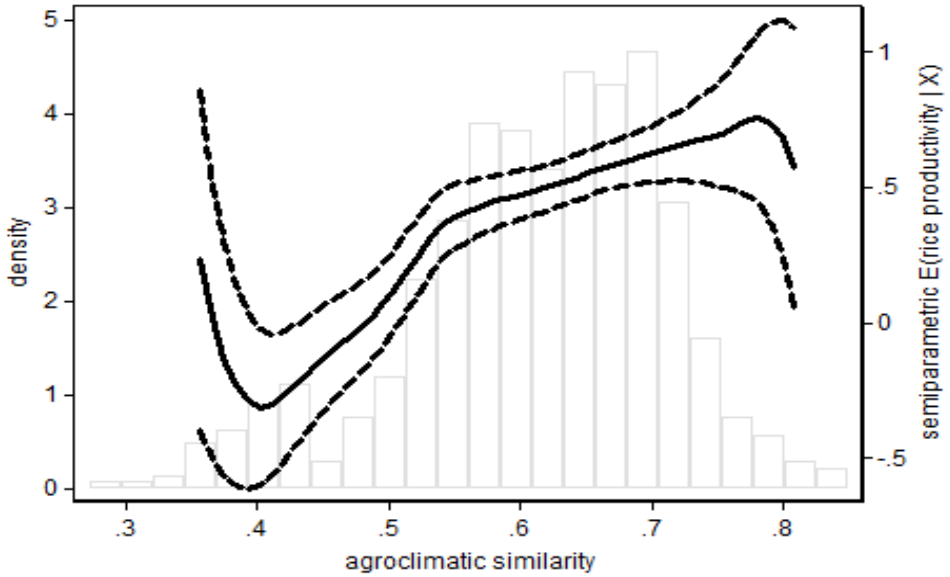
Notes: Panel A shows kernel densities of village-level agroclimatic similarity computed over all individuals—natives and immigrants—in the village separately for Transmigration settlements and all other Outer Islands villages. Panel B shows kernel densities of village-level agroclimatic similarity computed over all immigrants in the village separately for Transmigration settlements and all other Outer Islands villages. The agroclimatic similarity indices for village j, $\mathcal{A}_j$, are constructed according to equation (1) with π_{ij} in (a) being the share of the population in j from each origin district i including $i = j$, and in (b) being the share of the immigrant population in j from each origin district i excluding $i = j$. All indices are standardized to lie on the unit interval.

Figure 4: Individual Agroclimatic Similarity by Schooling: Transmigration Villages



Notes: This figure shows the kernel densities of standardized individual-level agroclimatic similarity, $\mathcal{A}_{ij}$ by level of schooling for all Java/Bali-born individuals living in Transmigration villages and who are between the ages of 15 and 65 and were older than 10 years old in the initial year of settlement. The schooling levels are as reported in the 2000 Population Census.

Figure 5: Baseline Specification: Semiparametric Evidence for Rice Productivity



Notes: This is based on semiparametric Robinson (1988) extensions of the main parametric specification in column 3 of Table 5 relating agroclimatic similarity to log rice productivity. The dashed lines correspond to 90% confidence intervals based on clustering of standard errors at the district level. The local linear regressions use 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 in the figure for presentational purposes.

Tables

Table 1: Agroclimatic Diversity in Java/Bali (Origins) and the Outer Islands (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)

Notes: This table reports summary statistics for each of the variables included in our agroclimatic similarity index. The mean and standard deviation for the given variable are computed over all villages in Java/Bali (Outer Islands) in columns 2-3 (4-5). Sample sizes vary slightly across measures, but the full coverage includes 40,518 villages in the Outer Islands and 25,756 in Java/Bali. See Appendix A for details on data sources and construction.

Table 2: Summary Statistics: Transmigration Villages

	Mean	Std. Deviation	No. of Villages
Demographic Characteristics			
total population (2000)	2,041	(1,283)	814
population per square km (2000)	140	(651)	814
Java/Bali-born population share	0.39	(0.19)	814
Transmigrant ethnicity population share	0.69	(0.29)	814
average years of schooling	4.00	(0.90)	814
Economic Characteristics			
farming employment share	0.69	(0.24)	814
any rice production in village	0.74	(0.44)	814
rice output per hectare (tons)	2.52	(2.81)	600
total agricultural productivity (revenue weighted avg.)	1.00	(2.65)	770
rice is highest revenue crop	0.28	(0.45)	814
Δ log(1+light intensity), 1992-2010	0.16	(0.63)	814
Δ village area with any lights, 1992-2010	0.05	(0.24)	814
Similarity			
$\mathcal{A}_j$: agroclimatic similarity index $\in [0,1]$	0.67	(0.14)	814
$\mathcal{L}_j$: linguistic similarity index $\in [0,1]$	0.59	(0.07)	814

Notes: This table reports summary statistics for Transmigration villages. The number of villages varies across rows depending on data availability. See Appendix A for details on data sources and construction. The similarity indices have been standardized to lie between zero and one. All agricultural outcomes are as observed in the 2001-2 growing season. Rice output per hectare has been winsorized above 20 tons/ha. Total agricultural productivity is winsorized at the fourth maximum order statistic to account for three extreme outliers. All results in the paper are robust to alternative cutoffs or not winsorizing at all.

Table 3: Agroclimatic Similarity and Predetermined Development Proxies (Destinations)

Dependent Variable	agroclimatic similarity
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)
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: */**/*** denotes significance at the 10/5/1 percent level. Each cell corresponds to a regression of agroclimatic similarity on the given variable in the row, island fixed effects, and the predetermined village-level control variables described in the text. Potential yields are obtained from FAO-GAEZ. The variables beginning with "log district population, 1978" are (i) based on data from the 1980 Population Census (available on IPUMS International), (ii) measured at the district level based on 1980 district boundaries, (iii) computed using the sampling weights needed to recover population summary statistics from the sample-based Census, and (iv) 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). Standard errors in parentheses are clustered at the (1980) district level for the Census variables and allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999) for the potential yield variables (which constructed from 10km $\times$ 10km geographic grids).

Table 4: Quasi-Gravity Regression of Migration from Java/Bali to the Outer Islands

Dependent Variable:	$Pr(migrant_{ij} > 0)$		$\ln(migrant_{ij})$	
	(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 Controls (Table 3), Destinations	No	Yes	No	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. This table regresses the stock of migrants from origin district i in Java/Bali residing in Outer Islands village j in the year 2000 on the agroclimatic similarity between i and j and the inverse log great circle distance between i and j. The unit of observation is an origin district i (of which there are 119) by destination Transmigration village j. The dependent variable in columns 1-2 is an indicator equal to one if there are migrants from i in j. The dependent variable in columns 3-4 is the log number of migrants from i in 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 reported in Table 3. Standard errors are two-way clustered by birth district and destination district.

Table 5: Agroclimatic and Linguistic Similarity and Rice Productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
agroclimatic similarity	0.204	0.182	0.210	0.151	0.166	0.438		0.221	0.218
	(0.064)***	(0.045)***	(0.075)***	(0.057)***	(0.068)**	(0.080)***		(0.062)***	(0.061)***
$\cdots \times$ potential rice yield						-0.502			
						(0.065)***			
$\cdots \times$ tercile 1 wetland share $\in [0, 0.16]$							0.372		
							(0.021)***		
$\cdots \times$ tercile 2 wetland share $\in (0.16, 0.66]$							0.108		
							(0.054)**		
$\cdots \times$ tercile 3 wetland share $\in (0.66, 1.0]$							0.089		
							(0.058)		
linguistic similarity								0.134	0.074
								(0.051)***	(0.078)
$\cdots \times$ small initial cohort									0.107
									(0.055)*
Number of Villages	600	600	600	600	600	599	600	600	600
R^2	0.252	0.149	0.277	0.367	0.400	0.260	0.279	0.257	0.259
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{X}_j$)	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin Province Migrant Shares	No	No	Yes	No	Yes	No	No	No	No
Log Weighted Avg. Distance to Origins	No	No	Yes	No	Yes	No	No	No	No
Weighted Avg. Predetermined Controls (Table 3), Origins	No	No	Yes	No	Yes	No	No	No	No
Predetermined Controls (Table 3), Destinations	No	No	No	Yes	Yes	No	No	No	No
Predetermined Demographics and Schooling	No	No	No	Yes	Yes	No	No	No	No

Notes: */**/*** denotes significance at the 10/5/1 percent level. The dependent variable in all specifications is log rice output per hectare. Agroclimatic and linguistic similarity are normalized to have mean zero and a standard deviation of one. All regressions include island fixed effects. All regressions except in column 2 predetermined village-level control variables described in the text. "Origin Province Migrant Shares" are four variables capturing the share of the Java/Bali-born population hailing from the given province. "Log Weight Avg. Distance to Origins" the weighted log great circle distance between j and all Java/Bali districts i with weights equal to the share of the Java/Bali-born population from i. "Predetermined Controls, Destinations" are all of the variables reported in Table 3, and "Weighted Avg. Predetermined..." are those same variables observed in the origins i weighted by the share of j born in i. "Predetermined Demographics and Schooling" are age, gender, and schooling shares for each of the Java/Bali-born and Outer Islands-born populations residing in j and born before the program. "Wetland share" denotes the fraction of agricultural land in j that is wetland. "Small initial cohort" is an indicator equal to one if the village received below the median number of transmigrants placed in the initial year of settlement. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

Table 6: Occupational Sorting within Transmigration Villages

Dependent Variable	**Panel A**: $\mathbf{Pr}(\mathbf{Occupation} = \dots)$ **Farming**			**Trading/Services**		
Age	All **(1)**	Young **(2)**	Old **(3)**	All **(4)**	Young **(5)**	Old **(6)**
individual agroclimatic similarity	0.0090	0.0119	0.0079	-0.0037	-0.0050	-0.0032
	(0.0052)*	(0.0057)**	(0.0053)	(0.0027)	(0.0028)*	(0.0027)
individual linguistic similarity	-0.0139	-0.0153	-0.0134	0.0175	0.0154	0.0183
	(0.0161)	(0.0179)	(0.0155)	(0.0067)**	(0.0068)**	(0.0067)***
Number of Individuals	566,956	175,546	391,410	566,956	175,546	391,410
Dependent Variable Mean	0.622	0.489	0.682	0.099	0.089	0.103

Dependent Variable	**Panel B**: $\mathbf{Pr}(\mathbf{Occupation} = \mathbf{Farming} \dots)$ **Cash Crop**			**Food Crop**		
Age	All **(1)**	Young **(2)**	Old **(3)**	All **(4)**	Young **(5)**	Old **(6)**
individual agroclimatic similarity	0.0042	0.0148	0.0000	0.0047	-0.0029	0.0078
	(0.0067)	(0.0061)**	(0.0074)	(0.0078)	(0.0073)	(0.0082)
individual linguistic similarity	0.0056	0.0016	0.0080	-0.0195	-0.0169	-0.0213
	(0.0158)	(0.0129)	(0.0172)	(0.0221)	(0.0216)	(0.0225)
Number of Individuals	566,956	175,546	391,410	566,956	175,546	391,410
Dependent Variable Mean	0.210	0.175	0.225	0.412	0.314	0.456
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year of Settlement Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{X}_j$)	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. This tables regresses the linear probability that a Java/Bali-born individual living in a Transmigration village as recorded in the 2000 Population Census works in (Panel A): cash or food crop farming (columns 1-3) or trading/services (column 4-6); (Panel B): cash (columns 1-3) or food crop farming (columns 4-6). Columns 1 and 4 include all Java/Bali-born individuals. Columns 2 and 5 restrict to individuals who were less than 10 years old at the time of the initial settlement in their village. Columns 3 and 6 restrict to individuals aged 10 years and greater at the time of the initial resettlement. 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) individual-level controls, including age interacted with a male dummy, married dummy, indicators for seven schooling levels, Java/Bali indigenous ethnic group dummy, immigrant from Java/Bali within the last five years, immigrant from another Outer Islands province within the last five years, immigrant from district within the same (Outer Islands) province within the last five years, and indicators for seven religious groups. Results are similar omitting the individual-level controls. Standard errors are clustered at the district level.

Table 7: Agroclimatic Similarity and Other Development Outcomes

Dependent Variable	**revenue weighted total agricultural productivity**			**Δlight intensity**	**coverage**
Crops	Cash	Food	All	—	—
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
agroclimatic similarity	0.024	0.066	0.025	0.054	0.022
	(0.031)	(0.038)*	(0.052)	(0.029)*	(0.011)**
Number of Villages	712	681	770	814	814
R^2	0.422	0.525	0.387	0.137	0.130
Dep. Var. Mean (Levels)	1.025	0.608	0.996	0.158	0.054
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{X}_j$)	Yes	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. The dependent variable in column 3 is the log of revenue weighted total agricultural productivity as defined in Section 4. Column 2 reconstructs that measure for the subset of all 37 crops that can be readily classified as food crops, including rice, cassava, maize, sweet potato, tubers, taro, soybean, and potato. Column 1 reconstructs the measure for the remaining cash crops, primary among which are palm oil, rubber, cocoa, coffee, and groundnuts. Results are robust to alternative classifications. The dependent variables in columns 4-5 are the two measures of growth in nighttime lights capturing, respectively, the change in log(1 + light intensity) and the fraction of the village with any light coverage between 1992 and 2010. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

Table 8: Against the Grain: Neighborhood vs. Origin Effects in Rice Land Allocation

Dependent Variable	Rice/Staples		Pr(Rice/Staples>0.5)	
	(1)	**(2)**	**(3)**	**(4)**
share of rice Ha in main staple Ha, neighbors	0.157	0.158	0.164	0.166
	(0.023)***	(0.023)***	(0.025)***	(0.025)***
share of rice Ha in main staple Ha, Java/Bali origin		0.021		0.036
		(0.008)***		(0.012)***
Number of villages	694	694	694	694
Dep. Var. Mean	0.684	0.684	0.707	0.707

Notes: */**/*** denotes significance at the 10/5/1 percent level. The dependent variable is farmland area planted with rice as a fraction of area planted with the three main staples of rice, maize, and cassava. In columns 3-4, the share is transformed into a binary outcome equal to one if the share of rice is greater than 50%. The "share of rice hectares (Ha) in main staple Ha, neighbors" is the average share across all villages in the given district excluding Transmigration villages. The "share of rice hectares (Ha) in main staple Ha, Java/Bali origin" is a weighted average of the shares prevailing in the origin districts of Java/Bali with the weights being the share of Java/Bali-born immigrants in the given village from the given origin district. Both variables have been normalized to have mean zero and standard deviation one. All regressions include the usual predetermined village-level control variables and island fixed effects. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

Table 9: Average Treatment Effects of the Transmigration Program

Dependent Variable	(1)	(2)	(3)	(4)
	Panel A: **Demographic Outcomes**			
log population density	-0.390	0.556	0.799	0.769
	(0.118)***	(0.132)***	(0.220)***	(0.170)***
Java/Bali-born population share	0.321	0.355	0.352	0.348
	(0.017)***	(0.018)***	(0.018)***	(0.019)***
transmigrant ehtnicity population share	0.484	0.538	0.516	0.558
	(0.027)***	(0.029)***	(0.046)***	(0.037)***
	Panel B: **Economic Outcomes**			
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 total agricultural productivity	-0.051	-0.193	0.023	0.134
	(0.083)	(0.136)	(0.159)	(0.142)
log light intensity growth	-0.316	0.012	0.035	0.021
	(0.064)***	(0.044)	(0.076)	(0.066)
light coverage growth	-0.126	0.011	0.025	0.019
	(0.025)***	(0.016)	(0.034)	(0.023)
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: */**/*** denotes significance at the 10/5/1 percent level. Each cell reports the coefficient from a regression of the given dependent variable on an indicator for whether the village is a Transmigration village. Panel A outcomes are as observed in the 2000 Population Census. Panel B agricultural outcomes are as observed in 2001-2 growing season. Column 1 comprises all Outer Islands villages (with non-missing data). Column 2 restricts to our quasi-experimental design including only Transmigration and control/RDA sites and conditions on the predetermined village-level characteristics that explain (sequential) site selection. Column 3 is a double robust specification that (i) reweights controls by normalized $\widehat{\kappa} = \widehat{P}/(1-\widehat{P})$ where $\widehat{P}$ is the estimated probability that the village is a Transmigration settlement and (ii) controls for the predetermined village-level characteristics. Column 4 is a control function specification based on a Blinder-Oaxaca decomposition developed in Kline (2011). All specifications include island fixed effects. Sample sizes vary across outcomes (depending on data availability) and columns but include as many 31,185 villages in column (1), and 814 treated villages and 668 controls in column 2-4. Standard errors clustered by district in parentheses and estimated using a block bootstrap in column 3 to account for the generated $\widehat{\kappa}$ weights.