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

April 9, 2014

Abstract

Between 1979 and 1988, the Transmigration Program in Indonesia relocated two million voluntary migrants from the Inner Islands of Java and Bali to the Outer Islands. We use this large-scale policy experiment to identify the importance of skill transferability for economic development. We proxy for migrants' skill transferability using two novel measures of economic proximity that capture the agroclimatic and linguistic similarity between migrants' origins and destinations. The plausibly exogenous assignment of migrants to destinations allows us to provide the first causal estimates of the impact of migrants' economic proximity on local development outcomes in receiving areas. We show that this quasi-experimental variation provides a novel solution to identification problems in multi-sector Roy models. Transmigration villages exhibit significantly higher agricultural productivity if they were assigned migrants from regions of Java/Bali with more similar agroclimatic endowments and indigenous languages. Limited adaptation—in terms of occupational sorting, crop choices, and ex post migration—may explain why initial origin-by-destination match quality matters over the long-run. We also use a model to show that economic proximity proxies for migrants' adjustment costs and hence is a measurable source of comparative advantage. Overall, our findings have implications for the design of resettlement programs and debates about adaptation to agroclimatic change. Using an exogenous budgetary shock and a place-based evaluation strategy, we find that poor matching of migrants to destinations may explain the absence of significant average treatment effects of this large resettlement program on local economic development.

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

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

[*]Acknowledgement: We thank Gilles Duranton, Gordon Hanson, Dilip Mookherjee, Ben Olken, and seminar participants at Boston University, University of Michigan, and the Federal Reserve Board for helpful suggestions. We thank Peter Timmer

1 Introduction

Skill transferability is central to understanding how individuals sort across firms, occupations, industries, and locations. These choices, in turn, have implications for the aggregate distribution of productivity. We study the transferability of skills across locations within agriculture, a sector which employs 1.3 billion people globally (World Bank, 2008b) and is at the core of ongoing debates about world income inequality.[1] We do so in the context of a large-scale, rural-to-rural resettlement program in Indonesia known as Transmigration.[2] Between 1979 and 1988, this seven-billion-dollar program relocated two million voluntary migrants (transmigrants) from the Inner Islands of Java and Bali to the Outer Islands. We examine skill transferability using novel measures of economic proximity that capture the similarity between transmigrants' origins and destinations in terms of agroclimatic and linguistic characteristics. The plausibly exogenous assignment of migrants across destinations allows us to provide the first causal estimates of the impact of migrants' economic proximity on local development in receiving areas.

Our quasi-experimental variation in spatial labor allocation provides a novel solution to an "extremely hard" identification problem in multi-sector Roy models (Combes et al., 2011). This problem was first highlighted by Heckman and Honore (1990) and spans multiple fields in economics.[3] Consider a Roy (1951) assignment model that matches farmers to farms. With two types of farms (e.g., Lowlands and Highlands) 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. We want to estimate the effects of economic proximity on productivity. However, this poses several identification challenges. First, we show in Section 3 that under some assumptions, if farmers sort based on comparative advantage, the econometrician would only observe two of the four outcomes. Second, omitted variables that influence both location choice and productivity may be correlated with economic proximity. The exogenous assignment of transmigrants addresses both problems.

We develop a model to explain why economic proximity is positively related to both the transferability of migrants' skills as well as their comparative advantage at the destinations. A farmer's human capital is specific to characteristics of her birth location. Farmers can transfer this human capital more successfully if destinations are more similar to their birth locations (transferability increases with economic proximity). Therefore, conditional on observably identical destinations, migrants from similar origins have greater comparative advantage relative to migrants from dissimilar origins.

When we aggregate our individual-level model to the village level, we show how to obtain consistent estimates of the impact of economic proximity on local development. Our main village-level dataset combines newly digitized data from a 1998 census of Transmigration villages, with outcome variables from Indonesia's 2003 Village Potential survey and individual data from the 2000 Population crop-specific revenue weights), and light intensity growth.[4]

Our empirical analysis compares Outer Island 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. The identifying assumption is that our proximity measures are uncorrelated with unobservable determinants of productivity. We show that the spatial allocation of transmigrants is uncorrelated with observable determinants of productivity, and does not follow the typical gravity patterns found in settings where migrants selectively choose destinations.

We estimate a parameter akin to an "assimilation elasticity" that relates economic proximity to migrants' transferability of skills, similar in spirit to the trade elasticity that relates physical proximity to the transportability of goods. We are among the first to estimate an assimilation elasticity that relates quasi-experimental variation in economic proximity to the transferability of migrants' human capital.[5] In many labor markets, it is "notoriously difficult" to measure the ability of workers to transfer skills across sectors (Bauer et al., forthcoming). An advantage of the agricultural context is that farming in different environments requires very different production methods, and we observe many characteristics that capture these differences.[6] Using a detailed geospatial dataset containing predetermined village-level measures of topography, climate, soil characteristics, and local languages, we measure the differences between individuals' birth locations and the 1,021 Transmigration villages in the Outer Islands.

We construct two novel measures of economic proximity that take advantage of the vast differences in agricultural systems and ethnolinguistic diversity across Indonesia.[7] Our first measure, agroclimatic similarity, is higher when origins (in Java/Bali) and destinations (in Outer Islands) have more similar agroclimatic features and hence growing conditions. We compare Transmigration and other villages in the Outer Islands and find that non-Transmigration villages exhibit less dispersion in agroclimatic similarity, which is consistent with Heckman and Honore (1990). Moreover, the distribution of agroclimatic similarity in non-Transmigration villages displays a rightward shift relative to Transmigration villages, suggesting that spontaneous migrants choose to move to destinations that are similar to their origins. We apply a similar approach to measure linguistic similarity using *Ethnologue* data on the native homeland and familial classification of Indonesia's more than 700 languages.

Our preferred estimate implies that an increase in the agroclimatic similarity index by one standard deviation leads to a 21.5 percent increase in rice productivity.[8] Villages with a high share of Java/Bali

[4] Rice is the main crop in Indonesia and is grown by nearly 80 percent of farmers. Rice productivity is an important outcome because one of the key objectives of the Transmigration program was to expand rice output in the Outer Islands for food security reasons. Light intensity growth has been shown to proxy for economic growth at a cross-country level (Henderson et al., 2012), and we use it here to measure changes in village-level economic activity from 1992 to 2010.

[5] Friedberg (2000), Lubotsky (2007), and Abramitzky et al. (forthcoming) study the economic assimilation of migrants to Israel and the United States (see Borjas, 1999, for a survey of the literature). Other studies that examine the relationship between

migrants from origins with similar agroclimatic characteristics exhibit higher yields compared to observably identical villages that have a high share of Java/Bali migrants from dissimilar origins. Linguistic similarity also increases productivity. Additionally, both proximity measures have positive impacts on broader economic development captured by light intensity growth and revenue-weighted yield per hectare across all crops. Our estimates for light intensity growth suggest that a one standard deviation increase in agroclimatic similarity leads to an additional 0.15 percentage points of village-level income growth annually between 1992 and 2010.[9]

These results suggest the specificity of farming skills has long-term consequences. We investigate possible margins of adaptation to agroclimatic and linguistic dissimilarity including occupational sorting, crop choices, and ex post migration. We find some evidence of adaptation, but the effects are small. For example, a one standard deviation increase in agroclimatic similarity leads to a 0.9% greater likelihood of Java/Bali migrants choosing farming as their primary occupation while a one standard deviation increase in linguistic similarity leads to a 1.7% greater likelihood of migrants being traders (an occupation where language is important). This pattern of occupational choice is consistent with theories of sorting based on comparative advantage, but the magnitudes are small.

We contribute to a growing literature that examines comparative advantage in agriculture.[10] Our key innovation is to identify the economic proximity between migrants' origins and destinations as a proxy for labor adjustment costs at the destinations and hence a measurable source of comparative advantage. Costinot et al. (2012) simulate a structural trade model that allows farmers to adjust costlessly, so comparative advantage (as proxied by potential yields) can evolve as farmers adapt to climate change. Our findings show that farmers may face difficulties adjusting to new agroclimatic conditions, even after a decade of experience.[11] Migrant skills acquired in dissimilar origins may be a low quality match to the production environment in the destination. Because higher similarity reflects a better match quality, villages assigned a higher share of migrants from agroclimatically and linguistically similar origins experienced a "shock" of higher quality-adjusted labor endowments. Such villages therefore have greater comparative advantage at farming than villages assigned a high share of migrants from dissimilar origins. This provides a useful lens to study the role of comparative advantage in shaping the spatial distribution of productivity with mobile labor.

We also provide novel insights on how the mobility and spatial (mis)allocation of labor can affect economic development.[12] Recent studies emphasize difficulties faced by farmers in making optimal migration decisions (Michalopoulos, 2012; Atkin, 2013; Bryan et al., 2013; Munshi and Rosenzweig, 2013).

West Bengal by 20 percent. Adamopoulos and Restuccia (forthcoming) use a calibrated general equilibrium model to show that farm size restrictions reduced agricultural productivity in the Philippines by 7 percent.

[9]In our sample, the annualized growth rate in light intensity is 0.8 percentage points, which implies income growth of 0.4

Our key innovation is to identify agricultural productivity losses due to limited skill transferability and adaptation. Barriers to the transferability of farming skills may constrain decisions on whether and *where* to move.[13] In decomposing agroclimatic similarity into the three sub-components of topography, soil, and climate, we find that topographic similarity is a key determinant of rice productivity. This is consistent with agronomic research suggesting, for example, that adjustment to highland farming systems is quite difficult for migrant farmers from lowland areas. Helping migrant farmers acquire more general skills (such as language) may help them overcome these barriers.[14] Overall, our results suggest that even in the absence of policy barriers to internal mobility, skill-specificity may constrain labor reallocation across regions that differ in (potential) productivity and hence perceived arbitrage opportunities.

Finally, our paper also contributes to the literature on identification in multi-sector Roy models (Heckman and Honore, 1990). As discussed in Section 3, previous studies have relied on panel data, instrumental variables, and control function approaches to address biases due to endogenous sorting based on unobservable comparative advantage. Our measures of economic proximity capture an observable source of comparative advantage, and the resettlement process generates plausibly exogenous variation in migrants' locations that enables credible reduced form identification in a cross-section.

Our findings have important policy implications, given the growing need to relocate millions of rural households affected by climate change, infrastructure development, and industrialization, especially in developing countries.[15] As a policy exercise, we estimate long-run average treatment effects of the Transmigration program on local economic development by comparing Transmigration villages against planned villages that were never assigned transmigrants. These counterfactual, almost treated villages exist because the program was abruptly halted due to budget cutbacks following the sharp drop in global oil prices in the mid-1980s. Using a place-based evaluation approach akin to Busso et al. (2013) and Kline and Moretti (forthcoming), we find null average treatment effects on agricultural productivity and light intensity growth. These null impacts stem in part from the heterogeneous effects of agroclimatic similarity. Semiparametric regressions for Transmigration villages identify important non-linearities with the largest elasticity estimates for the bottom tercile of agroclimatic similarity and smaller elasticities for the top tercile. This suggests that barriers to transferability are most significant in places with very low agroclimatic similarity. An important lesson for the design of resettlement schemes is that optimizing the person-to-place match is important, and it is especially important to avoid very bad matches.

The remainder of the paper proceeds as follows. Section 2 provides background on the Transmigration program and describes the data. Section 3 develops our theoretical framework and empirical strategy in the context of a Roy model. Section 4 presents our main results. Section 5 concludes.

2 Indonesia's Transmigration Program: Background and Data

Indonesia is the world's fourth most populous country, with over 230 million people living on more than 6,000 different islands. The archipelago is home to more than 700 distinct ethnolinguistic groups and remarkable agricultural biodiversity.[16] Like many countries, the spatial distribution of its population has historically been highly skewed, 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, despite containing only 7.3 percent of the nation's total land area. The remaining population is scattered across the *Outer Islands*, consisting of the vast islands of Sumatra, Sulawesi, and Kalimantan, as well as Maluku, Nusa Tenggara, and Papua in Eastern Indonesia.

Indonesia's Transmigration program was designed to alleviate these perceived population pressures. Over many decades, the program relocated millions of poor and landless agricultural households (hereafter, transmigrants) from rural areas of Java and Bali to rural areas of the Outer Islands. By moving farmers to previously unsettled areas, planners hoped that the program would also increase national agricultural output, especially rice production (Kebschull, 1986; MacAndrews, 1978). Our study focuses on the most intensive waves of the program that took place between 1979 and 1988, during President Suharto's third and fourth five-year plans (or *Repelita*).[17]

Participation in the program was almost entirely voluntary.[18] Participating settlers were given free transportation to their new transmigration villages. Prior to resettlement, program officials cleared land at destination sites, preparing it for agricultural use, and connected these sites to the road network. They also built houses for transmigrants, and each household received two hectare plots of agricultural land that were often allocated by lottery. Settlers received property rights to their homes and land, and they were given provisions for the first few growing seasons, including seeds, tools, and food (Otten, 1986). In some cases, the government provided temporary agricultural extension services, social infrastructure (e.g, schools, mosques, and health facilities), and support for the development of cooperatives and other social institutions. Additionally, between 10 and 30 percent of the land at each site was reserved for members of the indigenous population, who would move from nearby areas to take advantage of access to new land.

We identify 1,021 Transmigration villages established between 1979 and 1988 using new data that we digitized from Indonesia's Transmigration Census, produced in 1998 by the Ministry of Transmigration (MOT). The locations of these villages are depicted in Figure 1. More than half of the Transmigration sites are located on the island of Sumatra (566 out of 1,021), but many are also found on Kalimantan (223) and Sulawesi (161), with much smaller numbers in Maluku and Nusa Tenggara. The spatial variation in

[16] As Donner (1987) notes, "Conditions often change from one place to another, even on the small islands, and the rugged relief

destination sites exposed transmigrants to many different types of growing conditions, some of which were unfamiliar to settlers who had learned how to farm on Java or Bali. In the next subsections, we describe the different margins of selection in the program and discuss how they relate to identification in our empirical work.

2.1 Selecting People, Places, and Assignments

Individual Selection. To participate in the program, transmigrants had to be Indonesian citizens in good physical health. The program targeted entire families for resettlement, and according to official guidelines, couples had to be legally married, with the household head between 20 and 40 years of age. The youngest member of the family could be no younger than 6 months, and the oldest not more than 60 years. Most program participants were landless agricultural laborers, who were often very poor, with no savings, few assets, and very little schooling (Kebschull, 1986).[19] Most had always lived in the village where they were born and had strong social ties to their communities.

Individual Assignments. Once selected, participants exerted little, if any, control over their eventual destination in the Outer Islands. Many previous studies argue that even just prior to departure, transmigrants were ill-informed about the geographical location, native ethnic group, and agricultural systems in the areas where they were sent. For instance, in Kebschull's pre-departure survey, approximately 30 percent of transmigrants could not name the village where they would be settled, 68 percent did not know the nearest town, and 82 percent knew nothing about the local agroclimatic conditions.[20] Another study, focusing on earlier waves of the program, found that transmigrants had "little idea that the quality of soil in the Outer Islands was different from that at home, nor did they have any clear understanding that these new conditions required different techniques of farming" (Guinness, 1977, p. 108). In Section 4.3, we show that there is little evidence that transmigrants were assigned to settlements on the basis of perceived complementarities between their skills and local conditions at destinations.

Site Selection. The MOT used a three-stage process to select agriculturally viable settlement areas. In the first phase, potential recipient villages were identified using large-scale maps, which captured basic information about elevation, vegetation, soil types, market access, and locations of existing settlements. Then, planners used aerial reconnaissance and detailed mapping to screen sites based on climate, hydrology, present land use, forest status, and agricultural potential. Areas with slopes greater than 8 percent of the net area were deemed unsuitable for the program. Finally, after suitable, "recommended development areas" (RDA) were found, MOT officials conducted detailed surveys of topography, hy-

2.2 Origin-by-Destination Match Quality: Agroclimatic and Linguistic Similarity

Most transmigrants expected to pursue the same sort of farming activities they had been practicing in their origin villages. However, the arbitrary assignment mechanism placed many transmigrants in villages with unfamiliar agricultural conditions. Although agricultural extension services were supposed to help transmigrants adapt to their new environments, the quantity and quality of these services were often lacking.[21] In order to measure how easily farmers can transfer the skills they acquired in Java/Bali to the new settlements, we construct two similarity indices measuring proximity or match quality between origin and destination sites.

Agroclimatic Similarity. We consider first the agroclimatic similarity between transmigrant origins and destinations. Unfamiliar climatic and soil conditions could pose difficulties for transmigrants making use of their newly acquired land for farm production. Agroclimatic differences are particularly salient in the case of rice. On Java/Bali, it is more commonly grown on irrigated wetland (*sawah*), while in the the Outer Islands, rainfed dryland cultivation is more common (Geertz, 1963).[22]

We use data from the *Harmonized World Soil Database* (HWSD) and other sources to measure many agroclimatic characteristics, including (i) *topography* (elevation, slope, ruggedness, and altitude), (ii) *water access* (distance to rivers and the sea coast), (iii) *soil characteristics* (texture, drainage, sodicity, acidity, and carbon content), and (iv) *climate* (rainfall and temperature). These characteristics, which we measure at a very high spatial resolution, are fundamental components of agricultural and especially rice productivity (Moormann, 1978).[23] Table 1 presents the mean and standard deviation for each of the variables separately for the villages of Java/Bali and the Outer Islands. These summary statistics demonstrate the remarkable agroecological variation across transmigrants' (potential) origins and destinations. Note that the time-varying characteristics are measured pre-program and hence not subject to concerns about reverse causality. For instance, all of the soil type information is based on data from the Soil Map of the World (FAO 1971-1981). Let $\mathsf{x}_j = (x_{j1}, ..., x_{jG})'$ denote this $(G \times 1)$ vector of predetermined agroclimatic characteristic for location j.

Given location j's agroclimatic characteristics, x_j, the agroclimatic similarity of an individual's origin location i and her destination location j can be defined as:

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

where $\mathrm{d}(\mathsf{x}_i, \mathsf{x}_j)$ is the agroclimatic distance between district i and location j, using a metric defined on the space of agroclimatic characteristics.[24] 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, $\mathrm{d}(\mathrm{x}_i, \mathrm{x}_j)$ projects these differences in G dimensions onto the real line (by taking a simple mean across the absolute differences).[25] We multiply by (-1) so that larger differences correspond to lower values of agroclimatic similarity.

Using $\mathcal{A}_{ij}$, we construct an agroclimatic similarity index for location j by aggregating using population weights:

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

where $\pi_{ij} \in [0, 1]$ is the share of migrants residing in transmigration site j who were born in district i. Our preferred index uses individuals born in Java/Bali to calculate the migrant weights, π_{ij}.[26] To construct π_{ij}, we use the universe of microdata from the 2000 Population Census, which identifies each individual's district of birth, district of residence in 1995, 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 or near treated villages in the Outer Islands as potential transmigrants.[27] Overall, the index captures a weighted average of the agroclimatic differences between village j and the origin districts of transmigrants, with weights proportional to migrant shares. We refer to $\mathcal{A}_{ij}$ ($\mathcal{A}_j$) as individual- (village-) level agroclimatic similarity.

We use Figure 2 to illustrate how the index is constructed. We highlight several agroclimatic characteristics in two nearby destination villages on the island of Sumatra and the district that sent the largest share of transmigrants to each of them. Consider the village of *Telang Sari*, which is at elevation of 3m, has an average topsoil pH of 4.73, and 25% of its soil is fine textured. Figure 2 also shows that 23 percent of residents in *Telang Sari* were born in Kebumen in Central Java. This primary sending district is quite similar to *Telang Sari*. Its average elevation (94m), topsoil pH (5.41) and share of fine textured soil (29.5%) are all well within one standard deviation of the characteristics of *Telang Sari*.

To arrive at the agroclimatic similarity index of 0.5 for *Telang Sari*, we first convert these characteristics into z-scores, and then take the sum of the absolute differences in these characteristics, $\mathrm{d}(\mathrm{x}_i, \mathrm{x}_j)$, for each sending district i represented in *Telang Sari*. This gives us $\mathcal{A}_{ij}$. Then, to construct $\mathcal{A}_j$, we take averages by applying migrant weights for each sending district (e.g., 0.23 for Kebumen). Finally, we multiply by negative one and scale the index to lie on the unit interval.

In comparison to the relatively high agroclimatic similarity index for *Telang Sari* (0.5), Figure 2 also shows a second destination village, *Nunggal Sari*, which is only 36 kilometers away from *Telang Sari* and has similar agroclimatic characteristics as *Telang Sari* but a lower agroclimatic similarity index (0.4). This is because the main Java/Bali origin district represented in *Nunggal Sari* has much higher elevation (493m instead of 5m for *Nunggal Sari*), less acidic topsoil (6.62 instead of 4.85) and less coarse soil(57.6% instead of 14.3%).

guistic similarity index. To capture this 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.[28] Analogous to the agroclimatic similarity measure in (1), *linguistic similarity* for village j can then be summarized using the following index,

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

where $\pi_{\ell j} \in [0, 1]$ is the share of Java/Bali-born immigrants in j from ethnolinguistic group ℓ in Java/Bali, $branch_{j\ell}$ is the sum of shared language tree branches between ℓ and the language indigenous to village j, $max\ branch = 7$ is the maximum number of shared branches between any Java/Bali language and any native Outer Islands language, and ψ is a parameter, set to 0.5 as a baseline following Fearon (2003). As with others using these types of measures in the economics literature (e.g., Desmet et al., 2009; Esteban et al., forthcoming), 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 natives and transmigrants.

Figure 3 demonstrates the spatial variation—within and between islands—in our two village-level measures of similarity, $\mathcal{A}_j$ and $\mathcal{L}_j$. This variation arises from three sources: (i) the rich agricultural and ethnolinguistic diversity within Java/Bali (many x_i, languages ℓ), (ii) the diffusion of Transmigration settlements across the even more diverse Outer Islands (many x_j, languages j), and (iii) the arbitrary assignment of transmigrants across settlements (continuum of π_{ij}, $\pi_{\ell j}$). The spatial variation remains considerable even if we ignore the within-settlement distribution of origin districts and ethnicities by simply comparing the given site j to the average origin district and all ethnicities (i.e., setting π terms to 1). In sum, the variation is substantial, which is important for our ability to identify the importance of economic proximity using a natural experiment and a single cross-section.

Although other dimensions of origin-by-destination match quality may be important, we focus on agroclimatic and linguistic similarity as two salient dimensions of quality that were alluded to in several 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.[29] Moreover, as borne out further in Section 3.2, these two sources of heterogeneity are plausibly exogenous given the largely arbitrary assignment of transmigrants across settlements.

[28]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

2.3 Counterfactual Settlements: Exploiting a Policy Discontinuity

Another key feature of the Transmigration program, which we use for identification in Section 4.3, relates to time series variation in program intensity driven by variation in oil revenue. Figure 4 plots the world oil price alongside the number of Transmigrants resettled under Suharto, using the MOT Transmigration Census data described above. In *Repelita* III (1979-1983), the government was flush with oil revenue and substantially increased its funding for the program, and there was a big push to recruit more transmigrants, with a target of resettling 2.5 million people over the five year period. As shown in Figure 4, a total of 1.2 million people—approximately 1.5 percent of the total population of Java and Bali—were relocated through the program during this period.

Spurred on by the popularity of the program, the government set an ambitious target of relocating 750,000 households (3.75 million people) during *Repelita* IV (1984-1988). However, in the mid-1980s, global oil prices collapsed, 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.[30] Because of budget cuts to the program, numerous selected sites never received any transmigrants. We can construct a control group from this set of planned but unsettled villages using a place-based impact evaluation strategy based on the observable characteristics that influenced the process of site selection noted above (see Kline and Moretti, 2013). As detailed in Section 4.3, this enables us to recover unbiased estimates of the program's average treatment effects on village outcomes.

We identify control villages using the MOT's maps of recommended development areas (RDAs) constructed during the site-selection process. There were a total of 969 RDAs identified by the maps, though many were adjacent to one another. We digitally traced these RDAs using GIS software and overlaid the results onto maps of village boundaries in 2000. We define as controls those 907 villages that shared any area with the RDA polygons. These planned sites serve as a counterfactual for what would have happened to Transmigration villages had the program not taken place.

2.4 Summary Statistics: Demographic and Economic Outcomes

Our empirical work primarily relies on measures of economic outcomes at the *village* level. Ideally, we would want to estimate the impact of similarity on individual-level outcomes. However, datasets with useful individual outcome data either (i) do not contain enough Transmigration sites (e.g., the *Indonesia Family Life Survey* only covers 23 settlement villages) or (ii) cover Transmigration sites but do not contain information on migration or place of birth (e.g., annual *Susenas*). Our best individual-level dataset, the 2000 Population Census, contains information on migration and covers all settlement areas, but productivity outcomes, such as wages or agricultural yields, are not recorded.

ethnic groups from Java/Bali (largely reflecting the second generation of transmigrants born in the Outer Islands). 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.

Each village typically hosts transmigrants from numerous origins. The average Transmigration village comprises transmigrants from 44 different origin districts (out of 119) and hailing from three Java/Bali ethnicities (out of 8). The average origin district Herfindahl index in these villages is 0.15 suggesting that concentration is not high.

Moreover, the Census also includes data on schooling attainment and sector of work, which we use below to investigate selection and occupational choice across different cohorts. Educational attainment in Transmigration villages was quite low, with an average of less than 4 years of schooling. Native Outer Islands ethnics residing in Transmigration villages had slightly higher schooling than Java/Bali ethnics.

Second, we measure agricultural productivity using the triennial administrative census known as *Podes* (or Village Potential). The August 2002 round provides detailed information on agricultural activities including area planted and total yield for over one hundred crops in the 2001-2 growing season. Our main focus will be on rice, Indonesia's most important staple food that is consumed by nearly all households and is grown across the country.[31] As rice is the primary crop grown on Java/Bali, most transmigrants expected to be able to grow rice in their new villages (Kebschull, 1986), and policymakers envisaged the program as a means of increasing national rice output. From Table 2, we see that on average, 70 percent of residents were employed in farming. Nearly 75% of Transmigrant villages were growing some rice, and the average village produced 2.7 tons per hectare.

If areas were poorly suited to rice production, transmigrants could have chosen to grow other crops. Therefore, we also consider a broader measure of agricultural productivity. We follow Jayachandran (2006) and Duflo and Pande (2007) in constructing a revenue-weighted average yield across all crops grown in village j:

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

where $C_j \subseteq C$ denotes the subset of crops grown in village j among all crops for which we have price data, p is the unit price in 2001 from FAO/PriceSTAT, y_c is total (harvested) output, and a_c is area planted as reported in *Podes* 2003 (see Appendix A).

Broader economic development, beyond agricultural productivity, is captured using nighttime light intensity from the National Oceanic and Atmospheric Administration (see Henderson et al., 2012). Light intensity has been identified as a good proxy for local income within Indonesia over a period of rapid electrification beginning in the late 1980s (Olivia and Gibson, 2012). Changes in light intensity from 1992

3 Theoretical and Empirical Framework

This section lays out our conceptual framework for relating economic proximity to long-run development and productivity. We first explain what we mean by location-specific human capital and economic proximity, and how it serves as a measurable source of comparative advantage. We then derive our key estimating equation and discuss identification. As discussed in Section 2.2, we consider two salient measures of economic proximity: agroclimatic and linguistic. We will focus here on agroclimatic similarity, $\mathcal{A}_{ij}$, but the discussion readily generalizes to linguistic similarity, $\mathcal{L}_{ij}$.

3.1 Theoretical Framework

We adapt the classic Roy (1951) model with two sectors to a multi-location choice model where heterogeneous farmers sort across heterogeneous locations. Our setup is similar to Dahl (2002), but adapted to an agricultural setting.

There is a discrete set of J locations, indexed by $j = 1, ..., J$. The farming methods (production functions) are different in each location. Each location is differentiated by a bundle of characteristics which we denote using a fixed $(G \times 1)$ vector, x_j, as discussed in Section 2.2. Individual farmers, indexed by i, are born into a birth location, $b(i) \in \{1, ..., J\}$. Individuals acquire farming skills that are specific to local growing conditions at their birth locations, captured by $\mathsf{x}_{b(i)}$.[33] For rice, this would include, among others, knowledge of what types of varieties are best suited to local growing conditions.[34] Farming skills are not perfectly transferable across locations so that some knowledge may not be helpful when trying to farm in different environments. Hereafter, we denote $\mathsf{x}_{b(i)}$ with x_i to simplify notation.

For simplicity, 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. Using a correlated random coefficients framework, we can write the value of output per unit of land owned by farmer i in location j as:

$$y_{ij} = \gamma_i \mathcal{A}_{ij} + \mathsf{x}_j' \beta, \tag{3}$$

where $\mathsf{x}_j' \beta$ maps observable agroclimatic characteristics of location j into productivity, $\mathcal{A}_{ij} = (-1) \times \mathrm{d}(\mathsf{x}_i, \mathsf{x}_j) = -\sum_g |x_{jg} - x_{ig}|$ is our measure of agroclimatic similarity between locations, and γ_i captures the farmer's adaptability or adjustment costs. The transferability of migrants' skills between origins and each potential location depends both on how similar the growing conditions are between the two locations and also on how adaptable the migrant is. Conditional on γ_i (among farmers with similar adaptability), y_{ij} should be increasing in $\mathcal{A}_{ij}$ because skills specific to x_i are more complementary the more similar is x_j to x_i. Conditional on $\mathcal{A}_{ij}$, a higher γ_i means $\mathcal{A}_{ij}$ is an important predictor of producimigrant's origin does not matter and hence $\gamma_i = 0$.

Economic proximity ($\mathcal{A}_{ij}$) is an important source of comparative advantage because it reflects complementarity between a farmer's location-specific knowledge and local farm characteristics. For a given destination, farmers migrating from more similar origins will find it easier to transfer their farming skills, compared to farmers from dissimilar origins. Therefore, farmers from similar origins are comparatively advantaged relative to farmers from dissimilar origins. In this way, economic proximity proxies for a key source of comparative advantage among farmers (the match quality or complementarity between their location-specific knowledge and local agroclimatic endowments).[35]

As in Dahl (2002), we assume that location choice depends on productivity and preferences. 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. Productivity and taste differences determine how farmers sort across locations.

The stylized example described in the Introduction is a simplified version of this model, where there are only two types of farms (Highland and Lowland) and farmers are either born in highlands or lowlands. There are four potential outcomes: $Y_{HH}, Y_{LL}, Y_{HL}, Y_{LH}$, where Y_{HH} and Y_{LL} are outcomes associated with high economic proximity and Y_{HL} and Y_{LH} are farmers-to-farm matches that are associated with low economic proximity. 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 the econometrician would only observe two of the four outcomes, namely those associated with high economic proximity only: Y_{LL}, Y_{HH}. In this case of perfect sorting, there is no observed variation in economic proximity.

Returning to the multi-location Roy model, in equilibrium, farmers choose locations to maximize V_{ij}, and we denote farmer i's optimal location by $j(i)^*$. In this setting, each farmer i has J potential outcomes for agricultural productivity, which we can write as $y_{i1}, y_{i2}, \ldots, y_{iJ}$. As shown by Heckman and Honore (1990), it is difficult to identify the importance of comparative advantage because, as the theory of sorting based on comparative advantage implies, we do not observe all the potential outcomes for each farmer (we observe $y_{ij(i)^*}$ instead of all J of the $y'_{ij}s$). A common solution is to identify instruments that affect location choice but are excluded from the determination of productivity. The problem is that many variables jointly affect both location choice (V_{ij}) as well as productivity (y_{ij}). A second problem is related to the fact that we have a *multi*-sector Roy model. Each individual has J potential locations, instead of a binary choice (such as whether or not to work). In addition to needing an instrument that satisfies

3.2 Empirical Strategy

The empirical object of interest is an "assimilation elasticity" that relates the economic proximity between migrant farmers' origins and destinations to productivity at the destinations (we focus here on yield per hectare). Our key regression is at the village level, but it is instructive to begin at the individual level and then aggregate up. We derive the main estimating equation by augmenting the model above to allow for observable and unobservable determinants of productivity. In particular, let γ be the mean γ_i and write $(\gamma_i - \gamma)$ as the individual-specific deviation from the mean. After collecting terms, we obtain an individual-level equation:

$$y_{ij} = \underbrace{\gamma \mathcal{A}_{ij} + \mathrm{x}_j' \beta}_{\text{observable}} + \underbrace{(\gamma_i - \gamma)\mathcal{A}_{ij} + \mu_j^u + \omega_{ij}}_{\text{unobservable}}, \tag{5}$$

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

As mentioned in Section 2, we observe productivity data at the village level. The difficulty is that with sorting, our village-level estimating equation aggregates (5) over the non-random set of individuals, I_j, whose optimal location is j: $y_j \equiv \sum_{i \in I_j} y_{ij(i)}$. Our primary village-level estimating equation is:

$$y_j = \gamma \mathcal{A}_j + \mathrm{x}_j' \beta + \underbrace{\sum_{i \in I_j} (\gamma_i - \gamma)\mathcal{A}_{ij} + \mu_j^u}_{\text{unobservable}} + \omega_j, \tag{6}$$

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 elaborated in equation (1)) and ω_j is an idiosyncratic error term.[36]

The key parameter of interest, γ, measures the causal impact of average agroclimatic similarity on aggregate rice output per hectare for the village. Our regression conditions on observably identical destination villages (such as *Telang Sari* and *Nunggal Sari* in Figure 2), and compares villages that have a high share of Java/Bali migrants from similar origins (*Telang Sari*) against villages that have a high share of Java/Bali migrants from dissimilar origins (*Nunggal Sari*). As shown in Figure 2, *Telang Sari* (high $\mathcal{A}_j$) has higher rice productivity (2.9 tons per hectare) compared to *Nunggal Sari* (2.0 tons per hectare). This pair of outcomes reflects a positive relationship between agroclimatic similarity and productivity ($\gamma > 0$).

The γ parameter also sheds light on the role of comparative advantage in shaping the spatial distribution of aggregate productivity. As described in Sattinger (1993), the theory of comparative advantage describes the allocation of factor endowments (workers) to production units based on their relative cally similar origins experienced a "shock" of higher quality-adjusted labor endowments. Such villages therefore have greater comparative advantage at farming than villages assigned a high share of migrants from dissimilar origins.

The ability to quantify an observable source of comparative advantage is an important innovation of our research design. In a recent study on population resettlement in Germany, Bauer et al. (forthcoming), note that it is "notoriously difficult to measure" skill transferability between migrants' origins and (potential) destinations. We are able to do so because a wealth of agronomic research has identified and collected data for (predetermined) agroclimatic characteristics that are vital to farm output. Moreover, land and local climate characteristics change slowly, so that predetermined agroclimatic characteristics measured in the 1970s are still highly predictive of productivity in 2000. All land attributes and ethnolinguistic homelands are predetermined and hence unaffected by settler farming activities and any corresponding inflow of capital or labor. Hence, we can abstract from reverse causality concerns.

The key source of plausibly exogenous variation is the assignment of Java/Bali transmigrants to Outer Islands. As discussed in Section 2, there are multiple sources of variation in the village-level index $\mathcal{A}_j$: (i) variation in the absolute differences between predetermined agroclimatic characteristics in destinations versus origins, $\mathsf{d}(\mathsf{x}_i, \mathsf{x}_j)$, and (ii) variation in the share of Java/Bali migrants in destination village j who are from origin district i, π_{ij}. Our regression conditions on x_j and exploits variation in π's from origins with similar versus dissimilar x's.

Threats to Identification. We first show that the distribution of agroclimatic similarity is different among Transmigration villages compared to other villages in the Outer Islands. Panel A of Figure 5 plots the 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 economic proximity between origins and destinations. The distribution for non-Transmigration villages is shifted to the right. Second, there is greater dispersion in realized similarity in Transmigration villages. This is consistent with the discussion in Heckman and Honore (1990) that the Roy model has "no empirical content."[37]

To show how omitted variables could bias our estimate of γ, consider the case where a Transmigration village j has either low (L) or high (H) agroclimatic similarity. Then, we have $\gamma = \mathsf{E}(y_j|\mathsf{x}_j, \mathcal{A}_j = H) - \mathsf{E}(y_j|\mathsf{x}_j, \mathcal{A}_j = L)$. This equation clarifies why high and low proximity villages are not comparable in a typical setting. First, they could be different due to omitted natural advantages (μ_j^u). Second, farmers are likely to migrate only to locations where their skills are transferable so that the realized outcomes are 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 ethnolinguistic group ℓ and origin district i residing in Transmigration village j in 2000 is increasing in agroclimatic ($\mathcal{A}_{ij}$) and linguistic similarity ($\mathcal{L}_{\ell j}$) between i/ℓ and j. The following equation

$$migrants_{i\ell j} = \alpha + \chi_a \mathcal{A}_{ij} + \chi_\ell \mathcal{L}_{\ell j} - \chi_d \ln distance_{ij} + \tau_i + \tau_\ell + \upsilon_{i\ell j}, \tag{7}$$

is estimated using nonlinear count data estimators—namely, Poisson and negative binomial—to account for over-dispersion (i.e., many zero migration $i\ell j$ corridors). We also report OLS estimates with $\ln(migrants_{i\ell j})$ for villages with migrants in the given $i\ell j$ cell. In all cases, we cluster standard errors within i and restrict our definition of (trans)migrants in j to individuals born in Java/Bali.

In all specifications of equation (7) reported in columns 1-3 of Table 3, we can reject the null hypotheses that $\chi_a > 0$ and $\chi_\ell > 0$. This provides strong suggestive evidence that 12 to 20 years after the initial wave of resettlement, migrants from Java/Bali did not endogenously sort in (out) of more (dis)similar sites. Although migrant stocks tend to be somewhat higher in physically closer sites ($-\chi_d > 0$), linguistic and agroclimatic "distance" do not exhibit the same hypothesized gravity forces. Moreover, the non-Transmigration villages in columns 4-6 exhibit considerably stronger sorting on distance as well as linguistic similarity. We account for the relationship between distance and migrant stocks in Transmigration by conditioning on (π weighted) distance to Java/Bali (origins) in our results below.

A related concern is that agroclimatically similar destinations are initially assigned or subsequently attract different settlers along unobserved dimensions that are correlated with productivity. Although transmigrants are negatively selected on average, there is some variation across settlements in average schooling acquired by household heads before the program. We show in Table 4 that this variation in schooling of Java/Bali migrants is not correlated with $\mathcal{A}_j$ and $\mathcal{L}_j$. Neither similarity index has an economically or statistically significant relationship with predetermined schooling acquired by eligible individuals born in Java/Bali between 1930-1979. In Figure 6, we plot the kernel densities of individual-level agroclimatic similarity for all Java/Bali-born migrants in Transmigration villages by schooling. The distributions appear to be very similar across schooling levels.

Another concern is that destinations that are agroclimatically similar to Java/Bali may have unobservable natural advantages, which implies upward bias on γ because Java/Bali is known to be naturally advantaged for rice production, $Cov(\mu_j^u, \mathcal{A}_j) > 0$. Our assumption is that the correlation between μ_j^u and $\mathcal{A}_j$ is zero, conditional on x_j.

Using the planned but unsettled villages as counterfactuals, we show that the *unweighted* agroclimatic similarity with Java/Bali (i.e., an index that compares characteristics in destination j, x_j, to the average tivity in Java/Bali, our key measure of *weighted* agroclimatic similarity identifies the *added* productivity effect of being assigned relatively more migrants from regions of Java/Bali with similar endowments than from regions with dissimilar endowments (conditional on x_j).

We can also rule out concerns that improved planning between *Repelita*s III and IV led to the selection of more agroclimatically and ethnolinguistically similar sites towards the end of our study period. To test this, we simply regress the year of settlement, $t \in \{1979, \ldots, 1988\}$, on the *weighted* similarity indices, conditional on x_j. In OLS and ordered logit specifications, both indices have a negligible, statistically insignificant relationship with t. The same null results hold for the *unweighted* similarity indices.

Overall, the evidence suggests that our main measures of economic proximity, agroclimatic ($\mathcal{A}_j$) and linguistic ($\mathcal{L}_j$) similarity, are quasi-randomly distributed across Transmigration villages, even as observed nearly two decades after resettlement. Endogenous sorting and selection on comparative advantage matter in most migration contexts. However, the near-random initial assignment of transmigrants across sites had persistent effects on location choices over subsequent decades. The absence of long-run sorting is consistent with network effects, high costs of inter-island mobility, and location-specific skills limiting migration to nearby cities in the Outer Islands.

Other Approaches to Identification with Comparative Advantage. Overall, our research design builds upon existing approaches to comparative advantage in three ways. First, we show how to use quasi-experimental variation in origin-by-destination match quality to identify the importance of comparative advantage using a single cross-section. Other work in the literature uses panel data to compare stayers across time (as in Moretti, 2004a), or focuses on switchers (as in Suri, 2011; Combes et al., 2008; Gibbons et al., 2005). The key assumption then is that these sources of variation are orthogonal to sorting forces based on unobserved comparative advantage.

Second, we characterize *and* measure the underlying sources of comparative advantage. Rosenzweig and Zhang (2013) examine differences in comparative advantage with respect to skill and brawn between men and women, using differences in birthweight between Chinese twins. We focus on the agricultural context and argue that one can open the black box on comparative advantage and not only measure some of its key components but also relate those components to aggregate productivity. Costinot et al. (2012) also identify the sources of comparative advantage by crop and location (globally) using agronomic models of potential productivity in a single cross-section. Although our underlying intuition is similar, we take a reduced form approach to causal identification based on observed productivity whereas Costinot et al. take a structural approach allowing them to simulate how productivity (and welfare) will evolve if farmers have no adjustment costs and behave optimally.

Third, we can directly test several of the key assumptions underlying our identification of the causes

In sum, our empirical strategy offers a rich quasi-experimental design to a small but growing literature aimed at unpacking the causes and consequences of sorting based on comparative advantage in Roy model settings. We turn next to a discussion of empirical results based on this strategy.

4 Empirical Results

In this section, we present our main empirical results. We begin by showing that economic proximity matters for productivity over the long-run and provide evidence of heterogeneous effects consistent with our conceptual framework. Then, we show how limited adaptation can explain part of the persistent long-run effects of match quality that we observe in the data. Finally, we estimate the average treatment effects of the program using planned but unsettled villages as controls. Ultimately, we argue that the persistent, long-run effects of economic proximity in Transmigration villages may explain the limited average impact of the program on local economic development in the Outer Islands.

4.1 Long-Run Effects of Economic Proximity

Table 5 reports our main results: estimates of $\boldsymbol{\gamma} \equiv (\gamma_a, \gamma_\ell)$, the coefficients on economic proximity in the following regression based on equation (6):

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

where village-level agroclimatic ($\mathcal{A}_j$) and linguistic similarity ($\mathcal{L}_j$) are based on the Java/Bali migrant weights; $\mathbf{x}_j$ includes island fixed effects, the full set of predetermined controls elaborated in Section 2.2;[40] 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 75 kilometers of village j.

Columns 1 and 2 report results when each measure of proximity enters the regression separately, while column 3 reports results with both measures included. The coefficient estimates are stable across columns. This reinforces the notion that our similarity indices are uncorrelated and that the effect of one similarity index, conditional upon the other, is similar to the unconditional effect.[41] Column 3 reports our preferred estimates 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.

The top panel reports our main results on rice productivity (measured as log yield per hectare). The estimation sample includes 600 Transmigration villages with non-missing data for rice productivity. A one standard deviation increase in the agroclimatic similarity index leads to 21.5 percent increase in rice (average rice productivity is 2.75 tons per hectare, see Table 2). Our results imply that Transmigration villages with a higher share of Java/Bali migrants from origins with similar agroclimatic characteristics exhibit greater yield per hectare compared to observably identical villages with a lower share of these migrants.

We interpret these results as evidence that having a higher share of well-matched transmigrants (in terms of similarity with growing conditions at their origins) leads to greater productivity. The significance of the agroclimatic similarity coefficient implies barriers to transferring farming knowledge uniformly across locations. Finding effects 12 to 20 years later suggests a high specificity of farming knowledge to local growing conditions and high adjustment costs faced by farmers.

We also find that a one standard deviation increase in linguistic similarity improves rice productivity by 12.9 percent or 0.36 more tons per hectare for the average village. The positive effects suggest that villages with a higher share of farmers from linguistically similar origins have higher productivity, perhaps through learning from natives, as describe in several reports. Although we cannot reject the equality of the coefficients on agroclimatic and linguistic similarity ($p \approx 0.37$), the smaller effects for linguistic similarity point to smaller adjustment costs for farmers sent to linguistically dissimilar areas versus farmers sent to agroclimatically dissimilar areas. Our results, taken together, are consistent with farmers being credit constrained. Adapting to different growing conditions may entail purchasing new equipments, whereas learning a language within 12 to 20 years does not entail large fixed costs.[42]

Although rice is, by far, the most important crop in Indonesia, and expanding rice production was one of the program's main goals, it may not adequately capture economic well-being in these villages. In some of the Transmigration villages in our sample, rice production is not the primary economic activity. We therefore investigate two broader measures of economic development in Table 5.

The first outcome is total yield per hectare, calculated as the revenue-weighted average yield across crops grown in the village. If the Transmigration areas were poorly suited for rice production, farmers could have chosen to grow other crops.[43] Panel B of Table 5 shows that an increase of one standard deviation in agroclimatic similarity increases overall agricultural yields by 7.4 percent. This translates into an additional 0.1 tons/ha on average across all crops relative to a mean of 1.33 tons/ha. Linguistic similarity meanwhile has a similarly positive albeit slightly smaller and less precisely estimated effect.

Our second outcome is growth in light intensity between 1992 and 2010. Panel C shows that a one standard deviation increase in agroclimatic similarity raises long-run light intensity growth by 6.1 percent or around 0.3 percentage points annually relative to a mean of around 16 percent. Although insignificant, the standard errors fall substantially as we expand the spatial clustering radius beyond 100 kilometers (see below). Meanwhile, a one standard deviation increase in linguistic similarity increases light intensity growth by 9 percent, or around 0.5 percentage points annually. Using an approach similar annually according to this luminosity-based proxy for income growth. These effect sizes are not surprising given the limited degree of electrification in remote, rural areas of the Outer Islands where many of these Transmigration villages are located.

In columns 4-6 of Table 5, we examine the impact of similarity on these outcomes for all villages within 10 km of Transmigration sites. We do this because of potential uncertainty in geocoding the locations of Transmigration villages and because the transmigrant population shock is not confined to settlement village borders.[44] Overall agricultural productivity and light intensity growth impacts become weaker and less precise. However, the impacts on rice productivity remain economically and statistically significant as we expand the sample radius around Transmigration villages.

One advantage of our diverse agricultural setting and rich data is that we can identify which aspects of agroclimatic conditions are the main barriers to the transferability of skills. In Table 6, we repeat the estimation in the previous table, but we decompose our main agroclimatic similarity index into three sub-components of topographic, soil, and climate similarity indices. Topographic similarity is a key determinant of rice productivity. This is consistent with research suggesting that adjustment to farming in highlands could be quite difficult for migrant farmers from lowlands (Donner, 1987). Meanwhile, soil similarity appears most important for total agricultural productivity across all crops, and climatic similarity is an important determinant of broader economic development captured by light intensity growth.

Robustness. For our rice productivity results, we conduct numerous robustness checks, reported in Table 7.[45] Each row introduces a single change to the baseline specification in column 3 of Table 5, which is reproduced in row 1 for reference. The results are robust to controlling for variables that capture demographic differences. This includes controlling separately for the gender, age, schooling and occupation shares of Java/Bali and Outer-islands born residents in each village (row 2), controlling for the aggregate origin mix by controlling directly for the $\pi_{\ell j}$ terms used to construct linguistic similarity in equation (2) and four region-level aggregates of the (119) origin district i-specific π_{ij} terms used to construct $\mathcal{A}_j$ (row 6), and controlling for the share of natives which (partially) addresses aggregation bias (row 16). Additionally, rows 3 and 4 show that the results are not confounded by program features that could be correlated with similarity, by controlling for the scale and timing of the initial transmigrant influx.

The results are also robust to controlling for variables that capture heterogeneity across locations, including the log physical distance to Java/Bali-born migrants' origins, weighted by the π_{ij} terms (row 5),[46] a third-degree polynomial in the latitude and longitude of each village (row 7), dropping our natural advantage controls x_j (row 14), and controlling for the share of farmland with irrigation (row 17). Province fixed effects (row 8) reduce the coefficient on $\mathcal{A}_j$ by nearly half suggesting that some of the identifying variation occurs across rather than within provinces. However, the results are statistically old (and hence eligible to be relocated through the program), which addresses concerns that our baseline π'_{ij}s that include all Java/Bali migrants (calculated using birth locations in the 2000 Census) may not be capturing transmigrants. Finally, the results are also robust to trimming the top and bottom 1% of observed rice productivity (row 15).

In Appendix Figure B.2, we show that inference is largely robust to expanding the spatial HAC radius. If anything, our baseline choice of a 75 km radius is too conservative and leads to larger standard errors, particularly for the light intensity growth outcome in Panel C of Table 5. The estimated impact of agroclimatic similarity in column 3, for example, is statistically significant at the five percent level for radii $r \in [120, 320)$ km.[47]

We also address another concern with our main specification: aggregation bias. By regressing village-level outcomes on key similarity regressors that only apply to a subset of villagers (transmigrants), we risk misinterpreting the relationship between economic proximity and productivity. In Appendix Table B.3, we use the best available household-level survey data from a small random sample of 74 Transmigration villages to show that the main productivity effects of economic proximity that we identify in Table 5 are driven by (plausible) transmigrants rather than natives. Estimating an individual-level analogue to equation (8), we find that agroclimatic similarity has a positive effect on farm-level rice yields that is qualitatively and quantitatively very similar to the estimates of our main estimates of γ_a.

Where Does Economic Proximity Matter (Most)? In Table 8, we explore the heterogeneous impacts of agroclimatic and linguistic similarity across different types of Transmigration villages. We first examine the hypothesis that the effects of similarity depend on the scope for learning from native farmers about locally appropriate rice growing techniques. Linguistic similarity would likely be more important in villages where the initial native population is large (relative to the number of transmigrants); in contrast, agroclimatic similarity becomes more important when the initial native population was small.[48]

To test the hypothesis, we split the villages into two groups, based on whether they were assigned above- or below-median number of transmigrants, and separately re-estimate equation (8). 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. Under the assumption that program planners accounted for the native population size when they calculated the carrying capacity, conditional on x_j, a large (small) initial transmigrant population is indicative of a small (large) initial native population. Using transmigrants placed at the start of the program, as opposed to the actual numbers of Java/Bali-born in the village in 2000, is preferable given that it is not subject to any ex-post sorting that may have occurred (i.e., it is not a "bad control", Angrist and Pischke, 2009).

similarity is positive and significant in villages with below median numbers of initial settlers placed in column 2, but we cannot reject that agroclimatic similarity has a similar effect size. If transmigrants encountered relatively few natives with whom to communicate, then agroclimatic similarity within their initial cohort may have had a larger effect on long-run productivity.[49]

In Panel B, we examine how the effects of economic proximity on rice productivity vary with the scope for matching to land types (and production systems) common in one's origin as well as learning from natives about new cultivation methods. We present a simple test of this matching hypothesis by re-estimating equation (8) for three groups of Transmigration villages, with low, medium and high shares of wetland (the most prevalent type of land used to farm rice in Java/Bali). This reduces an otherwise high-dimensional vector of agroclimatic attributes into a single land quality measure that is particularly informative about variation in rice cultivation methods.

We find that agroclimatic similarity has the largest effect in areas with more dryland: a one standard deviation increase in similarity leads to nearly 35% increase productivity. This translates into an additional 0.7 tons/ha at the mean of 2.05 tons/ha in dryland areas. This large effect has two potential explanations. First, farmers from Java/Bali accustomed to wetland agriculture found it difficult to adapt to the dryland approaches in the settlement area (see Section 2.2).[50] 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 easily overcome given the strong natural advantages of wetland production systems.

Meanwhile, linguistic similarity has the largest effects in dryland areas. This is also consistent with productivity being higher in villages where the largely wetland farmers from Java/Bali could more easily communicate with and relate to the nearby indigenous population. That linguistic similarity matters in wetland areas as well (column 3) is consistent with the possibility of learning being important given that the irrigated wetland practices in Java/Bali may differ from the rainfed wetland practices prevalent in the Outer Islands.

Summary. Overall, our results in Tables 5 and 8 are consistent with a small but growing literature in development examining skills and preferences that may be specific to certain agroclimatic settings. Like Michalopoulos (2012) and Atkin (2013), we find a high degree of specificity associated with farmers' origins. Our key contribution is to identify a causal effect of economic proximity to origins on productivity outcomes at the destinations. These results have important implications for the literature on the spatial wage disparities and arbitrage opportunities for rural migrants. Farmers are potentially constrained in where to move because their productivity and consumption are tied to human capital and preferences acquired at the origin. One lesson for policymakers is that resettlement policies targeted at farmers growth and development would be limited due to insufficient planning and hasty preparation of settlement areas prior to the arrival of transmigrants. These authors often found that lack of familiarity with local agricultural practices and difficulty learning from native farmers posed a serious challenge in the early stages of the resettlement process. We show that these early indications of the importance of match quality hold over the long-run.

4.2 Adaptation

The results above show that villages assigned relatively more dissimilar migrants exhibit weaker economic outcomes, even after more than a decade. Given all of the potential margins for adjustment and adaptation to new environments, it is surprising that similarity to Java/Bali characteristics matters so much after so many years. Here, we investigate three potential margins of adjustment to dissimilarity: switching occupations, switching crops, and switching locations.[51] In each case, we find that although some adjustment occurred, it was not large enough to offset the effects of dissimilarity.

Occupational Choice. The first margin of adjustment we consider is whether transmigrants assigned to dissimilar villages switch occupations. 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 market). The theory of comparative advantage predicts that individuals assigned to agroclimatically similar villages are more likely to remain as farmers (as they were in Java/Bali) and those assigned to linguistically similar villages are more likely to switch into trading and services.

We test these predictions in Table 9 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 and linguistic distances. The flexible set of individual- and village-level controls ensures that we are comparing the effects of economic proximity on occupational choices across otherwise observably identical individuals in observably identical villages.[52] 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.[53] 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

[51] We focus on these three salient modes of adaptation to dissimilarity. In a companion paper exploring the nation-building aspect of the program, we find that linguistic dissimilarity may hasten adoption of the national language, *Bahasa* Indonesia.

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. Across all ages, agroclimatic similarity increases the likelihood of farming and decreases the likelihood of trading. A one standard deviation increase in individual agroclimatic similarity leads to a 0.9 percentage point (pp) higher probability of an individual reporting farming as their primary occupation. Meanwhile, a one standard deviation increase in linguistic similarity is associated with an imprecisely estimated 1.5 pp lower probability of farming but a statistically significant 1.7 pp higher probability of trading/services. This pattern is consistent with sorting into occupations based on comparative advantage (as proxied by the complementarity of migrants' skills acquired in the origin and the production environment in the destination.)

Comparing across columns, we find no significant differences in the patterns of occupational choices across the young and old generation of transmigrants. This is consistent with intergenerational persistence in occupational choices.

Although the qualitative patterns in Table 9 suggest that dissimilarity encouraged some individuals to switch occupations, the effects are quantitatively small. For example, the 0.9% agroclimatic similarity effect in column 1 implies that only 5,104 individuals with low similarity (1 SD below the mean) switched out of farming. This is quite a limited effect given that more than 350,000 individuals in the sample are farmers. The effects of linguistic similarity are relatively larger. Overall, however, intersectoral mobility seems to have been limited, presumably due to labor market frictions or other adjustment costs. Such limited occupational sorting may explain in part why we find such large, persistent negative consequences of dissimilarity on aggregate productivity.

Crop Choice. Another potential margin of adjustment would be for households to remain as farmers but to switch the crops they grew. In Table 10, we examine the extent to which the Java/Bali migrants bring their preferences for growing rice with them to Transmigration sites. Following Michalopoulos (2012), we estimate the following regression for Transmigration villages,[54]

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

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

Table 10 demonstrates that Transmigration villages allocate cropland in a very similar manner to their neighbors ($\rho_1 > 0$ in all columns). However, columns 2 and 4 show that on average, the cropping patterns in Transmigration villages depart from those in neighboring Outer Islands villages in a manner reflecting the predominance in transmigrants' origin regions ($\rho_2 > 0$).[55] Consistent with Michalopoulos (2012), our results suggest that origin region cropping patterns explain about 15-20 percent of the patterns explained by spatial autocorrelation. Overall, these results provide evidence that although crop switching might have been possible, it was not fully embraced by transmigrants due to strong preferences for growing rice and replicating the basket of goods grown in their origin regions.

Table 11 provides more direct evidence on crop selection. We regress our similarity indices on binary variables that measure whether villages grew any amounts of certain food (rice, cassava, maize) and cash crops (rubber, palm oil, coffee, or cocoa). Maintaining the baseline specification in equation (8), we find that greater agroclimatic similarity stimulated entry into rice production. A one standard deviation increase in agroclimatic similarity increases the likelihood that farmers in the village grow any rice by 8.1 percentage points relative to a mean of 74 percent. However, agroclimatic similarity also induced some entry into cash crop production, particularly for coffee and cocoa—both of which are grown in several areas of Java/Bali. Hence, some of the productivity effects that we observe in Table 5 are due to an increase in the extensive margin of rice production in settlement areas.

In other words, the results in Table 11 are indicative of revealed comparative advantage in villages with a high share of transmigrants from agroclimatically similar origins. This is why we treat crop choice at the village-level as a measure of adaptation and use the revenue-weighted agricultural productivity measure in Table 5 to compare productivity outcomes across villages with potentially different patterns of comparative advantage across crops.

(Non-)Selective Migration Patterns. Another way in which farmers may adapt to initial low quality matches is by moving out of the village and perhaps returning to Java/ and Bali. While bias from return migration has been shown to be important in the literature (e.g., Abramitzky et al., forthcoming), we argue that this margin of adjustment is less important in our context.[56] 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, previously (mostly) landless agricultural laborers, were given land, and the property rights may play a role in tying them to the Transmigration villages.[57] Moreover, the property rights may have acted as a deterrent to subsequent large scale in-migration to the settlements given constraints on land availability, which informed the planners' initial determination of the size of the transmigrant cohort in each location.

We confirm that selective out-migration is indeed low. First, the 1998 Transmigration census reports the number of individuals initially placed as well as the population size when a Transmigration village was deemed independent enough that it no longer required official supervision (typically within 5-10 years of placement). We regressed the ratios of these two population sizes on our (weighted and unweighted) similarity measures and find statistically insignificant effects of similarity. If there were selective out-migration from dissimilar villages, these coefficients would be positive and significant. Second, we used the 1985 inter-censal survey (*Supas*) to estimate an upper bound on the number of return migrants to Java/Bali. Our calculations suggest there were 1,800,264 individuals in Java/Bali in 1985 who reported being in a different district in 1980. Of these migrants, only 2 percent (36,000) reported living in an Outer Island district that had transmigration villages. This upper bound indicates a low rate of short-term return migration, compared to the roughly 1,298,000 transmigrants who were moved in this period.[58] Finally, recall that the gravity results in Table 3 showed that longer-term *ex post* sorting patterns is not positively correlated with agroclimatic or linguistic similarity. We also show that these similarity measures are uncorrelated with population size and the Java/Bali-born migrant share in Transmigration villages in 2000.

4.3 Average Treatment Effects

In this section, we investigate the policy-relevance of our key findings on economic proximity by linking them to new estimates of the long-term, average impact of the Transmigration program on local economic development outcomes. Obtaining estimates of the overall treatment effects of the program requires an alternative identification strategy based on a place-based evaluation approach. We first develop that approach, then present key estimates of the average treatment effect (ATE) of the program, and finally relate the weak ATE estimates to the above findings on match quality and adaptation.

Exploiting the Policy Discontinuity. In order to identify the ATE of the Transmigration program, we make use of unanticipated budget cutbacks that left numerous planned sites (RDAs) across the Outer Islands unrealized (see Appendix Figure B.3). We use these "almost treated" villages as counterfactuals for what would have happened in Transmigration villages had the migrants and accompanying capital never arrived. This setup lends itself to the following equation:

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

where T_j is a treatment indicator equal to one for Transmigration villages and zero for planned but unsettled RDAs, and x_j is the usual vector of predetermined controls from equation (8). The key parameter