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*

December 2014

Abstract

We use a natural experiment in Indonesia to provide causal evidence on the role of location-specific human capital in shaping the spatial distribution of productivity. From 1979–1988, the Transmigration Program relocated two million migrants from rural Java and Bali to new rural settlements in the Outer Islands. Villages that were assigned migrants from regions with more similar agroclimatic endowments exhibit higher rice productivity and nighttime light intensity one to two decades later. While we find some evidence of migrants' adaptation to agroclimatic change, our results suggest that regional productivity differences may overstate the potential gains from migration.

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

Keywords: Spatial Labor Allocation, Internal Migration, Comparative Advantage, Agricultural Adaptation

*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 Urban Workshop, 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`.

†Corresponding author: 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

Geographic mobility is a core feature of the development process. Throughout history, soil and climate conditions have shaped migration and the diffusion of human capital and technology. Steckel (1983) and Diamond (1997) document a striking tendency for migrants and technologies to diffuse east–west rather than north–south in the historical process of settling the agricultural frontier. Griliches (1957) and Comin et al. (2012) highlight a similar pattern of spatial diffusion within agroclimatic zones. These long-run patterns suggest that similarity in agroclimatic conditions between locations may play an important role in determining the transferability of human capital and, hence, the distribution of productivity across space. Yet, we have limited causal evidence of these relationships because skill transferability is difficult to measure, migrants endogenously sort into places where their skills are transferable, and these spatial diffusion processes are slow and often confounded by time trends.[1]

This paper uses a remarkable policy experiment in modern Indonesia to provide causal evidence on the role of location-specific human capital and skill transferability in shaping aggregate productivity. Between 1979 and 1988, the Transmigration Program in Indonesia relocated two million voluntary migrants (hereafter, transmigrants) from the Inner Islands of Java and Bali to *newly created* agricultural settlements in the Outer Islands. We develop a novel proxy for skill transferability based on the similarity in agroclimatic conditions between two locations. Atkin (2013) and Michalopoulos (2012) provide fascinating evidence that migrants in India and Africa tend to grow and consume crops that are predominant in their native origin. Using the plausibly exogenous assignment of two million transmigrants across settlements, we identify large causal impacts of location-specific human capital on productivity at the destination, suggesting some skills may not be easily transferable across space. The exogenous assignment addresses a pervasive identification problem in the study of migration, and our measure of agroclimatic distance allows us to quantify how and why skill specificity matters.

Our findings are important for several reasons. First, recent debate questions whether labor is spatially misallocated and whether there exist potential productivity gains through labor reallocation (e.g., Gollin et al., 2014; Munshi and Rosenzweig, 2014; Young, 2013). If some skills are not easily transferable across locations, then spatial productivity gaps may not, in fact, represent labor market arbitrage opportunities. Second, understanding how shifting agroclimatic conditions affect farmer productivity is important in light of climate change. Many rain-fed, subsistence farmers in developing countries may lack the resources needed to adapt to agroclimatic change. Moreover, extreme weather events heighten the risks of population displacement with the estimate for South Asia alone exceeding 60 million people (IPCC, 2014; Stern, 2007). Third, our findings provide policy lessons for the optimal design of resettlement policies.[2] Natural disasters, conflicts, and infrastructure development continue to displace millions annually, necessitating resettlement (World Bank, 2004). Recognizing that some vulnerable groups may lack the resources to move voluntarily, various governments have begun planning for resettlement, as a last resort policy response to displacement (see IPCC, 2014; de Sherbinin et al., 2011).

The Transmigration program provides a rich empirical context for studying the important relation-

[1] See Bauer et al. (2013) and Autor (2013) for discussions on measuring and identifying the importance of skill transferability.

[2] 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 resettlement programs in the United States and Europe (Bauer et al., 2013; Beaman, 2012; Edin et al., 2003; Glitz, 2012; Sarvimäki et al., 2010).

ship between skill transferability and productivity (Becker, 1962) and also a unique lens into the historical process of settling the (agricultural) frontier. The program was designed to alleviate overpopulation concerns in rural Java/Bali and to develop the Outer Islands. The government provided households with free transport to the new settlements, housing, and two hectare farm plots assigned by lottery.

We leverage the unprecedented spatial scope and plausibly exogenous relocation process of the program for identification. 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. 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 assignment of transmigrants to new settlements. We show that our skill transferability proxy is largely unrelated to predetermined development outcomes, potential agricultural productivity, and individual schooling. Moreover, the spatial distribution of these Java/Bali-born migrants does not follow gravity patterns typically associated with endogenous sorting.

A key innovation of this study is our proxy for skill transferability. Farming often requires location-specific production methods and associated technical know-how (Griliches, 1957). Our proxy, *agroclimatic similarity*, is higher when the agroclimatic endowments (and hence growing conditions) between migrants' origin and destination regions are more similar. We construct this measure using detailed geospatial data capturing topography, climate, and predetermined 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 Island villages is more similar to transmigrants' languages. Both measures exhibit rich variation because Indonesia is home to over 230 million people from 700 ethnolinguistic groups, living on more than 1,000 different islands. Importantly, agroclimatic and linguistic similarity are uncorrelated, consistent with the ad hoc planning process generating exogenous variation in our similarity measures.

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. Using a multi-location Roy model, we show that agroclimatic similarity provides a novel and exogenous measure of comparative advantage.[3] Farmers can transfer their human capital more successfully if destinations are more similar to their birth locations. Hence, for a given destination, migrants from similar origins have greater comparative advantage and are more productive relative to migrants from dissimilar origins.

Despite being one of the largest resettlement policies ever implemented, we know relatively little about the Transmigration program due to a lack of data covering the scope of the program. We address this by digitizing a 1998 census of Transmigration villages and the planning maps used to identify settlements in the Outer Islands. Our main dataset combines these novel data sources with the geospatial data mentioned above, data on individuals' birth districts and ethnicities from a 2000 Population Cen-

[3]The identification problem of endogenous sorting based on unobservable comparative advantage was first highlighted by Heckman and Honore (1990) and spans multiple fields in economics. 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; Lagakos and Waugh, 2013; Qian, 2008; Suri, 2011), and trade (Costinot et al., forthcoming).

sus, and village-level agricultural activity from a 2002 administrative census. Our primary village-level outcome is rice productivity. We focus on rice because it was the focal crop of the program and grown by a majority of farmers in both the Inner and Outer Islands. This allows us to measure productivity losses in the Outer Islands due to the imperfect transferability of skills acquired in the Inner Islands. Rice is also the basic staple for Indonesia and more than half of the world (Nguyen, 2002; Peng et al., 1995), is grown on 144 million farms worldwide (more than for any other crop, Mohanty et al., 2013), and is the crop expected to be most vulnerable to climate change. We also investigate productivity for other crops and nighttime light intensity in 2010 (a proxy for local income, 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 village-level rice 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. We show further that the productivity gains from skill transferability are larger in adverse growing conditions. Also, the largest effects are found in the bottom tercile of agroclimatic similarity, suggesting a concave adjustment process. In contrast, agroclimatic similarity has null effects on the productivity of cash crops. Since most of these crops were not grown in Java/Bali during the time of the program, skills acquired in Java/Bali should indeed be less important for their productivity. This result serves as a placebo test of agroclimatic similarity as a proxy for skill transferability.

These findings for Indonesia's most important staple crop 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 Bowl. These barriers are particularly salient in developing countries, where an extensive literature documents the importance of local agroclimatic conditions in this adjustment process (e.g., Conley and Udry, 2010). The large productivity losses of agroclimatic dissimilarity that we estimate for a major staple crop like rice suggest that skill transferability plays an important role in adjusting to agroclimatic change, particularly for rain-fed, subsistence farmers. This may imply added costs of climate change to the extent that existing projections do not fully incorporate the difficulty of adjusting to (abrupt) changes in growing conditions.

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, cash crops generate slightly more revenue in low similarity villages. This is consistent with Costinot et al. (forthcoming) who use a simulated 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 percentage point (p.p.) greater likelihood of Java/Bali migrants choosing farming as their primary occupation while a one standard

[4]Our index is scaled between zero and one, with a relatively large standard deviation of 0.14.

deviation increase in linguistic similarity leads to a 1.8 p.p. 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 show that agroclimatic similarity still has positive effects on the level of economic development in 2010, as proxied by nighttime light intensity. This proxy measure is increasingly used in studies exploiting highly localized identifying variation as we do here (e.g., Hodler and Raschky, forthcoming; Michalopoulos and Papaioannou, 2014). Our estimates imply that a one standard deviation increase in agroclimatic similarity leads on average to 1.8-5.2 percent greater income by 2010 (based on local income elasticities of light intensity, Olivia and Gibson, 2013). Coupled with the large effects on rice productivity and the evidence on adaptation, these results suggest that the adjustment process was costly and may be incomplete.

Our study contributes to the growing literature on migration and the 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). 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 imply more limited gains from labor reallocation than might be inferred from the productivity differences—and hence perceived arbitrage opportunities—across locations.[6]

We conclude with two policy exercises that demonstrate the importance of matching people (skills) to places (production environment) when designing resettlement schemes. First, we approximate an optimal reallocation of transmigrants across settlements on the basis of agroclimatic similarity and find that the program could have achieved 27 percent higher aggregate rice yields. Second, we use a policy discontinuity to estimate average treatment effects of the Transmigration program by comparing settlement villages against planned settlement areas 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 Busso et al., 2013; Kline and Moretti, 2014), we find null average impacts on local development outcomes, which stem in part from the persistent effects of agroclimatic similarity in treated villages.

Although the Transmigration program was unique in many ways, our findings offer general and policy-relevant insight into the (re)settlement process. Migrants' ancestral origins and the long-run pro-

[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 Schönberg, 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]For example, Morten and Oliveira (2014) interpret the large wage differentials across Brazil as evidence of underexploited labor market arbitrage due to transportation frictions.

cess of settling the frontier have had persistent impacts on today's economic landscape (Ashraf and Galor, 2013; Putterman and Weil, 2010). The unique features of Transmigration—the speed, scale, and scope of resettlement, the remoteness of the new villages, and the common institutional contexts and initial conditions for *all* settlements—are precisely what allow us to isolate the causal impact of skill transferability on economic development in a way that has not been feasible in slowly changing historical contexts. Our focus on the agricultural sector remains important today given that it employs 1.3 billion people globally (World Bank, 2009) and is at the core of ongoing debates about world income inequality (see Caselli, 2005). With mounting pressure to resettle millions of vulnerable, subsistence farmers—individuals not unlike Indonesia's transmigrants—our findings highlight an important aspect of well-designed relocation and assistance packages.

The remainder of the paper proceeds as follows. Section 2 provides background on the Transmigration program. Section 3 describes the sample construction and presents our key proxies for skill transferability and development outcomes. Section 4 develops our theoretical framework and empirical strategy in the context of a Roy model. Section 5 presents our main results. Section 6 reports the policy exercises. 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), 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. Rice is the single most important crop grown and consumed across the archipelago. Food security, and in particular rice self-sufficiency, has been an overarching policy goal throughout our study period (Kebschull, 1986; McCulloch and Timmer, 2008).

2.1 Program Background

Indonesia's Transmigration program was designed primarily to alleviate these perceived population pressures. Over several decades, the program relocated many transmigrants from rural areas of Java and Bali to rural areas of the Outer Islands. Planners hoped that the program would increase national agricultural output (especially rice) by moving farmers to unsettled areas, and also promote nation building by integrating diverse ethnic groups (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 development period (or *Pelita*).[7] At that time, the government chose to support rainfed food crops because Indonesia was the world's largest importer of rice (the primary staple), and annual crops were the quickest to establish, promoted early

[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 (see World Bank, 1982, 1984).

self-sufficiency, and were the crops with which farmers in Java/Bali had centuries of experience (Geertz, 1963). All Transmigration settlements were initially under central government jurisdiction with administration transferred to local governments after 5–10 years (World Bank, 1988).

The Transmigration program was one of the largest resettlement program during its time and involved complex logistics in both the Inner and Outer Islands. Participation in the program was almost entirely voluntary.[8] Transmigrants were given free transport to 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 to be transported in groups (e.g., 50 households at a time) to the Outer Islands. Program officials identified land reserves that could be cleared, prepared for agricultural use, and connected 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. Additionally, a small amount of land at each site (about 10 percent) was reserved for members of the indigenous population, who could move from nearby areas to take advantage of access to new land.

2.2 Selecting People, Places, and Assignments

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

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 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]

Transmigrants are very similar to the target beneficiaries of potential resettlement programs (IPCC, 2014). The government-sponsored transmigrants are more comparable to stayers than to typical non-sponsored or spontaneous migrants. On average, Java/Bali-born individuals who moved to Transmigration villages had 0.5 fewer years of schooling compared to stayers in Java/Bali (based on the 2000 Population Census discussed in Section 3). 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.

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. We use this multi-stage site selection process later to identify planned but untreated sites.

Numerous reports indicate that the process was not as detailed as planners had hoped. For example,

[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 percent (3.7 percent) of Indonesia's (Outer Islands) population lived on Papua in 2010.

[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.

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 median Transmigration village has Java/Bali migrants from 46 sending districts (out of 119) and three different ethnic groups (out of eight). The origin district Herfindahl index in the median village is 0.12, suggesting that concentration is not high. Here, we provide 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, planners had little interest in or resources for matching transmigrants on the basis of agro-climatic conditions. Many planners believed that Javanese and Balinese farmers had superior farming skills.[10] 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. Planners also expected to provide agricultural extension services as well as irrigation to help farmers adjust at the new settlements, but these plans were never realized due to sharp budget cutbacks in the mid-1980s. Moreover, matching transmigrants' skills to destinations would have required a large amount of data on individual farming skills and up-to-date information on actual growing conditions in available settlements—details largely unavailable at the time.

Third, the fact that there were only four transit camps (one for each province of Java/Bali) ensured a rich mix of origin districts at the destination sites. Each transit camp would comprise transmigrants from many origin districts. Furthermore, motivated by the nation-building goals of the program, planners often explicitly assigned groups of migrants from each of the four provinces to a single settlement (with province specific housing blocks, Levang, 1995). Having only four transit camps also meant that each camp was quite crowded so that transmigrants could not stay long at the camps. In our interviews with transmigrants, most were transported to the Outer Islands within a few days. 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 could not choose their destination in the Outer Islands (Levang, 1995).[11] Previ-

[10] This came up in several of our discussions with officials involved in the program at the time. These Java-centric views are thoroughly discussed in Dove (1985).

[11] In theory, transmigrants could opt-out if they did not like the assigned destination. However, in practice, this was difficult because they had already sold all their assets by the time they travelled to the transit camps and most were at the transit

ous studies argue that 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, and most transmigrants expected to pursue the same sort of (rice) farming activities they had been practicing in their origin villages.

Finally, as we argue at greater length in Section 4.2, it appears that it was difficult for transmigrants to migrate ex post. This was perhaps due to the illiquidity and imperfections of rural land markets (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), and as discussed above, the relatively poor government-sponsored transmigrants may not be as mobile as typical migrants.

2.3 External Validity Concerns

The Transmigration program is globally recognized as one of the largest resettlement schemes ever implemented. Surprisingly, we know relatively little about such a large scale program, perhaps due to a lack of data. There are other examples of large government-sponsored resettlement programs that were implemented with various goals, including population redistribution and agricultural development similar to Transmigration.[12] Due to corruption, the lack of property rights and the large number of individuals in need of resettlement, purely compensatory subsidies often do not suffice in developing countries (Cernea, 1999), which is why many governments have resorted to resettlement.

Regardless of the objectives, resettlement has affected millions of households, cost billions of dollars, and is growing in importance due to the millions expected to be displaced due to extreme weather events, infrastructure development, as well as conflict. The International Organization of Migration estimates that 200 million people may become environmentally-induced migrants by 2050, and various governments have started planning for resettlement, as some of these displaced persons may need government assistance in moving (IPCC, 2014). Many of these migrants are rainfed, staple subsistence farmers who, like transmigrants, lack the resources to move on their own.

3 Data: Measuring Skill Transferability and Its Effects

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 (MOT). These villages, which span the vast Outer Islands of the country, 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),

camps for a few days only, which made it harder to wait for their preferred destination.

[12]Other examples of government-sponsored resettlement schemes include the Polonoroeste program in Brazil that relocated 300,000 migrants between 1981 and 1988 at a cost of US$ 1.6 billion (Hall, 1993), villagization programs in Ethiopia that relocated 440,000 households between 2003 and 2005 with the plan to resettle 1.5 million by 2013, the resettlement of 4 million migrants in Mozambique between 1977 and 1984, and another 43,000 households that were relocated following floods in the 2000s. Additional resettlement programs can be found in India, China, Vietnam, as well as refugee settlement programs in Finland, Germany, Sweden and the United States.

with smaller numbers on Maluku and Nusa Tenggara. The wide geographic scope in destination sites exposed transmigrants to a range of economic conditions. Below, we first discuss how we proxy for skill transferability across locations. We then describe the development outcomes we use to identify the importance of skill transferability.

3.1 Proxies for Skill Transferability

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 Schónberg (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). 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.[13]

Additionally, we capture linguistic similarity between origins and destinations, building on related measures in Desmet et al. (2009), Esteban et al. (2012), and Fearon (2003). 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.

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).[14] All land attributes and ethnolinguistic homelands are predetermined and hence unaffected by settler farming activities and any corresponding inflow of capital or labor. For instance, all of the soil type information is based on data from 1971–1981. Since land and local climate characteristics change slowly, agroclimatic characteristics measured in the 1970s are still highly predictive of productivity in 2000. We can therefore abstract from reverse causality concerns.

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. The differences between Inner and Outer Islands are partic-

[13] 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."

[14] 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).

ularly 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.[15]

Given a G-dimensional vector of agroclimatic characteristics, $\mathbf{x}$, 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 location i and location j, using a metric defined on the space of agroclimatic characteristics.[16] 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 the sum 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 across i 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 π_{ij} is the share of migrants residing in transmigration village 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 other weights and distance metrics in robustness checks. We use $\mathcal{A}_j$ in our main village-level analysis but occasionally use $\mathcal{A}_{ij}$ for individual-level analyses. Therefore, we refer to $\mathcal{A}_{ij}$ ($\mathcal{A}_j$) as individual-(village-) level agroclimatic similarity.

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.[17] *Linguistic similarity* for village j can then be represented as

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

[15] 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.

[16] 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.

[17] 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.

where $\pi_{\ell j}$ is the share of immigrants in j from ethnolinguistic group ℓ in Java/Bali (according to the 2000 Census), $branch_{j\ell}$ 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 Island 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 (see Figure 2) exposed transmigrants to many different types of agroclimatic and linguistic settings. Table 2 reports summary statistics based on the 2000 Census, showing that the average Transmigration village has around 2,000 people (or 140 people per square kilometer), close to 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).[18]

3.2 Key Development Outcomes

We study the impact of skill transferability on local economic development at the *village* level. First, we measure agricultural productivity using the triennial administrative census known as *Podes* (or Village Potential). The August 2002 round was combined with the Agricultural Census and 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 productivity for several reasons. As noted above, rice is one of the most important crops in the world, it is Indonesia's most important staple food, and its growth was a major policy goal. Crucially, rice was the primary crop grown in Java/Bali during the Transmigration period, making it the ideal crop for studying the transferability of skills acquired in Java/Bali.[19] As the program promoted rainfed food crops and especially rice, most transmigrants expected to be able to grow rice in their new villages. From Table 2, we see that on average, close to 70 percent of residents were employed in farming (based on the 2000 Population Census). Average rice yields are 2.5 tons per hectare among villages with any production. We winsorize yields at 20 tons/ha to account for possible misreporting, but our results are robust to alternative cutoffs.

We also construct cash crop productivity as a placebo outcome. As a proxy for skill transferability, agroclimatic similarity should only affect productivity for crops in which Java/Bali-born farmers have growing experience. In the late 1970s, less than five percent of farmers in Java/Bali were growing cash crops, according to the 1976 and 1980 (inter-)Census. One of the most important of these crops is tobacco, which is often produced in rice-growing areas of Java/Bali but is not grown in the Outer Islands.

To aggregate across cash crops, we construct an average cash crop productivity by taking a revenue-weighted average log tons per hectare across crops. We follow Jayachandran (2006) and normalize the productivity of each crop to mean one for comparability. We construct revenue weights based on average

[18]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.

[19]According to the 1983 *Podes*, rice was grown in 88 percent of villages in Java/Bali, 77 percent in Sumatra, 84 percent in Kalimantan, and 63 percent in Sulawesi. Also, 88 percent of villages reported rice as their primary staple.

national unit producer prices in 2001-2 from FAO/PriceSTAT.

The revenue weights allow us to compare productivity across crops in potential revenue terms. However, this average productivity approach using national prices is subject to the strong caveat that we are unable to measure the local prices and substantial inter-crop differences in unit production costs that would be needed to construct ideal weights based on profits. While the median village grows 8 crops, rice is the most important crop across villages. Grown in nearly 75 percent of villages (more than any other crop), rice is the modal top revenue value crop and is among the top three in nearly 65 percent of villages. Other food crops such as maize and cassava as well as cash crops such as palm oil, rubber, and cocoa are important in revenue terms after rice (see Appendix Table B.1).

We capture broader economic development over the medium-run 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). The level of nighttime light intensity in 2010 serves as our main proxy for overall economic growth at the village level. We can assign a growth interpretation because the Transmigration villages had no inhabitants and hence no lights prior to the program. By 2010, nearly 25 percent of Transmigration villages recorded any nighttime lights. By comparison, 12 percent of the 12.5 km^2 geographic grids across sub-Saharan Africa studied in (Michalopoulos and Papaioannou, 2014) have recorded lights in 2007 and 2008. As in other studies, we account for the prevalence of zeros by considering two measures of light activity: (i) $\ln(c+light\ intensity)$ for some small constant c where $light\ intensity \in [0, 57.6]$ in our villages, and (ii) the fraction of the village covered by any lights.

In summary, we have seven main data sources. These include maps to capture light intensity, agroclimatic attributes (HWSD), temperature and precipitation data (UDel), and linguistic homelands (WLMS), as well as the 2000 Population Census, the 2002 Village Census *Podes*, and the 1998 Transmigration Census. We also use several auxiliary datasets, including the FAO-GAEZ data on potential agricultural yields by crop, a 2004 survey (*Susenas*) that 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 these data sources in Appendix A.

While a major benefit of the large spatial scope of the program is the rich variation in agroclimatic attributes, it also poses data constraints. Transmigration villages represent less than five percent of the more than 60,000 villages in Indonesia. As a result, coverage limitations make it difficult to study productivity effects at the individual level. Datasets with useful individual outcome data contain too few observations (e.g., the *Indonesia Family Life Survey* includes only 50 households with Java/Bali-born migrants in Transmigration villages settled during our study period). Our best individual-level dataset, the 2000 Population Census covers all settlement areas and contains the abovementioned demographic characteristics as well as schooling, but does not record productivity outcomes such as wages or agricultural yields. Nevertheless, our granular village-level maps and administrative censuses provide strong measures of local productivity.

4 Theoretical and Empirical Framework

This section lays out our conceptual framework. We first explain how agroclimatic 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.

4.1 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. For now, we assume everyone is a farmer and all farmers grow rice. 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 3.1. 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). Hereafter, we denote $\mathbf{x}_{b(i)}$ with $\mathbf{x}_i$ to simplify notation. 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.[20]

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, $\mathcal{A}_{ij}$ is our measure of agroclimatic similarity between locations. If skills are *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 (acquired at origins) 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

[20] 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.[21] Such villages therefore have greater comparative advantage at farming than villages assigned a high share of migrants from dissimilar origins.

Having described the determinants of productivity, we next characterize the location choice process to motivate why resettlement programs might provide useful natural experiments. 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, 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. This is quite difficult because location choice (V_{ij}) and productivity (y_{ij}) are often confounded (Combes et al., 2011).[22] Moreover, it is challenging to find an instrument capable of generating a strong "first stage" for *each* of the J potential locations in addition to satisfying the exclusion restriction (Dahl, 2002). This can be easily seen in a stylized two-sector Roy (1951) assignment model with 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 into locations based on comparative advantage, then we would only observe two of the four outcomes, namely those associated with high similarity: 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 free to sort. With this in mind, we derive our key estimating equation and discuss threats to identification.

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

[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, 2014, for recent examples).

4.2 Empirical Strategy

Our goal is to estimate the elasticity of aggregate productivity with respect to agroclimatic similarity. 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 at the individual and village level:

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

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

We obtain our village-level estimating equation by aggregating across i:

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

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)), $\eta_j^u \equiv \sum_{i \in I_j} \eta_i^u$ is unobserved demographic characteristics summed over a non-random set of individuals (I_j) whose optimal location is j, and ω_j is an idiosyncratic error term.[23]

The key parameter of interest, γ, measures the semi-elasticity of aggregate agricultural productivity with respect to average agroclimatic similarity for the village. The common identification concerns are that endogenous location, crop and occupation choices may be confounded with unobservable determinants of productivity. The ideal experiment to estimate skill transferability across locations in the agricultural context should (i) randomly assign farmers from many origins to many destinations (to abstract from endogenous location choices),[24] and (ii) ensure that all migrants remain farmers growing the same crops at the origins and destinations (to address selection due to endogenous occupation and crop choices). Moreover, we need farmers growing the same crops at the origins and destinations to interpret agroclimatic similarity as a proxy for the transferability of skills acquired at the origins.

Our research design approximates this ideal experiment. The broad spatial scope and exogenous relocation process from the Transmigration program generate 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 serve to tie the first generation movers to farming. We study productivity effects on rice, which is the ideal focal crop for reasons discussed in Section 3.2. Because transmigrants' farming skills acquired in Java/Bali were mostly specific to rice, we view productivity of other non-rice and, in particular, cash crops as placebos. Below, we provide evidence that the program gives rise to plausibly exogenous variation in agroclimatic similarity. In Section 5.2, we characterize ex post adaptation responses and show that selection via crop,

[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]We need farmers from *many* origins assigned to *many* destinations to estimate the average elasticity for the population. This is easiest to see in the stylized two-by-two example, where the full set of potential outcomes are y_{LL}, y_{HH}, y_{LH}, y_{HL}. If we only observe farmers from lowland origins assigned to highland and lowland destinations, we would worry that the elasticity we estimate may not be representative of skill transferability for farmers from highland origins. Likewise, we may be concerned if we only observe farmers from lowland and highland origins assigned to lowlands only.

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 ($\mathbf{x}$ 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 weights from origins with similar versus dissimilar $\mathbf{x}$.[25]

Appendix Figure B.1 helps to illustrate. We highlight a few agroclimatic characteristics in two nearby Transmigration villages in Sumatra and the district that sent the largest share of migrants to each. Our index aggregates across all sending districts but we focus on the primary sending district in order to simplify the figure. Consider the village of *Telang Sari*, which has an agroclimatic similarity index of 0.5 and has low elevation and low topsoil pH. Its primary sending district (Kebumen in Central Java) also has low elevation and low topsoil pH. By contrast, the nearby village of *Nunggal Sari*, which also has low elevation and low topsoil pH, has a lower agroclimatic similarity index of 0.4, because its primary sending district (Karanganyar in Central Java) has high elevation and high topsoil pH.

We 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 policy gives rise to more low similarity realizations). 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]

We also show that the distribution of agroclimatic similarity across the 814 Transmigration villages is approximately what would be observed under random assignment. 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. Based on 10,000 random similarity indices, we cannot reject that the means and standard deviations of the random and actual distributions are equal.

Balance Checks. Our main identification threat is that high and low similarity villages are not comparable because of unobserved differences in demographic compositions (η_j^u) and unobserved differences

[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.

in natural advantages (μ_j^u) that are correlated with our outcomes. The latter is potentially problematic because destinations that are agroclimatically similar to Java/Bali may have unobservable natural advantages given that Java/Bali is known to be naturally advantaged for rice production.

We first show in Table 3 that pre-program correlates of productivity are not correlated with agroclimatic similarity. 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 (i) potential agricultural productivity based on auxiliary agronomic data from the FAO, as well as (ii) Census-based measures of 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] Recall 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 yields of rice as well as other major food and cash crops. This rules out first-order concerns about unobserved natural advantages.[28] 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 predetermined measures of development in surrounding villages. 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 ($\rho = -0.03$), which is consistent with the plausibly exogenous assignment of heterogeneous people to heterogeneous places.[29] Overall, the evidence suggests that agroclimatic similarity is balanced across Transmigration villages, even as observed up to two decades after resettlement.

5 Empirical Results

We begin by reporting large average effects of skill transferability on rice productivity. We relate this to recent work on location-specificity of consumption and growing preferences of migrant farmers. We turn next to heterogeneity analysis to identify where similarity matters most. We then explore possible mechanisms of adaptation and report results for broader economic development that suggest incomplete adjustments within our study period. Finally, we rule out additional threats to identification, including ex post sorting.

[27]We use the FAO's GAEZ data (see footnote 14) 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]It is also important to note that in agronomic terms, newly cleared wetland is generally of lower quality than long-tilled wetland whereas the opposite holds for newly cleared dryland. To the extent that Java/Bali has relatively more wetland, agroclimatic similarity is plausibly negatively correlated with the unobservable quality of newly cleared land in settlements.

[29]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.

5.1 Effects of Skill Transferability on Rice Productivity

Panel A of Table 4 reports our main estimates of γ, the semi-elasticity coefficient on agroclimatic similarity in the following regression based on variants of our baseline estimating equation (6):

$$y_j = \alpha + \gamma \mathcal{A}_j + \mathbf{x}_j' \boldsymbol{\beta} + \nu_j, \tag{7}$$

where village-level agroclimatic similarity ($\mathcal{A}_j$) is based on the Java/Bali migrant weights, and $\mathbf{x}_j$ includes island fixed effects as well as the full set of predetermined agroclimatic endowments elaborated in Section 3.1;[30] 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.[31] 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.[32] 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. This baseline estimate is important because rice is an important crop and staple in Indonesia and around the world, expanding rice production was one of the program's main goals, and rice is particularly vulnerable to climate change.

Our quasi-experimental estimate of productivity losses due to agroclimatic dissimilarity complements recent evidence 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. We drop natural advantage controls ($\mathbf{x}_j$) in Column 2 to address the concern that agroclimatic similarity is only picking up spurious correlation with unobserved natural advantages. 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 rice productivity (i.e., all variables reported in Table 3). This addresses concerns that the π_{ij} used to construct the agroclimatic similarity index is correlated with

[30] We report results based on linear controls, but the findings are robust to a nonlinear specification based on indicators for the deciles of each component of agroclimatic similarity. In a slight abuse of notation, the $\mathbf{x}_j$ vector in the regression also includes 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. None are material to the results.

[31] Inference is largely robust to varying the bandwidth up to 500 kilometers or clustering by district boundaries.

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

unobserved determinants of productivity at the origins (and hence absolute advantage). The province-level migrant shares additionally capture variation across the four transit camps in Java/Bali. Column 4 adds predetermined controls at the destinations (i.e., all controls in Table 3) as well as controls for demographic characteristics, 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 (87 in total). The effects change slightly but are not statistically significantly different from column 1. We retain this demanding specification in subsequent tables discussed below.

We can also rule out endogeneity concerns associated with the facts that not all villages and not all individuals produce rice. We deal with the former by running 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.[33] The latter concern is that even after controlling for observed demographic compositions (as we do in columns 4-5), the lower rice productivity in low similarity villages is driven by the selection of unobservably higher ability individuals out of rice farming. In Appendix B.1, we show that the degree of selection needed to explain the productivity effects is quite large, when compared to the estimated effects of agroclimatic similarity on crop choices. Additionally, following Altonji et al. (2005) and Bellows and Miguel (2009), we calculate that selection on unobservables would have to be at least 10 times greater than selection on observables, to explain the 16.6 percent effect on productivity in column 5.[34] We further address aggregation bias by using *Susenas* data for a small sample of Transmigration villages that includes household-level rice productivity and find similar results (see Appendix Table B.3).

Cash Crops as a Placebo Test. In Panel B of Table 4, we show that agroclimatic similarity has a fairly precise zero effect on cash crop productivity. The productivity measure is revenue-weighted across 28 cash crops based on the approach described in Section 3.2. In column 1, the 95 percent confidence interval for the effect of a one standard deviation change in agroclimatic similarity ranges from -0.04 to 0.09 tons per hectare (a narrow range relative to a mean of one ton per hectare and a standard deviation of 2.7). The small effect size holds across the additional specifications in columns 2-5 considered in Panel A.

These null effects for cash crops serve as a placebo check, lending further support to the view that the large rice productivity effects reflect transferability of skills acquired in Java and Bali. Most key cash crops were not grown in Java and Bali during the 1970s and 1980s, as discussed in Section 3.2. Hence, our proxy for skill transferability should have no effect on cash crop productivity given that transmigrants had not acquired these crop-specific skills prior to moving. Incidentally, the null effects provide further evidence that agroclimatic similarity is not merely proxying for unobservable

[33] Appendix Table B.5 shows that a one standard deviation increase in agroclimatic similarity increases the likelihood that the village has any rice production by 8.8 percentage points relative to a mean of 74 percent. However, formal Tobit decompositions (available upon request) suggest that the majority of the rice productivity effects in levels are due to an increase in the intensive margin of productivity (i.e., among villages growing any rice).

[34] Altonji et al. consider an empirical model with a bivariate normal structure while Bellows and Miguel develop the same test for a linear model relaxing the joint normality assumption. We implement this approach by dividing the estimate with the most controls (column 5) by the difference between the estimate with island fixed effects but without controls (column 2) and the estimate with controls. The larger the magnitude of this ratio, the more unlikely that the effect is driven by selection on unobservables. This implementation follows Nunn and Wantchekon (2011), and we find ratios that are similar or larger in magnitude than these three papers. The ratio for the specification in column 5 is 10.93. The ratios for columns 1, 3, and 4, range from 4.87 to 9.17.

land quality, market access or any other determinants of agricultural productivity (common across crops). If this was the case, then agroclimatic similarity should lead to productivity effects for cash crops.

Heterogeneity: Where Does Similarity Matter Most? Having found large average productivity effects for rice, we show in Table 5 that agroclimatic similarity is more important in places with adverse growing conditions. In column 1, we interact agroclimatic similarity with the FAO-GAEZ measure of potential rice productivity.[35] 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.[36]

Column 2 interacts agroclimatic similarity with indicators for three groups of Transmigration villages with low, medium and high shares of wetland as observed in 2002. This reduces an otherwise high-dimensional vector of agroclimatic attributes into a single land quality measure that is informative about variation in cultivation methods and potential productivity and is also uncorrelated with agroclimatic similarity (albeit not predetermined). 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.[37] Second, adaptation to wetland production is relatively easy even for farmers accustomed to dryland methods in Java/Bali. In this context, agroclimatic differences can be easier to overcome given the strong natural advantages of wetland production systems.

The heterogeneous effects across growing conditions are also in line with results in columns 3-6 where we consider individual sub-components of agroclimatic similarity. We report results for a few key agroclimatic characteristics (others are available upon request). Column 3 shows that elevation similarity has null effects on rice productivity, which suggests that our main results are not merely driven by lowland Java/Bali farmers having difficulty adapting to highland growing conditions. However, topographic similarity does matter in that a one standard deviation increase in ruggedness similarity leads to a large 17.4 increase in productivity in column 4. Similarity in soil characteristics such as drainage and carbon content also have large effects on rice productivity in columns 5 and 6, respectively. In sum, although the wetland–dryland breakdown is salient, a range of agroclimatic features—each of which are imperfectly correlated with wetland share—matter for understanding skill transferability.

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 (7):

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

where $g(\cdot)$ is a partially linear function that relates agroclimatic similarity to the outcome y_j using the

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

[36] Although both agroclimatic similarity and potential productivity have large effects on actual rice productivity, having a high quality match is twice as important as being in a high quality place. In a parsimonious specification with only island fixed effects, the elasticity of rice productivity with respect to agroclimatic similarity is two times larger than the elasticity with respect to potential rice productivity.

[37] Donner (1987) succinctly captures this possibility: "The Javanese transmigrants, mostly experienced in growing wet-rice, were promised irrigated land in the new settlements, but found only dryland and had to change to rain-fed cultivation."

approach in Robinson (1988). Figure 4 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. 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 (with high expected return) are particularly costly near subsistence.

The shape of $g(\cdot)$ in Figure 4 also clarifies how our natural experiment provides novel insights into the importance of sorting. In particular, the density for non-Transmigration villages in Panel B of Figure 3 coincides with the flatter region in the semiparametric estimate in Figure 4. This is consistent with spontaneous migrants sorting into destinations where their skills are easily transferable. Without the program-induced skill mismatch, our study would, as in most migration settings, lack the "empirical content" to say anything about the productivity implications of sorting based on comparative advantage.

Finally, the semiparametric estimate provides important policy lessons. 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 small effects. The concave shape suggests that it is most important to avoid very bad matches rather than achieving the best match. Second, 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 dissimilarity seen in the lower tail of Figure 4. We revisit policy questions in Section 6.

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.

5.2 Adaptation and Broader Development

We investigate four adaptation mechanisms: learning, switching occupations, crop choice, and ex post migration. We find relatively more support for learning and crop adjustments, especially switching to cash crops. Finally, while we find some adaptation response, there remain sizable differentials between high and low similarity villages in nighttime light intensity, a proxy for local income used in several recent studies. This suggests that adjustments were costly and perhaps incomplete.

Learning. An extensive literature documents the importance of learning in the agricultural context (see Foster and Rosenzweig, 2010). This rich literature provides evidence of several types of learning models. 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, BenYishay and Mobarak (2014) find that farmers are most persuaded by information provided by other farmers who face comparable agricultural conditions.

In Table 6, we provide additional evidence on learning mechanisms within Transmigration villages. In column 2, we augment the specification in column 5 of Table 4 (reproduced here in column 1) with three measures of variation in ethnic and origin district composition within Transmigration villages: the ethnic fractionalization index across the eight transmigrant ethnicities, the Herfindahl index (HI) for origin district population shares, and the number of origin districts. Each of these measures of transmigrant diversity could have direct effects on rice productivity that are independent of or confounded with the effects of agroclimatic similarity. However, we find no economically or statistically meaningful impacts. Moreover, the effect of agroclimatic similarity is unchanged from the original specification in column 1. This suggests that the main productivity effects are driven by the agroclimatic match. This provides further evidence of the local learning channel. Although other measures of village-level agroclimatic similarity—median, 75th percentile, and share with maximum $\mathcal{A}_{ij}$ match—also matter for rice productivity, it is difficult to distinguish these other effects from the average with which they are highly correlated. Indeed, when included separately alongside average similarity, each measure has a small and insignificant effect while the average retains its overall significance.

While data constraints make it difficult to isolate social learning mechanisms among transmigrants, we are able to provide evidence consistent with social learning from natives. Column 3 of Table 6 shows that a one standard deviation increase in linguistic similarity increases rice productivity by 25 percent. As discussed in Section 3.1, our linguistic similarity index in equation (2) measures the structural proximity between languages native to Java/Bali and languages native to the Outer Islands. Column 4 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. We assume a larger scope for learning from natives in places with a smaller transmigrant stock in the initial year[38] and find that linguistic similarity is indeed more important in these villages. Overall, these results echo case studies of Transmigration settlements that discuss the importance of learning from natives (e.g., Donner, 1987).

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

[38]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 note that agroclimatic similarity is uncorrelated with the size of the initial transmigrant population, conditional on $\mathbf{x}_j$.

into trading and services.

We test these predictions in Table 7 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.[39] Columns 1-3 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 the sample 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 (p.p.) higher probability of an individual reporting farming as their primary occupation. Meanwhile, a one standard deviation increase in linguistic similarity is associated with 1.8 p.p. higher probability of trading/services. However, the effects are quantitatively small. For example, the 0.9 p.p. effect for agroclimatic similarity 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 that the young transmigrants are relatively more adaptable but there are no statistically significant differences in the patterns of occupational choices across generations. 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 explore this possibility using two complementary approaches.

First, in Table 8, we show that agroclimatic similarity has null effects on labor allocation albeit qualitatively significant effects on relative revenue across crops. Column 1 indicates that agroclimatic similarity has a fairly precise null effect on the share of farmers whose primary occupation is growing cash crops (according to the 2000 Census).[40] Column 2 shows that a one standard deviation increase in agroclimatic similarity leads to a 4.7 p.p. increase in the share of rice in total agricultural revenue based on the measure described in Section 3.2. Column 3 meanwhile shows the opposite with a 3.9 p.p. decline in the revenue share of cash crops (relative to a mean of 60 percent). In an attempt to summarize across crops, we show in column 4 that agroclimatic similarity has a small and statistically insignificant effect on revenue-weighted average agricultural productivity across all crops. This is not surprising given that cash crops have a substantially higher *potential* revenue weight than rice and agroclimatic similarity

[39] The individual-level controls include gender, married, years of schooling, residence five years ago (Java/Bali, other Outer Islands province or district), an indicator for belonging to a native Java/Bali ethnic group, and indicators for religion. All but the last are interacted with age. The village-level controls are the same as those used in column 1 of Table 4.

[40] This null results also holds at the individual-level in regressions as in Table 7.

has no effect on cash crop productivity. Indeed, a simple decomposition exercise suggests that the null productivity effect of agroclimatic similarity on cash crops (see Panel B of Table 4) with a high revenue weight of 0.6 offsets the large productivity effect on rice with a lower revenue weight of 0.27 and can explain the null result in column 4.[41]

However, there are at least three reasons why the relative weights between rice and cash crops could be smaller. First, 65 percent of farmers grow food (and primarily rice) crops, suggesting that employment shares may be the more relevant weighting factors, but unfortunately, we do not have employment shares by individual crops. Second, there are large differences in fixed and variable input costs of production across rice and cash crops. In turn, these likely imply smaller differences between cash and food crops in annual profits, which would arguably be the most ideal weights in terms of capturing welfare. Finally, due to data constraints, the revenue weights are based on national prices, which may understate the importance of non-export crops like rice in local agricultural income. Also, a brief liberalization of rice imports in the early 2000s temporarily pushed down the relative price ratio of rice to cash crops.

In Table 9, we provide a second piece of evidence on the crop adjustment mechanism. Adapting an approach developed by Michalopoulos (2012), we identify the extent to which transmigrants bring their preferences for growing rice with them to the Outer Islands. In particular, 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.

While $\rho_1 > 0$ across all specifications in Table 9, 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

[41]The decomposition follows from the product rule. Total productivity is calculated as $\omega_R \ln y_R + \omega_C \ln y_C + \mathcal{O}$ where ω_R and $\ln y_R$ are the revenue weights and log productivity for rice, ω_C and $\ln y_C$ are the analogues for cash crops, and $\mathcal{O}$ is the weighted average for all other (food) crops. The effect of a one standard deviation increase in agroclimatic similarity ($\mathcal{A}$) on total productivity is then the sum of the effects for each crop:

$$\frac{\mathrm{d}\omega_R}{\mathrm{d}\mathcal{A}} \ln y_R + \omega_R \frac{\mathrm{d} \ln y_R}{\mathrm{d}\mathcal{A}} + \frac{\mathrm{d}\omega_C}{\mathrm{d}\mathcal{A}} \ln y_C + \omega_C \frac{\mathrm{d} \ln y_C}{\mathrm{d}\mathcal{A}} + \frac{\mathrm{d}\mathcal{O}}{\mathrm{d}\mathcal{A}}.$$

We use estimated effects of agroclimatic similarity on revenue weights and productivity and evaluate this equation using revenue weights and log productivity for the average village. The key is that the revenue weights (calculated using national prices) are low for rice and high for cash crops, so that the large productivity effect for rice (0.2) is weighted down to 0.05 (0.27×0.2) and the small productivity effect for cash crops (0.024) is now relatively higher at 0.015 (0.602×0.024).

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.

(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.[42] 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. First, transmigrants are not as mobile as 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 were mostly landless agricultural laborers who were given land (without property rights for the first 5-10 years), which may have played a role in tying them to the Transmigration villages. 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, which could also explain why we did not see large return migrant flows in the early years.

We confirm that selective out-migration is indeed low. First and foremost, as detailed in Section 5.3, a quasi-gravity regression shows that longer-term ex post sorting patterns are uncorrelated with agroclimatic similarity. Moreover, agroclimatic similarity is uncorrelated with population size and the Java/Bali-born migrant share in Transmigration villages in 2000 (results available upon request). Second, 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.

Light Intensity as a Proxy for Income. The preceding discussion shows that learning and crop adjustments appear to be the more important adaptation mechanisms with less evidence of occupational adjustments and ex post migration. Having identified strong effects of agroclimatic similarity on rice productivity, it is important to ask whether these adaptation responses can undo the effects of dissimilarity over time. We investigate whether skill transferability has persistent effects on overall development by using nighttime light intensity in 2010 as the best available proxy for local income.

Table 10 reports statistically and economically significant positive effects of agroclimatic similarity on light intensity. Although transmigrant farmers adapted in several ways to low quality matches, the results in columns 1–4 imply that such adaptation was costly and may be incomplete. Using the baseline specification from Table 4, column 1 shows that a one standard deviation increase in agroclimatic similarity leads to a 1.6 percentage point increase in the share of the village that has any nighttime light coverage relative to a base of a 8.1 percent. This estimate substantially increases when including the

[42] If return migration was greater in dissimilar villages and return migrants were more unproductive (which is why they returned), 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).

full set of additional controls in column 2. This is perhaps because agroclimatic similarity is negatively correlated with district-level manufacturing intensity and electrification (i.e., in other villages) before the Transmigration villages were established (see Table 3). Not controlling for these variables, which are mechanically positively correlated with luminosity, biases us against finding positive effects on light intensity. To be conservative, we report the smaller estimates. Note that in all columns, we include year of settlement fixed effects in order to fix initial conditions and assign a growth interpretation to the estimates.

Beyond the extensive margin, higher agroclimatic similarity also leads to growth in the intensive margin of light intensity. Column 3 suggests that a one standard deviation increase in similarity translates to 3.6 percent greater light intensity. Again, the effect increases substantially in column 4 to 10.4 percent when adding the full set of controls. Given that only 24 percent of Transmigration villages had any light coverage by 2010, the results in columns 1-2 rule out concerns that our estimates using the conventional measure of intensity are driven by nonlinearities stemming from the zeros.[43] Using an approach similar to Henderson et al. (2012), Olivia and Gibson (2013) estimate that a one percent increase in light intensity 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 a one standard deviation increase in agroclimatic similarity increases medium-run village-level income by 1.8 to 5.2 percent. These are economically meaningful effects given the low average light intensity in the full sample and suggest that agroclimatic similarity has persistent effects on the broadest possible measure of local development.

5.3 Robustness Checks and Other Threats to Identification

We address additional concerns about identification here. First, we provide additional support for the plausible exogeneity of agroclimatic similarity. Second, we demonstrate the robustness of our key rice productivity results to aggregation bias, alternative specifications of the agroclimatic similarity index, and confounding program features.

Correlation with Schooling. We begin by addressing the 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 in Figure 5, 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.[44]

Gravity Test for Sorting. Next, we use a quasi-gravity specification to show that transmigrants did not

[43] The results are also robust to other approaches for dealing with the zeros and censoring considered in the literature such as binary outcome or Tobit specifications (Hodler and Raschky, forthcoming; Michalopoulos and Papaioannou, 2014).

[44] Taking a more parametric approach, Appendix Table B.2 shows that village-level agroclimatic similarity is uncorrelated with schooling (and other demographic characteristics) among Java/Bali immigrants after conditioning on $\mathbf{x}_j$. Additionally, the lack of correlation at the individual-level is robust to controlling for other demographic characteristics including age, gender, migration, and marital status.

endogenously sort into (out of) those sites in which they (do not) have high agroclimatic similarity. 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 11, we use OLS to estimate variants of the following equation

$$f(migrants_{ij}) = \alpha + \lambda_a \mathcal{A}_{ij} - \lambda_d \ln distance_{ij} + \mathbf{z}_j' \boldsymbol{\zeta} + \tau_i + \upsilon_{ij}, \tag{8}$$

where τ_i are origin fixed effects (FE), and $\mathbf{z}_j$ 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). Note that this specification is akin to regressing the migrant shares π_{ij} on $\mathcal{A}_{ij}$, origin i FE, and destination village j FE. In Table 11, we parameterize the destination FE using the $\mathbf{x}_j$ as in our village-level regressions in Table 4, but results are unchanged when instead including village FE.

In all specifications of equation (8), 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 into (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 2.2), which we account for directly in our main results by controlling for (weighted) distance. 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 equation (8) at the $j\ell$ level for the ethnolinguistic groups ℓ indigenous to Java/Bali.

Further Robustness Checks. 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 (rows 2 to 4). We further address aggregation bias by controlling for the share of natives and overall population density (row 5). We control for location using polynomials of latitude and longitude (row 6). We also consider alternative definitions of our agroclimatic similarity index based on different distance metrics and migrant weights (rows 7 to 10). None of these changes affects our key finding of a statistically and economically significant effect of agroclimatic similarity on rice productivity.

6 Impact of the Transmigration Program: Policy Exercises

In this section, we conduct two policy exercises . First, we use simulations to show that a reallocation of transmigrants to maximize agroclimatic similarity could have large aggregate effects on rice productivity. Second, we use a policy discontinuity and place-based evaluation approach to provide the first causal estimates of the average impact of the Transmigration program on local economic development. Ultimately, we argue that the persistent effects of agroclimatic similarity may explain the limited average impact of the program on local development.

These two exercises demonstrate the aggregate implications of origin-by-destination match quality for program effectiveness. Despite the growing policy relevance of resettlement, there remains a dearth of causal evidence on the medium- to long-run impacts of resettlement programs, especially in developing countries (IPCC, 2014). The findings below fill that gap.

6.1 Optimal Reallocation of Migrants

One of the original goals of the Transmigration program was to reallocate labor in order to narrow the agricultural and especially rice output gap between the Inner and Outer Islands. We attempt here to quantify the aggregate output losses from the poor matching of transmigrants' farming skills to local growing conditions. We use the baseline rice productivity results in column 1 of Table 4 and reassign transmigrants to destinations to maximize agroclimatic similarity, and hence, rice output. As discussed 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 in terms of its complexity (Fischer et al., 1986). However, we can approximate the optimal solution using a greedy assignment algorithm, in which similarity is sequentially maximized, village-by-village.

Using this algorithm, we find that aggregate rice production could have been 27 percent higher if individuals had been assigned in a more optimal manner. While this may not be a global optimum, the solution is computationally feasible and represents an approach to the problem that could be carried out by future resettlement planners. Indeed, this type of agroclimatic assignment mechanism would address an important challenge recognized in the World Bank's Operational Policy (4.12) on resettlement, namely that "people are relocated to environments where their productive skills may be less applicable and the competition for resources greater."

6.2 Average Treatment Effects

As mentioned in Section 2.2, 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.[45] 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

[45] 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.

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).

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 (7). 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).

In Table 12, column 1 compares 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 and cluster standard errors at the district level. Sample sizes vary across outcomes and columns (depending on data availability) but include as many as 31,185 villages in column 1, and 832 treated villages and 668 controls in columns 2-4.[46] As detailed in Appendix B.2, 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 (0.77 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 (available upon request). 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.

[46]We exclude control villages that are within 10 km of Transmigration settlements to minimize bias from spillovers.

We find similar null results for both revenue-weighted average yield across all crops as well as light coverage and intensity in 2010.

Our results on the persistent consequences of origin-by-destination mismatch and incomplete 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 publicly funded 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 development impacts we find in Table 12.

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 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 occupational adjustment or ex post migration. We find null effects on cash crop productivity and total agricultural productivity. 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 nighttime light intensity point to costly and perhaps incomplete adjustment over the medium-run period in this study. We relate these results to recent work on location-specific crop and staple consumption preferences as well as related research on farmers' adaptation responses to agroclimatic changes.

Our findings shed new light on the importance of comparative advantage in shaping the spatial distribution of aggregate productivity. A growing literature argues that labor is misallocated across locations (e.g., Munshi and Rosenzweig, 2014), sectors (e.g., Gollin et al., 2014), and occupations (e.g., Hsieh et al., 2013). Our natural experiment suggests that some of these productivity gaps may be explained by barriers to transferring skills and ultimately adjusting to new economic environments. Our focus on rural-to-rural migration is important, given that rural-to-rural flows are 1.5 to 2 times greeter than rural-to-urban flows (Young, 2013). Quantifying the welfare costs of these barriers is an important task for future research, especially in light of climate change, as discussed in the Introduction.[47]

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 treat-

[47] The estimated number of individuals displaced due to extreme weather events appear to be large, with costs borne disproportionately by vulnerable groups in developing countries, such as rain-fed, subsistence farmers. Climate change is expected to affect crop yields through its effects on the temperature, rainfall (and the associated soil hydrology), and pest ecology. Climate scenarios suggest that some of these changes could be abrupt, leaving limited time for farmers to experiment and adjust (Gollin, 2011; IPCC, 2014).

ment 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 suggest that complementary government inputs are crucial to help farmers mitigate the effects of dissimilarity. Although negatively selected from their rural cohort at the time, the government-sponsored migrants 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.

References

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

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Ashraf, Quamrul and Oded Galor, "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *The American Economic Review*, 2013, *103* (1), 1–46.

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

Au, Chun-Chung and J. Vernon 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, Samuel and Christopher Blattman, "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, forthcoming.

Beaman, Lori 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.

Bellows, John and Edward Miguel, "War and Local Collective Action in Sierra Leone," *Journal of Public Economics*, 2009, *93* (11), 1144–1157.

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

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

Busso, Matias, Jesse Gregory, and Patrick 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.

Cernea, M. M., *The economics of involuntary resettlement: Questions and challenges*, World Bank, 1999.

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

__ , __ , and __ , "The Identification of Agglomeration Economies," *Journal of Economic Geography*, 2011, *11*, 253–266.

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

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

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

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.

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

Dove, Michael R, "The Agroecological Mythology of the Javanese and the Political Economy of Indonesia," *Indonesia*, 1985, (39), 1–36.

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, Joan, Laura Mayoral, and Debraj 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 Schónberg, "How General Is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 2010, *28* (1), 1–49.

Geertz, Clifford, *Agricultural Involution: The Processes of Ecological Change in Indonesia*, Vol. 11, Berkeley, CA: University of California Press, 1963.

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, "Climate Change and Technological Innovation in Agriculture: Adaptation through Science," in Ariel Dinar and Robert O. Mendelsohn, eds., *Handbook on Climate Change and Agriculture*, Edward Elgar Publishing, January 2011, pp. 382–401.

__ , 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.

Hall, Anthony, "Making People Matter: Development and the Environment in Brazilian Amazonia," *International Journal of Contemporary Sociology*, 1993, *30* (1), 63–80.

Hardjono, Joan, "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, Chang-Tai, Erik Hurst, Jones I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," *NBER Working Paper No. w18693*, 2013.

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, "Summary for policymakers," in C.B. Field, V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White, eds., *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, UK; New York, NY, USA: Cambridge University Press, 2014, pp. 1–32.

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

Kline, Patrick M. and Enrico 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.

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

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

Levang, Patrice, "Tanah sabrang (la terre d'en face) la transmigration en indonesie: permanence d'une politique agraire contrainte." PhD dissertation, Ensa Montpellier 1995.

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

Lucas, Robert 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.

McCulloch, Neil and C. Peter Timmer, "Rice Policy in Indonesia: A Special Issue," *Bulletin of Indonesian Economic Studies*, April 2008, *44* (1), 33–44.

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

— **and Elias Papaioannou**, "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, 2014, *129* (1), 151–213.

Mohanty, S., R. Wassmann, A. Nelson, P. Moya, and S.V.K. Jagadish, "Rice and Climate Change: Significance for Food Security and Vulnerability," *International Rice Research Institute*, 2013.

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

— **and Jaqueline Oliveira**, "Migration, roads and labor market integration: Evidence from a planned capital city," *Unpublished Manuscript*, 2014.

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*, 2014.

Nguyen, N. V., "Global Climate Changes and Rice Food Security," *International Rice Commission, FAO, Rome, Italy*, 2002.

Nunn, Nathan and Leonard Wantchekon, "Leonard," The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 2011, *101* (7), 3212–3252.

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.

Peng, S., K. T. Ingram, H. U. Neue, and L. H. Ziska, *Climate Change and Rice*, Springer, 1995.

Putterman, Louis and David N. Weil, "Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality," *The Quarterly Journal of Economics*, 2010, *125* (4), 1627–1682.

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.

— , *Involuntary Resettlement Sourcebook, Planning and Implementation in Development Projects*, International Bank for Reconstruction and Development, 2004.

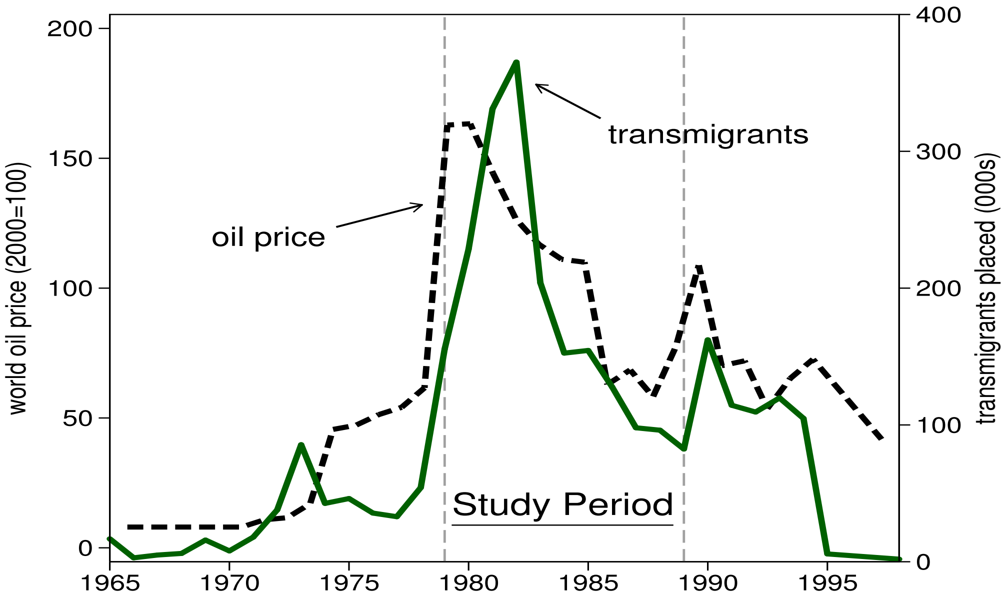
— , *Doing Business 2009: Country Profile for Indonesia*, Washington, DC: World Bank, 2008.

— , *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.

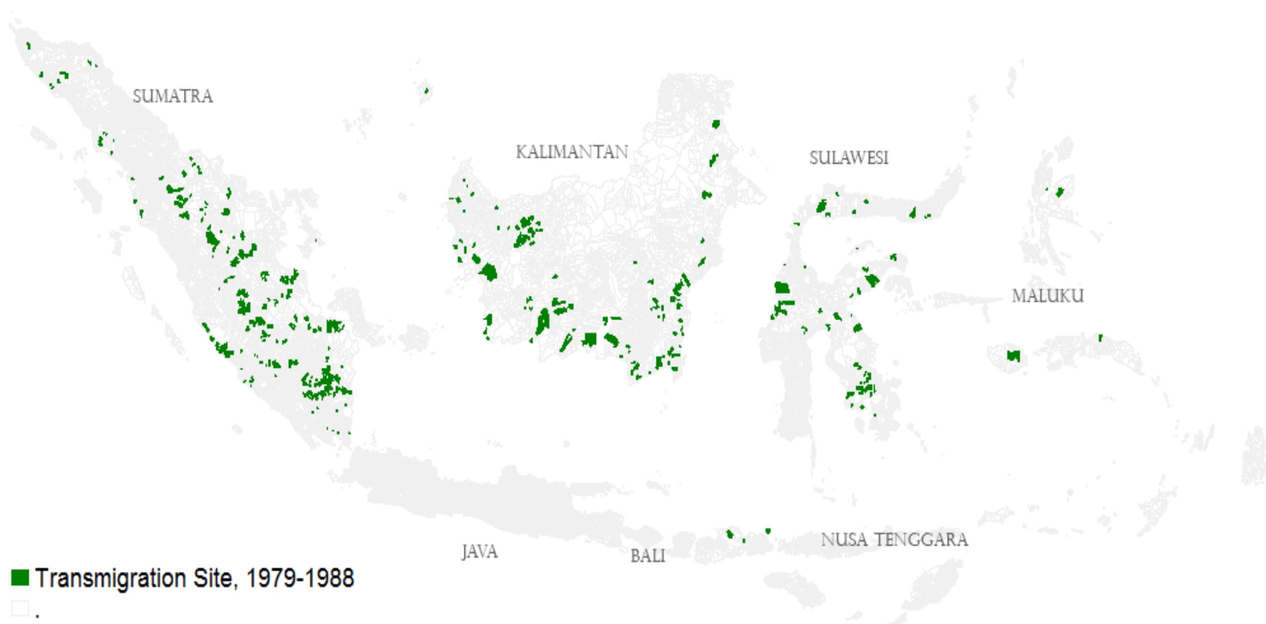
Figures

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.

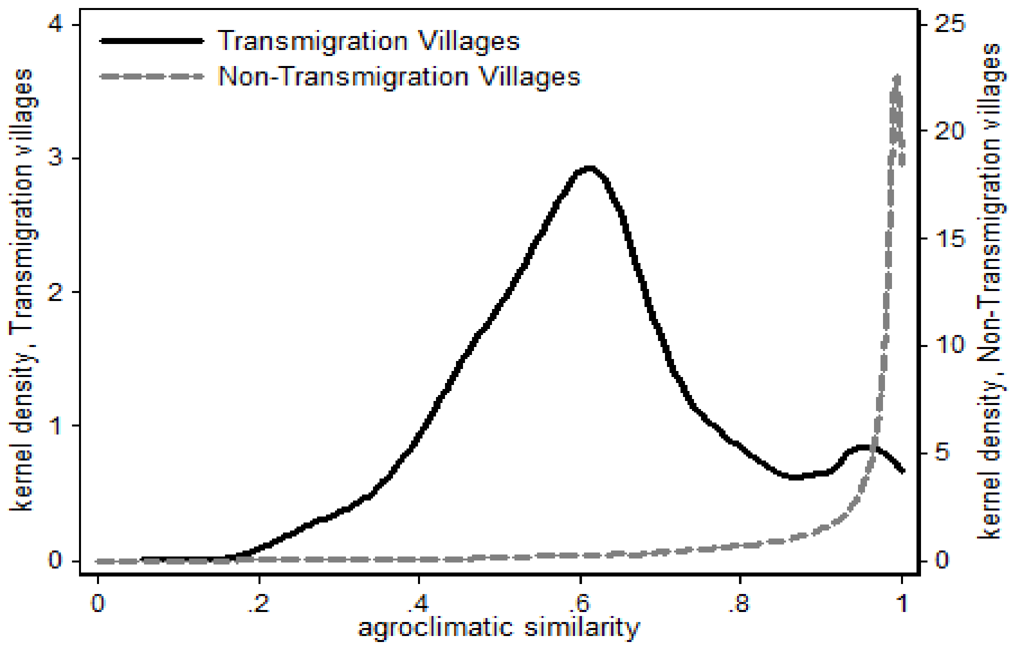
Figure 2: Map of Transmigration Villages



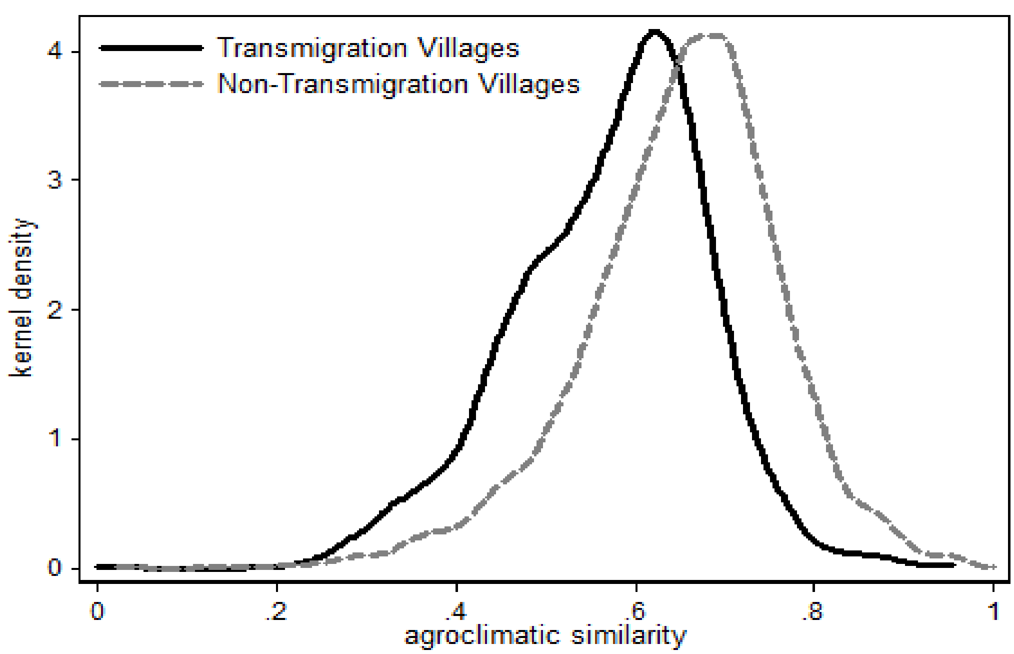
Notes: The figure shows all Transmigration villages settled in 1979–1988 based on our digitization and mapping of the Transmigration Villages in the 1998 MOT Census.

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

(a) All Individuals (Natives and Immigrants)

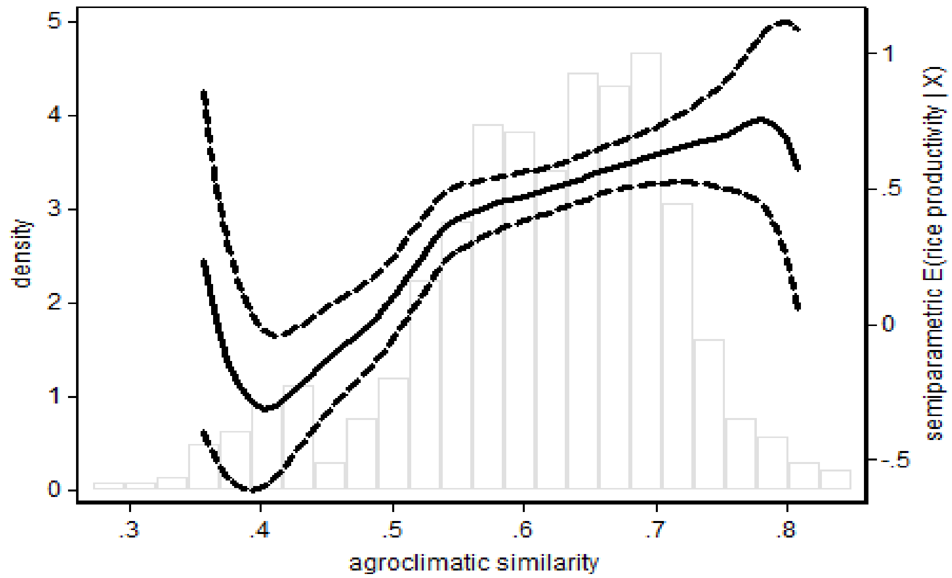


(b) All Individuals (Immigrants Only)



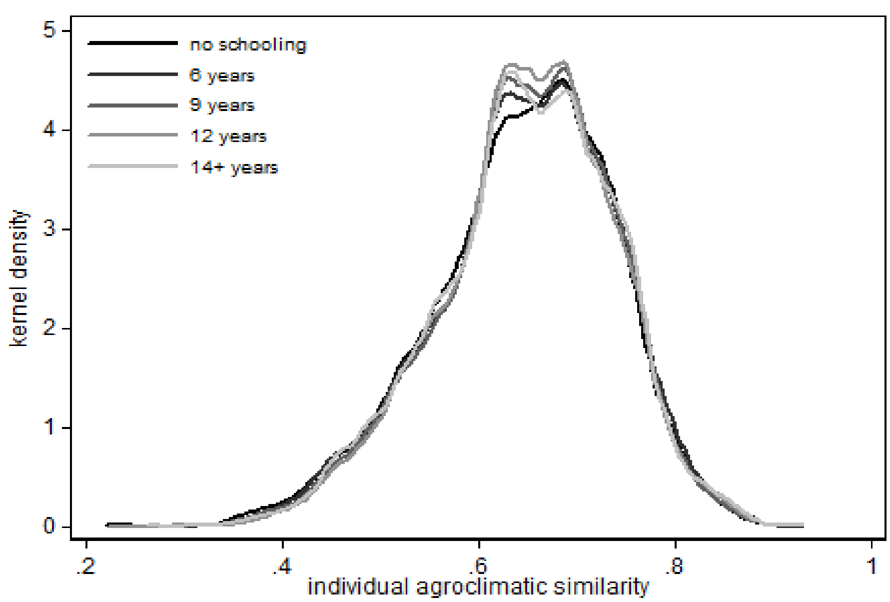
Notes: Panel A shows kernel densities of village-level agroclimatic similarity computed over all individuals—natives and immigrants—in the village separately for Transmigration settlements and all other Outer Islands villages. Panel B shows kernel densities of village-level agroclimatic similarity computed 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: Baseline Specification: Semiparametric Evidence for Rice Productivity



Notes: This is based on semiparametric Robinson (1988) extensions of the parametric specification in column 1 of Table 4 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 for presentational purposes.

Figure 5: 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.

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 (tons/ha)	1.00	(2.65)	770
cash crop productivity (tons/ha)	1.02	(3.03)	712
log light intensity, 2010	0.24	(0.59)	814
village area with any lights, 2010	0.08	(0.22)	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 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. Cash crop and total agricultural productivity are each 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. The number of villages differs for rice and total agricultural productivity as a result of missing or zero production of the given crops. See Appendix A for details on data sources and construction.

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 district-level population summary statistics, 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.

Table 4: Effects of Agroclimatic Similarity on Rice Productivity

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	Panel A: **Rice Productivity**				
agroclimatic similarity	0.204	0.182	0.210	0.151	0.166
	(0.064)***	(0.045)***	(0.075)***	(0.057)***	(0.068)**
Number of Villages	600	600	600	600	600
R^2	0.149	0.032	0.178	0.281	0.318
	Panel B: **Cash Crop Productivity (Placebo Test)**				
agroclimatic similarity	0.024	-0.007	0.039	-0.006	-0.021
	(0.031)	(0.014)	(0.048)	(0.096)	(0.071)
Number of Villages	712	712	712	712	712
R^2	0.054	0.008	0.095	0.126	0.164
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{x}_j$)	Yes	No	Yes	Yes	Yes
Origin Province Migrant Shares	No	No	Yes	No	Yes
Log Weighted Avg. Distance to Origins	No	No	Yes	No	Yes
Weighted Avg. Predetermined Controls (Table 3), Origins	No	No	Yes	No	Yes
Predetermined Controls (Table 3), Destinations	No	No	No	Yes	Yes
Predetermined Demographics and Schooling	No	No	No	Yes	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. The dependent variable in panel A is log rice output per hectare with a mean of 2.5 tons/ha, and in Panel B is the revenue-weighted log cash crop productivity with a mean of 1.0 tons/ha. The latter is calculated using crop-specific revenue-weights for 28 cash crops, primary among which are palm oil, rubber, cocoa, coffee, and groundnuts (see Appendix A). Agroclimatic similarity is normalized to have mean zero and a standard deviation of one. All regressions include island fixed effects and except in column 2 also include 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 Weighted Avg. Distance to Origins" is 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. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

Table 5: Heterogeneous Effects of Agroclimatic Similarity on Rice Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
agroclimatic similarity	0.424					
	(0.112)***					
$\cdots \times$ log potential rice yield	-0.536					
	(0.175)***					
$\cdots \times$ tercile 1 wetland share $\in [0, 0.16]$		0.355				
		(0.079)***				
$\cdots \times$ tercile 2 wetland share $\in (0.16, 0.66]$		0.141				
		(0.059)**				
$\cdots \times$ tercile 3 wetland share $\in (0.66, 1.0]$		0.059				
		(0.120)				
elevation similarity			0.017			
			(0.040)			
ruggedness similarity				0.174		
				(0.070)**		
soil drainage similarity					0.177	
					(0.086)**	
soil carbon similarity						0.114
						(0.067)*
Number of Villages	599	600	600	600	600	600
R^2	0.327	0.340	0.313	0.319	0.319	0.315
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{x}_j$)	Yes	Yes	Yes	Yes	Yes	Yes
Added Controls in Column 5 of Table 4	Yes	Yes	Yes	Yes	Yes	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. The dependent variable in all specifications is log rice output per hectare. All similarity regressors are normalized to have mean zero and a standard deviation of one. Log potential rice productivity is based on the FAO-GAEZ measure described in the text. We lose one observation relative to baseline after taking logs. Retaining this village and using potential productivity in levels or adding a small constant inside the logarithm does not affect the results. "Wetland share" denotes the fraction of agricultural land that is wetland in 2002. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

Table 6: Agroclimatic Similarity, Social Learning, and Rice Productivity

	(1)	**(2)**	**(3)**	**(4)**
agroclimatic similarity	0.166	0.156	0.150	0.146
	(0.068)**	(0.064)**	(0.061)**	(0.061)**
within-Java/Bali ethnic fractionalization		-0.032		
		(0.053)		
Herfindahl Index, Java/Bali origin district shares		0.039		
		(0.061)		
number of Java/Bali origin districts		-0.017		
		(0.068)		
linguistic similarity			0.258	0.214
			(0.088)***	(0.099)**
$\cdots \times$ small initial cohort				0.084
				(0.036)**
Number of Villages	600	600	600	600
R^2	0.318	0.320	0.330	0.325
Island Fixed Effects	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{x}_j$)	Yes	Yes	Yes	Yes
Added Controls in Column 5 of Table 4	Yes	Yes	Yes	Yes

Notes: */**/*** denotes significance at the 10/5/1 percent level. The dependent variable in all specifications is log rice output per hectare. All continuous regressors are normalized to have mean zero and a standard deviation of one. "within-Java/Bali ethnic fractionalization" equals $1 - \sum_{e=1}^{8} (N_{ej}/N_j)^2$ where N_{ej} is the number of individuals in 2000 from transmigrant ethnic group e, and N_j is the total transmigrant ethnic population in village j. The Herfindahl index equals $\sum_{i=1}^{I} \left(N_{ij}/N_j^d\right)^2$ where N_{ij} is the number of Java/Bali-born migrants from district i and N_j is the number of Java/Bali-born migrants. "Small initial cohort" in column 4 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 7: Occupational Sorting within Transmigration Villages

Dependent Variable	$\mathbf{Pr(Occupation = \ldots)}$ **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
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 farming (columns 1-3) or trading/services (columns 4-6). Columns 1 and 4 include all Java/Bali-born individuals between the ages of 15 and 65. 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 8: Agroclimatic Similarity and Crop Adjustments

Dependent Variable:	share of cash crop farmers	revenue weight on		total agric. productivity
		rice	cash crops	
	(1)	(2)	(3)	(4)
agroclimatic similarity	0.001	0.047	-0.039	0.014
	(0.022)	(0.017)***	(0.022)*	(0.079)
Number of Villages	770	770	770	770
R^2	0.448	0.410	0.360	0.187
Dep. Var. Mean (Levels)	0.348	0.273	0.602	0.996
Island Fixed Effects	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{x}_j$)	Yes	Yes	Yes	Yes
Added Controls in Column 5 of Table 4	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 controls are as in Column 5 of Table 4. The sample of villages is restricted to those with agricultural output data in *Podes* 2002. The dependent variable in column 1 is the share of farmers whose primary occupation is farming cash crops in the 2000 Population Census, which has separate occupational entries for food and cash crop farming. The dependent variable in columns 2 (3) is the share of rice (cash crops) based on the approach described in Section 3.2. The dependent variable in column 4 is the measure of revenue-weighted agricultural productivity building on that same approach and normalizing the mean tons/ha to be one across all crops for comparability (results are similar without weighting). Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

Table 9: Against the Grain: Neighborhood vs. Origin Effects in Rice Land Allocation

Dependent Variable	$Rice/Staples$		$\mathbf{Pr}(Rice/Staples > 0.5)$	
	(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 10: Agroclimatic Similarity and Nighttime Lights in 2010

	coverage		intensity	
	(1)	**(2)**	**(3)**	**(4)**
agroclimatic similarity	0.016	0.043	0.036	0.104
	(0.007)**	(0.008)***	(0.017)**	(0.025)***
Number of Villages	814	814	814	814
R^2	0.125	0.253	0.137	0.283
Dep. Var. Mean (Levels)	0.081	0.081	0.242	0.242
Island Fixed Effects	Yes	Yes	Yes	Yes
Predetermined Village Controls ($\mathbf{x}_j$)	Yes	Yes	Yes	Yes
Added Controls in Column 5 of Table 4	No	Yes	No	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 controls are as in Table 4 with the addition of indicators for the year the village was established. The dependent variables are the two measures of nighttime lights capturing, respectively, the fraction of the village with any light coverage and log(1 + light intensity) in 2010. Standard errors in parentheses allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999).

Table 11: Quasi-Gravity Regression of Migration from Java/Bali to the Outer Islands

Dependent Variable:	$\mathbf{Pr}(migrants_{ij} > 0)$		$\ln(migrants_{ij})$	
	(1)	**(2)**	**(3)**	**(4)**
agroclimatic similarity	0.0027	0.0015	-0.0004	0.0001
	(0.0066)	(0.0069)	(0.0200)	(0.0220)
$(-1)\times$ log distance	0.1262	0.1272	0.1287	0.2036
	(0.0192)***	(0.0238)***	(0.0597)**	(0.0753)***
Observations	96,866	96,866	37,446	37,446
Dep. Var. Mean (Levels)	.39	.39	16.8	16.8
Birth District (Java/Bali) Fixed Effects	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes
Year of Settlement Fixed Effects	Yes	Yes	Yes	Yes
Individuals Placed in Year of Settlement	Yes	Yes	Yes	Yes
Predetermined 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. Results are similar using destination district or village fixed effects. Standard errors are two-way clustered by birth district and destination village.

Table 12: 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, 2010	-0.500	0.009	0.009	0.001
	(0.082)***	(0.052)	(0.099)	(0.075)
percent any light coverage, 2010	-0.187	0.008	0.018	0.009
	(0.030)***	(0.017)	(0.033)	(0.025)
Treatment/Control Only	No	Yes	Yes	Yes
Geographic Controls	No	Yes	Yes	Yes
Reweighting	No	No	Yes	Yes
Blinder-Oaxaca	No	No	No	Yes

Notes: */**/*** 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 the 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 columns 2-4. Standard errors are clustered by district in parentheses and are estimated using a block bootstrap in column 3 to account for the generated $\widehat{\kappa}$ weights.