Unity in Diversity?
How Intergroup Contact Can Foster Nation Building*

Samuel Bazzi† *Boston University and CEPR*

Arya Gaduh‡ *University of Arkansas*

Alexander Rothenberg§ *RAND Corporation*

Maisy Wong¶ *University of Pennsylvania*

February 2018

Abstract

Ethnic divisions complicate nation building, but little is known about how to mitigate these divisions. We use one of history's largest resettlement programs to show how intergroup contact affects long-run integration. In the 1980s, the Indonesian government relocated two million migrants into hundreds of new communities to encourage interethnic mixing. Two decades later, more diverse communities exhibit deeper integration, as reflected in language use and intergroup marriage. Endogenous sorting across communities cannot explain these effects. Rather, initial conditions, including residential segregation, political and economic competition, and linguistic differences influence which diverse communities integrate. These findings contribute lessons for resettlement policy.

JEL Classifications: D02, D71, J15, O15, R23

Keywords: Cultural Change, Diversity, Identity, Language, Migration, Nation Building

*We thank Alberto Alesina, Oriana Bandiera, Toman Barsbai, Giorgio Chiovelli, Raquel Fernandez, Paola Giuliano, Dilip Mookherjee, Nathan Nunn, Daniele Paserman, Ben Olken, Imran Rasul and seminar participants at Boston University, Georgetown University, Harvard University, the Kiel Institute, McGill University, MIT, Notre Dame, University of Colorado Denver, University of Southern California, University of Toronto, Wellesley College, the Barcelona Graduate School Summer Forum, the 2017 DIAL Development Conference, the 2017 Midwest International Development Conference, the NBER Political Economy meeting, the Ninth International Migration and Development Conference, the 2016 Annual European Conference of ASREC, and the Northeast Universities Development Consortium 2016 Conference for helpful suggestions. Maisy Wong is grateful for financial support from the Research Sponsors Program of the Zell/Lurie Real Estate Center and the Wharton Global Initiatives program. A previous version of this paper circulated under the title "Unity in Diversity? Ethnicity, Migration, and Nation Building in Indonesia." Jeremy Kirk, Gedeon Lim, JoonYup Park, and Xuequan Peng provided excellent research assistance. All errors remain ours.

†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.

[The] central challenge of modern, diversifying societies is to create a new, broader sense of 'we'.
—Robert Putnam, *The 2006 Johan Skytte Prize Lecture*

1 Introduction

Uniting people from diverse cultures is a founding principle of many nation states.[1] Throughout history, leaders have introduced policies to foster a national identity that would sustain an "imagined political community" in which citizens remain connected by shared history and values, despite never meeting one another (Anderson, 1983). However, with rising geographic mobility, there are concerns that growing local diversity may encourage a narrower sense of "we" and undermine this nation-building objective (Putnam, 2007).[2] The recent refugee crisis has also stoked debate over how to design resettlement policies to facilitate the integration of diverse groups (Bansak et al., 2018).

The key contribution of this paper is to show how local diversity influences integration and contributes to an intergenerational process of nation building. Social theorists offer competing views. Some argue that exposure to new cultures provokes backlash and may incite conflict (Blumer, 1958; Huntington, 2004). Others posit that negative sentiments may dissipate as intergroup relationships develop over time with greater contact (Allport, 1954). Alternatively, diversity may engender social anomie or isolation, which limits integration (Alesina and La Ferrara, 2000; Algan et al., 2016). Empirically, however, it is difficult to identify whether contact strengthens intergroup relationships in the long run because (i) identities and social relationships change slowly and tend to be confounded with time trends, (ii) diverse communities are often unstable due to tipping forces (Schelling, 1971), and (iii) persistent diversity is often confounded by favorable geography and endogenous sorting (Michalopoulos, 2012). We use a large-scale policy experiment in Indonesia to address these identification challenges.

Indonesia's Transmigration program, one of the largest resettlement efforts in history, provides an ideal setting to understand how intergroup contact can foster nation building. After independence, the government faced urgent pressures to forge an Indonesian identity that would unite diverse groups across the archipelago and overcome secessionist tendencies. Policymakers viewed resettlement as part of a broader effort to integrate more than 700 ethnolinguistic groups geographically segregated throughout history. From 1979 to 1988, the Transmigration program assigned two million voluntary migrants (hereafter, transmigrants) from the Inner Islands of Java and Bali to new settlements across the Outer Islands.[3] Each settlement was endowed with the same public resources and included a mix of Inner and Outer Islanders with the goal of weakening salient ethnic divisions through contact.

We exploit the haphazard assignment of transmigrants across settlements to isolate plausibly exogenous, long-run variation in local diversity. Institutional and capacity constraints limited planners' abilities to systematically assign transmigrants. Moreover, a lottery was used to distribute farm plots

[1] For example, "United in Diversity" is the motto for the European Union, *E pluribus unum* (out of many, one) for the United States, and "Unity in Diversity" for South Africa and Indonesia. There are numerous historical examples of efforts "to form French citizens" (Weber, 1976), "to make Italians" (Duggan, 2007), and to create "one kind of man, Indonesian" (Hoey, 2003).

[2] Alesina et al. (2017) and Miller (2012) discuss challenges of forging a shared identity within the European Union. More generally, migration pressures are growing among minorities within rich countries (see Frey, 2014, on the United States) and in newer migration corridors from poor to rich countries (Hanson and McIntosh, 2016). The U.S. is projected to become a majority minority nation by 2044 (Colby and Ortman, 2014) and the United Kingdom by 2066 (Coleman, 2010).

[3] The program had three goals: population redistribution, agricultural development, and nation building. In prior work (Bazzi et al., 2016), we investigate the agricultural productivity effects using a different empirical strategy than in the present study.

and assign housing to newly-arriving migrants. Imperfect land markets tied migrants to these initial plots, limiting *ex-post* sorting. Ultimately, the large scale of the program created nearly 900 communities along a continuum of policy-induced diversity. Using the 2000 Population Census, we show that even after two decades, these communities exhibit significantly greater ethnic diversity and less within-village ethnic segregation than other villages in the Outer Islands.[4] The persistence of many mixed Inner–Outer communities suggests that tipping did not neutralize the initial policy assignment. This allows us to study the effects of sustained interethnic contact in communities where local diversity did not arise as a result of endogenous sorting.

We further address endogenous sorting by developing an instrumental variables strategy that leverages the initial assignment. In particular, we instrument for the Inner-Island ethnic share in a village in 2000 using the initial stock of transmigrants assigned in the 1980s. Planners determined each settlement's potential population size by assessing the carrying capacity of available land using soil attributes and topography. Conditional on these predetermined natural advantages (and hence, on potential population), the larger the initial transmigrant stock, the greater the Inner-Island ethnic share today. We show that this strategy helps rule out endogenous *ex-ante* assignment and *ex-post* sorting of migrants to places with tolerant natives.

Our key measure of integration is the choice of language used at home, as reported in a 2006 household survey. Language is broadly seen across the globe as the most critical component of national identity (Pew Research Center, 2017). Indeed, policymakers view the national language, *Bahasa Indonesia* or Indonesian, as synonymous with Indonesian identity, widely promoting its use across economic and social domains.[5] Indonesian is rooted in the language of an ethnic minority (Malay), with as few as 5 percent speaking Malay when it was chosen as the national language in 1928. Although almost all Indonesians can speak it, less than 20 percent choose Indonesian as their primary language at home. Choosing to primarily speak Indonesian at home, regardless of one's initial motive, deepens the adoption of the national identity in this generation and the next, and ultimately advances nation building.

We complement language use at home with two other proxies for nation building: (i) a child's identification of Indonesian as her mother tongue from a 1995 auxiliary survey, and (ii) post-program interethnic marriage rates from the 2000 Population Census. Government officials viewed language and intermarriage as barometers for integration, and both measures are used as such throughout the literature on cultural change (e.g., Abramitzky et al., 2015; Giuliano and Nunn, 2017).

We draw upon the Lazear (1999) model of cultural change to explain how intergroup contact influences the intergenerational process of nation building. Individuals make integration choices according to preferences whose inputs include a baseline idiosyncratic utility and a national stock of identity "capital" common to all citizens. This national identity capital evolves with shared experiences—akin to models of brand preferences and habit formation—and can be reinforced by nation-building policies such as na-

[4]There are fourteen ethnic groups native to the Inner Islands, including the Javanese who represent 40 percent of the country's population. The Outer Islands have many distinct ethnic groups, and around 20 have more than 1 million members according to Population Census data from 2000.

[5]As Kramsch (1998) argues, "There is a natural connection between language and identity insofar as language often defines membership to a specific group to the exclusion of nonmembers. Through language the group manifests 'personal strength and pride' and a 'sense of social importance and historical continuity' and most of all belonging to an 'imagined community' that shares a common worldview and that commands allegiance to it. . . ." Simpson (2007a) notes, for example, that "Indonesian has also become positively valued as the primary shared component of the country's emerging national identity."

tional language promotion or compulsory schooling. Parents may choose to speak the national language at home to help their children develop economic or social relationships across groups. Over time, the more they speak Indonesian at home, the more their children build up the stock of Indonesian identity. Using auxiliary longitudinal survey data, we trace out this intergenerational process, linking parents' language and marriage decisions to children's national identity and integration later in life as adults.

Individual incentives to develop intergroup relationships should vary with relative group sizes, but this relationship can be nonlinear and ambiguous. To illustrate, consider a native Outer Islander in a village with an (exogenously-determined) Inner-Island share of either 10 or 90 percent. In the former, since Outer Islanders are dominant, she prefers to *segregate* and speak her native Outer-Island language. In the latter, where Inner Islanders are dominant, she may prefer to *assimilate* and speak the Inner-Island language at home. With a clear majority, she prefers a local ethnic language and is unlikely to *integrate* and speak Indonesian.

However, in diverse villages with a more equal split of Inner and Outer Islanders, it is unclear whether segregation, assimilation, or integration will prevail in equilibrium. On the one hand, each group may have enough co-ethnic members to sustain segregated communities. On the other hand, tipping the Inner-Island share above 50 percent gives rise to network externalities from assimilating with the majority group. Meanwhile, there are potential benefits of coordinating on a common national language. In particular, national integration could generate greater social surplus (above and beyond local assimilation) through spillovers to the broader community beyond the village.

Therefore, the relationship between Indonesian use at home and the Inner-Island ethnic share could have an inverted-U, flat, or U shape. Given the weak incentives to integrate at the two extremes (10 and 90 percent), the shape depends on which forces dominate in the middle. Diverse villages that foster intergroup contact can promote integration. Yet, it is precisely in such communities with two large groups where polarization dynamics are most likely to exacerbate divisions (Esteban and Ray, 1994). Facilitating conditions that encourage cooperative rather than conflictual equilibria in these settings is a major policy challenge.

We find a significant inverted-U shape for Indonesian use at home, suggesting that national integration is strongest in communities where Inner and Outer Islanders are in roughly equal proportion. This is consistent with the notion that, over time, intergroup contact can promote convergence towards a common national identity. The turning point around 45 percent Inner-Island ethnic share is relatively high, given that with roughly equal mixes, individuals can segregate with fellow co-ethnics. We find similar effects of diversity on national language use at home using other measures such as overall ethnolinguistic fractionalization (*ELF*) or *ELF* among Inner-Island groups.

Robustness checks support our interpretation of how diversity affects integration and nation building. For example, we use additional survey data from 1995 to show that the inverted-U shape is not explained by immigrants arriving after the initial year of settlement. Nor is it explained by intermarried parents or those reporting Indonesian as their mother tongue (and perhaps more predisposed to speak Indonesian). Moreover, we find similar inverted-U relationships for high and low education households, which suggests that ability differences are unlikely to confound the choice to use Indonesian at home. Finally, relaxing the exclusion restriction, the inverted-U relationship also appears in the reduced form.

We find similar inverted-U relationships for other measures of integration. First, we identify a strong

inverted-U relationship between the Inner-Island ethnic share and the claim to Indonesian mother tongue among children. This holds even for children whose mother does not speak Indonesian at home, capturing a deeper sense in which identity is changing among the next generation. Second, we find an inverted-U relationship for interethnic marriage, consistent with stronger intergroup relationships in mixed villages despite co-ethnic marriage markets being sufficiently large to maintain segregation.

We identify several policy-relevant mechanisms that explain why intergroup contact fostered nation building rather than conflict. These tests rely on the wide geographic scope of the program that scattered settlements across a variety of different contexts. First, the spatial configuration of settlements can affect integration by determining the likelihood of local intergroup contact. We find that when ethnic groups within villages are less segregated, national language use at home increases. The placement of villages themselves is also important, as diversity has more positive effects in more remote communities.

Second, we find that diversity's benefits can be shaped by the economic environment in important ways. To quantify the economic environment, we use a proxy for the degree of substitutability between Inner and Outer Islanders' skills. This measure of the similarity in agroclimatic characteristics between the migrants' origins in Java/Bali and their destination villages is an important determinant of agricultural productivity (Bazzi et al., 2016). Interestingly, we find an inverted-U shape in places with low agroclimatic similarity but a U shape in places with high similarity. This reversal suggests that diversity may discourage integration in the presence of competition between groups with similar skills.

Third, sociopolitical factors at the local and regional levels can determine how diversity affects integration. We find greater home use of Indonesian where the majority group (Inner or Outer) is itself more ethnically fragmented in the village and in diverse places where the linguistic distance between Inner and Outer Islanders is greater (hence, raising the cost to assimilate). Both conditions increase the returns to coordinating on the national language relative to assimilating with the local majority (Lazear, 1999).

Finally, we find stronger integration in places where transmigrants pose less of a regional political threat. Resettlement of politically dominant groups (e.g., the Javanese) in peripheral regions of the country is often viewed as threatening the political power of local ethnic minorities (see, e.g., Fearon and Laitin, 2011). We capture the extent of political competition by measuring whether the local native group is a dominant majority within the broader political unit. Our results suggest that the regional ethnopolitical balance is another crucial input in designing effective resettlement policies.

The final part of the paper assesses whether the Transmigration program was a successful vehicle for nation building at the local level. Our identification strategy here relies on counterfactual villages that were planned but abruptly halted due to budget cutbacks following a sharp drop in global oil prices in the early 1980s. We further refine the comparison using a reweighting approach common in place-based evaluations (Kline and Moretti, 2014). We find that individuals in Transmigration villages are around three times more likely to report Indonesian as the primary language at home. Moreover, intermarriage rates among young cohorts are three times higher than in control areas, and not all of this increase can be explained by greater diversity in the marriage market.

Our findings contribute to a nascent political economy literature on nation building (Alesina and Reich, 2015; Miguel, 2004). Many studies document adverse consequences of diversity (see Alesina and LaFerrara, 2005; Ray and Esteban, 2017, for a review), and another rich literature shows that members of the same ethnic group exhibit differing degrees of national affinity across, for example, the Niger–

Nigeria border (Miles and Rochefort, 1991). Yet, there is relatively little evidence on how diverse countries might effectively manage ethnic divisions.[6] By leveraging policy-induced variation, we identify how local diversity affects intergroup relationships in the long run and shed light on the social conditions that facilitate integration. These underlying mechanisms help inform policy debates over how to balance the competing effects of changes in diversity. Additionally, we clarify how a national language fosters convergence towards a new shared identity, echoing other work on language and nation building (Clots-Figueras and Masella, 2013; Fouka, 2016; Laitin and Ramachandran, 2015).

Moreover, we provide new insight into the intergenerational process of cultural change in diverse societies (Bisin and Verdier, 2011; Fernandez, 2011; Giuliano and Nunn, 2017). Putnam (2007) and Algan et al. (2016) argue that diverse places tend to be isolating. We show how the incentives to connect with other groups depend on local conditions and change nonlinearly with relative group sizes.[7] Our findings are consistent with the Desmet et al. (2016) model showing how local learning from intergroup contact can mitigate the ethnic divisions and conflict associated with aggregate diversity.

As resettlement programs and integration policies expand into new settings, our findings offer insight on the potential conditions in which ethnic diversity might be harnessed for improved social outcomes. Our evidence for long-run integration in diverse Transmigration villages remains striking given popular concerns that such large-scale resettlement was a classic case of cultural imperialism that would stoke ethnic conflict. Nevertheless, our findings resonate with theories of contact and cultural change and are consistent with a recent reappraisal of the program (Barter and Côté, 2015).

The rest of the paper proceeds in seven sections. Section 2 provides background on nation building efforts in Indonesia. Section 3 provides relevant details on the Transmigration program. Section 4 describes our main data sources on diversity and nation-building outcomes. Section 5 presents our core empirical results linking local diversity to identity formation and the intergenerational process of nation building. Here, we develop the main identification strategy and also discuss salient mechanisms underlying the main results. Section 6 presents estimates of the place-based impact of resettlement on language use and intermarriage in the Outer Islands. Finally, Section 7 concludes.

2 Nation Building in Indonesia

With a population of more than 250 million, Indonesia is the world's fourth most populous country. It is also among the world's most diverse, with more than 1,200 self-identified ethnic groups living on roughly 6,000 islands. The Javanese are the largest ethnic group, constituting 40 percent of the population, followed by the Sundanese with 15.5 percent (Ananta et al., 2013). Both groups originate from the Inner Island of Java. Nationally, Indonesia's index of ethnolinguistic fractionalization (*ELF*)—the proba-

[6] A few studies use (quasi-)experimental variation in diversity within college dormitories (Boisjoly et al., 2006), primary schools (Rao, 2018) or sports teams (Lowe, 2018) to show that intergroup contact fosters short-run increases in tolerance and out-group friendships. We differ in our focus on the long-run, intergenerational process of nation building and more persistent behavioral outcomes like marriage and language use. Other recent work examines how public media (Blouin and Mukand, 2016), bureaucrat assignments (Okunogbe, 2015), schooling (Bandiera et al., forthcoming), shared religious experience (Clingingsmith et al., 2009), and external enemies (Dell and Querubin, forthcoming) influence intergroup tolerance or nation building.

[7] Recent work examines cultural assimilation through language and marriage during the period of mass immigration in the United States (Abramitzky et al., 2016; Advani and Reich, 2015). In contrast, we highlight conditions under which groups are more likely to integrate towards a new shared identity, not unlike the forging of American culture by migrants on the historical frontier (Bazzi et al., 2017).

bility that any two residents belong to different ethnicities—is around 0.7, according to 2000 Population Census data. Despite such diversity, most Indonesians live in ethnically segregated communities. Of the more than 60,000 urban and rural villages in Indonesia, the median village has an *ELF* of 0.05.[8]

For most of its history, the peoples of the Indonesian archipelago were governed by a collection of independent kingdoms (*kerajaan*), many of which were separated from one another by immense waterways and dense vegetation. The absence of a common ruler, together with geographic isolation, enabled the persistence of many different cultures, religious practices, and languages throughout the region. The divide-and-rule strategy of the Dutch East India Company (VOC) pitted different kingdoms against each other and sharpened the distinctions among them. As such, by the end of the nineteenth century, the peoples of Indonesia had little shared history, apart from their experiences with Dutch colonialism.[9] For at least a decade after independence in 1945, political tensions threatened to derail nation-building efforts. Tensions often grew out of opposition to the increasing concentration of power in the capital, Jakarta, which many, particularly in the Outer Islands, associated with a growing dominance of the Javanese (Bertrand, 2004). These frustrations often coincided with ethnic grievances, fueling anti-Javanese sentiment and recurring secessionist threats from the Outer Islands (Mulder, 1996; Thornton, 1972).

This historical context presented Indonesia's political leaders with the problem of nation building. Anderson (1983, pp. 6–7) defines a nation as "an imagined political community . . . [where] members of even the smallest nation will never know most of their fellow-members, meet them, or even hear of them, yet in the minds of each lives the image of their communion". Nation building can be understood as the task of fostering this "image of communion"—where shared values and preferences among citizens would be strong enough to glue them together as a nation—through the promotion of a shared national identity (Alesina and Reich, 2015). This task is bound to be daunting in a nationally-diverse, yet locally-homogeneous country like Indonesia, especially given the little shared history among its peoples.

Officials took the first step toward this objective by politically agreeing on the meaning of a nation in the 1928 Second Youth Congress, where they pledged to create "*satu nusa, satu bangsa, satu bahasa*" (one fatherland, one nation, one language). This pledge defined the nationalist leaders' nation-building agenda, namely, the creation of a nation "unified by ties of common language, common outlook, and common political participation, a people enthusiastically severing its outworn ties to local traditions and loyalties . . . and the creation of . . . [an] all-Indonesian culture" (Feith, 1962/2007, pp. 34-35). National unity became one of the state ideology's Five Key Principles (*Pancasila*), and "*Bhinneka Tunggal Ika*" (Unity in Diversity) is the state motto inscribed on the nation's coat of arms. Indeed, as noted by Feith (1962/2007, p. 34), nation building "was probably the central goal which the nationalist leaders believed should and would be realized with the attainment of independence."

[8]The *ELF* index equals $1 - \sum_g \pi_g^2$ where π is the share of group g in the village. Villages (*desa* or *kelurahan*) comprise the lowest level of governance in Indonesia with an average population of over 2,000 (7,000) in rural (urban) areas in the early 2000s. They are the main administrative unit of analysis in our study.

[9]As noted by Ricklefs (2008, p.189): "In 1905. . . [a sense] of a common Indonesian identity or of common goals simply did not yet exist. Most Javanese, for instance, neither knew nor cared about what happened in Aceh, except for those who were fighting beside the Dutch to destroy its independence."

2.1 The Nation Building Process

Nation building is a process of promoting a shared national identity, where identity can be viewed as a type of "capital". Here, we clarify how intergroup contact at the local level can affect this process through its influences on individuals' social choices. We draw upon insights from the Lazear (1999) model on language and cultural assimilation, adding a common national identity that is persistent, following models of cultural evolution (Bisin and Verdier, 2011), habit formation (Becker and Murphy, 1988), and brand preferences (Bronnenberg et al., 2012).

To fix ideas, consider a discrete choice model in which individual i can choose between speaking her own native language at home (segregate) or a national language (integrate). Individual i's preference has an idiosyncratic component and a common component shared by all citizens in the country. This common component represents a stock of national identity "capital" that persists and builds upon past experiences. The more she speaks the national language, the more she (and her children) identifies as Indonesian, and the more important the national identity becomes as a determinant of social choices.

Individual i chooses the optimal language by weighing marginal costs and benefits. Costs depend on the linguistic distance between her native language and the national language. Benefits depend on the relative importance of idiosyncratic and national components of identity, as well as the potential gain from intergroup contact.[10] Benefits from contact include both economic and social exchanges that are better intermediated when all parties speak a common language. In a diverse village, she may choose to speak the national language if the utility gain from speaking it and from intergroup contact outweighs the costs of speaking a non-native language.

The insights are similar if we extend to many native languages and a common national language. In this case, the national language may be preferred if adopting it costs less than learning the other group's native language or if there are benefits for all groups to coordinating around a "neutral" language, such as in highly-fragmented villages with many minority languages (Lazear, 1999). There could also be spatial spillovers associated with speaking the national language (relative to the local majority's language), which increases the potential for economic and social relationships beyond the village.

Ultimately, the nation-building policy goal is to increase the stock of national identity "capital" to promote integration. In terms of language, this involves facilitating conditions that raise the benefits and reduce the costs of speaking the national language at home. Doing so not only increases the current stock of national identity capital but also the future stock as parents socialize the identity to their children, thereby fostering an intergenerational process of integration.

2.2 Identity and Language

National language policy can be an important vehicle to advance nation-building objectives.[11] Notably in the case of Indonesia, choosing a minority language as the national language anticipated some of its nation-building challenges. Almost all of the nearly 700 languages currently spoken in Indonesia belong to the Austronesian language family, but many are very different from one another, and the differences

[10] To the extent that interethnic contact engender conflict (see, for example, Esteban and Ray, 2011a), what matters is the net benefit of intergroup exchanges.

[11] This is a central theme of several edited volumes on the role of language policy in shaping national identity in a range of rich and poor countries across Europe (Barbour and Carmichael, 2000), Asia (Simpson, 2007b), and Africa (Simpson, 2008).

are particularly large across the Inner–Outer Island divide. The national language, *Bahasa Indonesia*, or Indonesian, is a modified version of Malay that originated along the eastern coast of Sumatra (and peninsular Malaysia) but had been used as a trading language in the archipelago for centuries. Prior to its recognition as the national language at the 1928 Youth Congress, Malay was only spoken as the native language by 5 percent of the population living under Dutch colonial authority, whereas nearly 40 percent spoke Javanese. By unanimously choosing a minority language, the delegates of the congress avoided the resistance of non-Javanese ethnic groups and signaled their commitment to political unity. Its status as the national language was later cemented in the 1945 Constitution.

Subsequent policies leveraged the national language as part of broader nation building efforts. Widely viewed as "a symbol of national unity and identification" (Sneddon, 2003), policymakers promoted widespread adoption of the Indonesian language, in the hope that "the more [the Indonesian people] learned to express themselves in Indonesian, the more conscious they became of the ties which linked them" (Alisjahbana, 1962). Indonesian was established as the language for official communication and was incorporated in the national curriculum of a rapidly expanding education system (Suryadinata, 1988). Many outsiders view Indonesia's national language policy as exemplary (Simpson, 2007a).[12] Nevertheless, despite nearly universal knowledge of Indonesian today, less than 20 percent of households use it as the primary language at home, according to the 2010 Census.

2.3 Nation Building Through Transmigration

Some policymakers also saw the nation-building potential of resettlement programs. Planners hoped that the Transmigration program would encourage intergroup contact between Inner and Outer Islanders and promote the formation of a national identity (Kebschull, 1986; MacAndrews, 1978).[13] In speeches and policy documents, officials allude to the program's role in enhancing integration. For instance, in 1985, the Minister of Transmigration, stated "By way of transmigration, we will try to ... integrate all the ethnic groups into one nation, the Indonesian nation. The different ethnic groups will in the long run disappear because of integration and there will be one kind of man, Indonesian" (Hoey, 2003).

However, the program reignited suspicions of a "Javanization" agenda in the Outer Islands (see, e.g., Hoshour, 1997; Mangunrai, 1977). Echoing discord from the early days of independence, there were questions of whether Transmigration was a vehicle for cultural imperialism over Outer-Island ethnic groups or a way for Suharto's government to solidify power in frontier regions (see, e.g., Charras and Pain, 1993; Levang, 1995). These concerns reflect the common unease among indigenous, "sons of the soil" minority ethnic groups experiencing rapid immigration of majority ethnic groups from the political and economic center of the country. From a nation-building perspective, though, it is important to note that the Javanese and other native Inner-Island groups were not more likely to be standard bearers of Indonesian identity at the time of the program. This is important to keep in mind as we aim to

[12] To cite Paauw (2009): "[No] other post-colonial nation has been able to develop and implement a national language with the speed and degree of acceptance which Indonesia has. No other national language in a post-colonial nation is used in as wide a range of domains as Indonesian, a feat made more impressive by the size and ethnic, linguistic and cultural diversity of Indonesia."

[13] Discussing the program objective, World Bank (1988, p. 5) noted: "[Transmigration] has been seen by national leaders as a tool for national integration ... [and] as a means of promoting cultural contact and building national unity."

understand how changes in diversity affect convergence toward new forms of shared identity.

3 Transmigration: Program Design and Implementation

Our analysis focused on the diverse communities created by the Transmigration Program during the third and fourth Five-Year Development periods (or *Pelita*) from 1979–1988 under Suharto. During this period, the Suharto government massively expanded the program to relieve perceived over-population pressures in the Inner Islands and underdevelopment in the Outer Islands. The program provided free transport to the newly created settlements as well as housing and two-hectare farm plots. A total of 1.2 million people were resettled in *Pelita* III, and an additional 3.75 million people were planned to be resettled in *Pelita* IV.

Transmigrants volunteered for the program. The program targeted nuclear 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 limited schooling (Kebschull, 1986). These government-sponsored migrants are more comparable to stayers than to typical voluntary migrants. On average, Java/Bali-born individuals who moved to Transmigration villages had around 0.7 fewer years of schooling compared to stayers in their origin district in Java/Bali (based on the 2000 Population Census). By comparison, those who moved to urban areas or other rural areas of the Outer Islands had 2.4–3.2 more years of schooling compared to stayers. Nevertheless, our main analysis compares transmigrants across resettlement villages, thereby differencing out the common component of selectivity across program participants. Below, we describe how the program provides plausibly exogenous variation in ethnic diversity across Transmigration settlements with further details on the program described in Online Appendix C.

3.1 Transmigrant Assignment and Ethnic Diversity

The program's rapid expansion in *Pelita* III and IV contributed to an as-if-random initial assignment of transmigrants, as planners endeavored to meet the annual target of settlers determined by the central government. As a result of the scale and pace of implementation, institutional coordination issues and information problems were rife, with many reports describing a "plan-as-you-proceed" approach (World Bank, 1988). Coordination problems between government agencies made it infeasible to systematically match transmigrants to settlements. One agency was responsible for recruiting transmigrants in the Inner Islands, while another was tasked with clearing sites in the Outer Islands. The allocation of transmigrants was driven by the coincidental timing of transmigrants' arrival to transit camps in Java/Bali and the opening of settlements in the Outer Islands. Participants could not choose their destinations and were often ill-informed about the conditions they would face (Kebschull, 1986).

Moreover, official guidelines stipulated that a certain portion of land in each new village was to be allocated specifically to Outer Islanders from nearby areas within the same province. This generated further variation in initial diversity. In 1979, this share—officially known as *Alokasi Pemukiman bagi Penduduk Daerah Transmigrasi* (APPDT)—was stipulated by the government to be 10 percent, and it was increased to 20–25 percent in 1982. However, in practice, these shares tended to vary across locations

(Clauss et al., 1988; Rigg, 2013).[14]

Once transmigrants and APPDT natives arrived, land and housing *within* settlements were assigned by lottery. Over time, some settlements merged with adjacent villages, and others were split along hamlet boundaries (the next administrative unit below the village). Settlements fell under the jurisdiction of the National Ministry of Transmigration, which enforced centralized rules governing the formation of villages, all of which would be endowed with the same initial institutions and public goods.[15] Together with the lottery, these rules help mitigate the concern that village boundaries are endogenously correlated with ethnic relations, an issue we revisit in robustness checks in Section 5.3.[16] As we show in Section 5.1, these program features helped induce persistent diversity across Transmigration villages that we use to study the long-run effects of sustained interethnic contact.

3.2 Site Selection and Carrying Capacity

According to the 1978 Transmigration Manual, planners were concerned about food security and developed models to estimate the carrying capacity of each settlement in terms of population. The goal was to ensure transmigrants could eventually use their two-hectare farm plots to produce enough food to overcome subsistence. Officials worked with agricultural experts to map the elevation, vegetation, soil types, hydrology, climate, market access, and locations of settlements. They determined the potential population of each settlement based on these site selection characteristics and the availability of arable land. In Section 5.1, we exploit this feature to develop an instrumental variable (IV) strategy that uses the initial number of Inner-Island transmigrants to instrument for the current Inner-Island ethnic share.

Before establishing new Transmigration settlements, planners used a three-stage process to select potential sites that were sparsely populated and had agricultural potential. First, potential regions were identified using large-scale maps capturing basic information about topography, market access, and existing settlements. Second, aerial reconnaissance identified "recommended development areas" (RDAs) based on agroclimatic conditions and nearby indigenous populations. Finally, local surveys of these conditions helped determine the total number of transmigrants to be allocated to the settlement. In Section 6, we use a place-based evaluation strategy and a subset of these planned settlements (RDAs) as counterfactuals to estimate the average impact of the program on demographic change and nation building in the Outer Islands.

[14] While there is limited information on APPDT, anecdotal evidence suggests there is some *de facto* geographic variation. For example, Tanasaldy (2012, p. 191) notes cases in West Kalimantan where villages had 50 or even 80 percent. The 2000 Population Census data bears out this variation, revealing an average 15 percent gap in the Outer-Island ethnic share between villages created during each period along with additional mean differences across broad regions.

[15] During this time, the creation of new villages followed the 1979 Village Law, a set of rules stipulated by the centralized government to ensure that most villages have minimum population sizes, similar institutions, and access to a basic set of public goods. By comparison, as part of a 2004 Village Law revision, adjustments to village boundaries could arise with more locally tailored adjustments and hence greater scope for endogenous ethnic segregation.

[16] Without the initial lottery assignment, endogenous settlement could lead to diverse hamlets (with tolerant people) and also segregated ethnic enclaves. When hamlets mature into villages, the initial settlement could then give rise to a spurious correlation between ethnic shares and unobserved tolerance.

4 Data

This section discusses our main data sources. First, we use a newly digitized 1998 Transmigration Census and planning maps to identify settlements and initial assignments. Second, we draw upon the 2000 Population Census to measure ethnic diversity and residential segregation. Third, we draw upon multiple datasets to measure social choices capturing integration. These include national language use at home as well as interethnic marriage and linguistic identity among children. Finally, we provide empirical evidence to motivate these outcomes as proxies for the intergenerational process of nation building. Appendix D provides further background on the data and variable construction.

4.1 Transmigration Census

To identify Transmigration villages, we digitized the 1998 Transmigration Census, produced by the Ministry of Transmigration (MOT). For each settlement, this dataset provides the number of transmigrant individuals and households assigned by the program as well as location information. Using these data, we identify 911 Transmigration villages (outside of Papua) established from 1979 to 1988. These villages received between 350 and 8,500 transmigrants (an average of 1,872) in the initial year of settlement. We also rely on several geospatial data sources (detailed in Appendix D) to measure the characteristics used to determine carrying capacity of settlements.

As shown in Figure 1, Transmigration settlements are scattered throughout the Outer Islands. More than half of these sites are located on the island of Sumatra, but many are found on Kalimantan and Sulawesi. This allows for a within-island analysis that accounts for the vast socioeconomic differences across these large island groups.

4.2 Ethnic Diversity

Our primary measure of ethnic diversity is the share of individuals in a village that report an Inner-Island ethnicity. This measure is directly influenced by the program, is the leading source of variation in diversity within Transmigration settlements, and captures one of the primary ethnic cleavages in Indonesia. The Inner–Outer divide is salient in Transmigration settlements, with natives referring to transmigrant ethnic groups from Java/Bali as *pendatang*, or newcomers. To construct measures of diversity, we draw on the universal coverage 2000 Population Census microdata that reports a single ethnic identity for every individual at a fine geographic level (census blocks).[17] Transmigrants and their descendants comprise nearly all of those reporting Inner-Island ethnicities in Transmigration settlements.[18]

Overall, the variation in ethnic diversity in Transmigration villages is explained largely by the Inner–Outer divide and ethnic divisions among transmigrants. We show this in Appendix Figure A.1, which depicts a stark inverted-U relationship between the Inner-Island ethnic share and overall ethnic diversity

[17] The 2000 Census allows individuals to self-identify with one of more than 1,000 different ethnicities. The 14 Inner-Island groups include all of those native to Java/Bali with the top four—Javanese, Sundanese, Madurese, and Balinese—comprising nearly 99 percent of Inner Islanders in study areas. Meanwhile, the top 50 Outer Island ethnicities in study villages comprise around 80 percent of Outer Islanders in these areas. Although many Outer Island groups are relatively small nationally, they are relatively large in many Transmigration areas. Note that the 2000 Census round was the first Census since 1930 to report ethnicity, and that the publicly available 2010 Census data only reports information on ethnicity at the district level.

[18] According to a Shapley decomposition, those born in Java/Bali explain 51 percent of the variation in the Inner-Island ethnic share while their second-generation kin—those born within the same district as reported in the Census—explain 41 percent.

captured by the *ELF* index. Using a Shapley decomposition, we find that the (quadratic) Inner-Island ethnic share explains 70 percent of the variation in the *ELF* index. Another 25 percent is explained by ethnic divisions among Inner-Island ethnicities (i.e., ELF_{Inner}) with the small remainder explained by divisions among Outer-Island ethnicities (i.e., ELF_{Outer}). These patterns suggest that diversity in a typical transmigrant village can primarily be explained by the influx of Inner Islanders. Local APPDT natives are typically a more homogenous group hailing from a smaller set of nearby Outer-Island villages than the transmigrants who, in a typical settlement, come from 45 origin districts across Java/Bali.

While we primarily focus on the Inner–Outer ethnic divide in Transmigration settlements, we also leverage the full granularity in ethnic identity, focusing on *ELF* and ELF_{Inner} in additional checks. Finally, as a way of proxying for the extent of local contact, we also study residential segregation between Inner and Outer Islanders. To measure segregation, we use the standard Bell (1954) isolation index applied here to census blocks within villages. This measure exploits fine block-level geographic detail to study within-village segregation patterns. Appendix Table D.2 provides summary statistics for these and other measures.

4.3 Integration and Nation Building Outcomes

We consider three outcomes aimed at capturing local integration as part of the long-run nation building process: language use at home, children's mother tongue, and interethnic marriages. These measures are in line with the view of language and marriage decisions as leading indicators of culture (see Abramitzky et al., 2015; Giuliano and Nunn, 2017).

Data and Measurement. Our main nation-building outcome is the use of Indonesian *at home*. Language use at home often reflects a fundamental socialization decision by parents to inculcate specific cultural preferences in children. By primarily using Indonesian at home, parents instill in their children a common national identity, instead of their respective ethnic identity. Importantly, unlike language ability or daily use, home language use is less likely to reflect purely economic motives.[19] We also observe whether individuals choose to primarily speak a different ethnic language at home. By revealed preference, Indonesian use at home can therefore be viewed as a choice to invest in a common national identity relative to ethnic identities. From the nation building perspective, though, it is less important to distinguish whether national language use at home was initially motivated by economic or social incentives. What matters is how policy-induced diversity shapes that choice.

For the main analysis, we use individual-level data from the 2006 National Socioeconomic Survey (known as *Susenas*), which asks the household head about his/her ethnicity and primary language used

[19] According to the 1995 Intercensal Population Survey described below, 95 percent of individuals in Transmigration areas reported ability to use Indonesian. One potential concern, though, is that parents choose to speak Indonesian at home primarily as a means of helping their children succeed in primary school. Beeby (1979) suggests that even as early as the 1970s, this would not have been a key concern: "Recent systematic research on students' achievements in four subjects in grade 6 throughout Indonesia has confirmed that constant or frequent use of *Bahasa Indonesia* in the home does give some advantage to students at this level; but the difference between them and the rest is much less than one might have expected. The advantage is, naturally, greatest in the language test, but disappears in the tests on mathematics and science." Moreover, to the extent that these parental inclinations are unrelated to diversity, this concern will not confound our interpretation. Below, we find similar results for household heads with high and low levels of education, which is also consistent with home language choices reflecting cultural values and preferences rather than mere ability.

at home. To increase the number of village observations given the limited sampling frame of *Susenas*, we construct a sample that includes villages just beyond the original Transmigration settlements.[20] Our analysis is ultimately based on 2,126 household heads in 135 villages, 22 of which constitute original Transmigration settlements and the remainder being within 10 km of those settlements.

Additionally, we use individual-level data from the 1995 Intercensal Population Survey (known as *Supas*) to examine social and familial transmission of identity. *Supas* has greater coverage than *Susenas* and also allows us to explore intrahousehold heterogeneity in primary home language as well as mother tongue. Notably, household members aged 5 and older respond to these language questions individually. Our *Supas* sample comprises 28,532 individuals in 422 villages, including 109 original Transmigration settlements and the remainder coming from within 10 km of those settlements. With *Supas*, we can address concerns about self-selection and endogenous sorting using its detailed migration questions. Its drawback, which prevents us from using it in our main analysis, is that the survey was conducted five years prior to ethnicity being collected in the 2000 Census. To account for this, in all analysis based on *Supas*, we use measures of ethnic diversity that are calculated based only on the 2000 village residents in the Census that do not report living elsewhere in 1995.

Finally, we investigate intermarriage rates between Inner- and Outer-Island ethnics. Intermarriage has long been viewed by social scientists as a key marker of assimilation (see Gordon, 1964), and as Babcock (1986) notes, officials in the Ministry of Transmigration monitored marriage between Inner and Outer Islanders as a barometer for successful integration. We use the Population Census to calculate intermarriage rates for 932,642 people spanning 832 Transmigration villages. For each household head, we can identify their marital status as well as the ethnicity of their co-resident spouse. We focus on young cohorts plausibly married after the program began.

Nation Building in the Long Run. We marshal evidence from two independent surveys with measures of group affinities to argue that national language use at home and growing up with multiethnic parents contribute to an intergenerational process of nation building.[21] First, using the 2009 *Asian Barometer* survey, a cross-sectional analysis shows that home use of Indonesian is associated with a stronger national affinity. Conditional on age, gender, education and region fixed effects, individuals that primarily or exclusively speak Indonesian at home are 10 p.p. more likely to choose the national identity over their ethnic identity, relative to a mean of 63 percent.[22]

Second, using panel data from the Indonesia Family Life Survey (IFLS), we show how Indonesian use at home and intermarriage may improve national unity and weaken ethnic attachment. The rich

[20] The survey covers a random sample of around 15–20 percent of villages. The samples are drawn proportional to district population size, and because Transmigration settlements are in underdeveloped areas, they tend to be undersampled relative to a simple random draw of villages. Our sample expansion is supported by evidence in Appendix B.II showing that the Inner-Island ethnic population exhibits significant spillovers to villages within 10–15 km of the original Transmigration settlements in the 1998 MoT Census.

[21] Although both surveys contain several rich outcomes of interest, their geographic coverage is either too sparse or unavailable at the village-level, which is required for empirical analyses of Transmigration settlements scattered across the archipelago. Moreover, *World Values Survey* data used, for example, by Desmet et al. (2016) also has questions on national affinity, intergroup tolerance, and preferences, but it does not include geographic identifiers at the village level and only has very coarse measures of Indonesian ethnicities.

[22] The question is stated as follows: "Let us suppose you had to choose between being an Indonesian and being [own ethnic group], which of these do you feel most strongly attached to?" The declarative responses include "Indonesian", "Own Ethnic group", and "Another Identity".

outcomes and long panel of the IFLS allow us to follow children growing up in households where Indonesian is spoken at home and observe their language and identity choices as they form new households more than a decade later. To do this, we begin by taking the sample of individuals observed in 1997 that subsequently formed new households by 2014. We construct an indicator for whether they spoke Indonesian daily at home during their childhood. Of all of the 1997 households in the IFLS, only 27 percent did so. We further control for whether the child's parents are from different ethnic groups. We then correlate these indicators with four nation-building outcomes that capture an array of social identity decisions and preferences in 2014. We control for age, gender, education, and village fixed effects to focus the comparisons across observably similar individuals in the same local community.

Overall, the results in Panel A of Table 1 consistently reveal that children who grow up speaking Indonesian at home tend to have significantly weaker ethnic attachments and are more likely to embrace the national identity. Individuals growing up in such households are nearly 50 percent more likely to speak Indonesian at home in their new households (column 1). They are more likely to self-identify with a different ethnicity in 2014 than in 1997, reflecting a more fluid self-concept of ethnic identity (column 2). Additionally, they are around 55 percent more likely to marry someone who is not a member of their own ethnic group (column 3) and are significantly less likely to show discriminative trust toward co-ethnics (column 4). Panel B shows that these patterns are not driven solely by individuals that grew up in multiethnic households. Controlling for parental interethnic marriages leaves the coefficient on daily Indonesian use at home as a child unchanged, while itself entering positively and significantly across all four outcomes. Together, these patterns show how using Indonesian at home may help break down ethnic attachments and hence play an important role in socializing new forms of shared national identity across generations.[23]

5 Local Diversity and Nation Building

In this section, we use variation in ethnic diversity across Transmigration areas to identify the long-run impact of local diversity on national integration, as proxied by the choice of home language. We organize our analysis as follows. First, we develop an identification strategy exploiting initial transmigrant assignments. Second, we present baseline results on diversity and language use at home. Third, we offer an array of identification and robustness checks. Fourth, we present several results clarifying the conditions under which diversity fosters integration. Finally, we document the role of community and familial diversity in shaping the intergenerational transmission of national identity.

5.1 Estimating Impacts of Diversity on Integration

To capture the relationship between local diversity and integration, we estimate the following semiparametric regression function:

$$y_{ij}^{\ell} = \alpha + g\left(diversity_j\right) + \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_{ij}, \tag{1}$$

[23] Desmet et al. (2017) offer rich insight into the distinction between cultural values and ethnic identity. Motivated by the IFLS results, we view ethnicity as a choice that evolves over time instead of a permanent group affiliation.

where y_{ij}^{ℓ} is an indicator for whether individual i in Transmigration location j primarily speaks language ℓ at home, $g(\cdot)$ is a function of Inner-Island ethnic share or overall *ELF* in 2000, and $\mathbf{x}_j$ is a vector of controls. We estimate equation (1) semiparametrically, following Robinson (1988). We use a local linear estimator with the Fan and Gijbels (1996) rule-of-thumb (optimal) bandwidth and an Epanechnikov kernel, but results are similar using other specifications.

In general, estimating this relationship with observational data presents several empirical challenges. First, people's strong tendency to self-segregate into ethnic enclaves often results in limited intergroup contact and limited variation in local diversity (McPherson et al., 2001). Second, diverse communities tend to be confounded with other omitted variables, such as market access, favorable environments, or a history of harmonious intergroup relations. Third, intergenerational identity transmission tends to be a slow process and confounded by time trends.

The Transmigration program allows us to address these empirical challenges. Features of the program discussed in Section 3.1 generated a continuum of ethnic diversity across the newly constructed villages, while migration frictions and land-market imperfections sustained these initial diversity shocks.[24] Figure 2 summarizes this persistent variation by plotting the kernel density of village-level *ELF* across Transmigration program (solid line) and non-program (dashed line) villages. Non-program villages tend to be less diverse; their *ELF* distribution is skewed to the right and has a relatively low mean. By contrast, the kernel density for Transmigration villages reveals a continuum of diversity.[25] Notably, there is a significant mass of Transmigration villages with relatively high *ELF*. In typical settings with greater labor mobility, segregation and tipping forces would have rendered such diverse places unstable.

To deal with potential endogeneity of diversity in 2000, we develop an instrumental variables (IV) strategy that isolates the initial, policy-driven variation in ethnic shares. We use a flexible function of the number of transmigrants placed from 1979–88 within 10 km of village j as an IV for the Inner-Island ethnic share in village j. To proxy for the policy rule determining potential population in each settlement, we control for characteristics used by planners to determine a site's carrying capacity such as soil attributes and topography (i.e., the $\mathbf{x}$ vector, see Section 3.2). Conditional on these characteristics and hence predicted potential population, a larger initial stock of transmigrants implies a higher Inner-Island ethnic share in 2000. By capturing *ex-ante* diversity, the IV helps rule out endogeneity arising from sorting after the initial program allocations. When considering overall diversity (*ELF*), we can augment the instrument set to include the *ELF* index among Inner-Island ethnicities, which is both plausibly exogenous and explains the bulk of remaining variation in the overall *ELF* as noted in Section 4.2.

Figure 3, estimated using a flexible first stage, shows that the initial assignment of transmigrants strongly predicts Inner-Island ethnic shares in 2000. This strong relationship is consistent with barriers to mobility making it harder for settlers to leave their initially-assigned communities. Together, these frictions limited tipping, as evidenced by the roughly (log-)linear relationship. Moreover, ethnic residential patterns *within* villages are similarly persistent. The isolation index for the older generation of household heads (i.e., initial settlers assigned houses by lottery) is correlated at around 0.9 with that of

[24] Because transmigrants were mostly landless laborers, the gift of land may have tied many of them to the new settlements. Migrants did not receive property rights immediately, and land sales were also limited by a 10 percent transaction tax (World Bank, 2008), presenting additional barriers to *ex post* migration as found in other contexts (see De Janvry et al., 2012).

[25] Figure A.2 presents similar patterns for residential segregation across ethnic groups within villages, where Transmigration villages exhibit significantly less segregation between Inner- and Outer-Island groups than do non-program villages.

the younger generation (i.e., initial settlers' children in new households).[26] This persistence is largely explained by typical inheritance norms in rural areas where land is passed on to children who then form their own households in plots adjacent to their parents.

The initial-transmigrants IV is not only strong but also plausibly excludable. The key identifying assumption is that planners did not initially create more diverse settlements in communities that were unobservably more receptive to integrating. Figure 4(a) provides initial supportive evidence, showing that our instrument is orthogonal to the linguistic similarity between the average Inner Islander and the indigenous Outer-Island group. Linguistic similarity adds a distance metric to ethnic differences and is based on language classification trees reported in *Ethnologue* (see Appendix D). The mostly flat curve mitigates the concern that planners systematically sent transmigrants to destinations based on linguistic and hence cultural proximity to transmigrants. Moreover, Figure 4(b) shows that the instrument is orthogonal to post-program immigration between 1995 and 2000. This addresses the concern that planners created more diverse communities in places that naturally attract diversity-seeking, tolerant individuals. Further results supporting instrument validity and consistent reduced-form evidence are presented below. We turn now to our baseline results.

5.2 Diversity and Language Use at Home

Figure 5 reveals a significant inverted-U relationship between national language at home and the share of Inner-Island ethnicities in the village. As discussed, the shape could be an U, inverted U, or a flat line, depending on whether the incentives to integrate, segregate, or assimilate prevail in the middle. An inverted-U shape is consistent with strong incentives to segregate or assimilate when there is a clear majority and strong incentives to integrate in places with an equal mix. Indeed, the fraction of the population speaking Indonesian as the primary language at home increases with the proportion of Inner-Island natives residing in the village but starts to decline right above 40 percent. The parametric test due to Lind and Mehlum (2010) suggests that the turning point and inverted-U shape are significant at the 1-percent level. Reassuringly, a similar analysis using the 1995 *Supas* data for another subset of Transmigration areas also shows an inverted-U relationship (Appendix Figure A.5).

The roughly symmetric shape reflects similarity between Inner and Outer Islanders in the elasticities of Indonesian use with respect to diversity.[27] For example, if Outer Islanders were more responsive to using Indonesian at home as diversity increases, we should see a steeper slope to the left, where the Outer Islanders are the majority. The symmetry and inverted-U relationship are inconsistent with the views that ethnic polarization is conducive to social disintegration, and the Transmigration program was simply a vehicle for Javanese cultural imperialism.

We further clarify the origins of the inverted U by analyzing language use at home separately for Inner and Outer Islanders in Figure 6. The top panel restricts to Outer Islanders only. The three panels from the left to the right reflect an Outer Islander's choice to speak Indonesian (integrate), an Inner Island language (assimilate), or an Outer Island language (segregate). First, the top-left panel shows

[26] We define the younger generation as those individuals who were under 16 years of age during the period of settlement in the 1980s, but results are unchanged using other cutoffs.

[27] The slightly flatter slope and higher intercept on the left is consistent with the higher levels of overall diversity among villages in the lower tail of the Inner-Island ethnic share compared to the upper tail, as seen in Appendix Figure A.1. The role of other ethnic divisions in explaining language use is made explicit in other results below based on the overall *ELF* index.

that Indonesian use at home among Outer Islanders increases with more exposure to Inner Islanders. This line does not flatten, which would be the case if the incentive to integrate dissipates once they have enough co-ethnic neighbors to sustain a segregated community. The linear patterns for Indonesian use at home by both groups explains the symmetric inverted U when pooling groups together. Moreover, the linearity for both groups, which we cannot reject using the Härdle and Mammen (1993) test, suggests that the inverted-U shape is not an artifact of more tolerant types from each group being disproportionately likely to reside in diverse communities. Otherwise, we would find inverted-U shapes for each group separately as well.

In the middle panel, we see that the Outer Islander's choice to assimilate and speak the Inner-Island language increases nonlinearly at around 50 percent, consistent with network externalities. Native Outer Islanders begin to take on Inner-Island languages at home once Inner-Island ethnic groups constitute the majority in the village. Finally, the top-right figure shows a clear linear and monotonic relationship with more Outer Islanders choosing to segregate and speak their own language at home as their share increases. This pattern suggests a moderate form of ethnic attachment. If villagers immediately self-segregate once they have enough co-ethnics, the propensity to speak their own language should increase nonlinearly towards one, mirroring the inflected shape for the out-group language. The bottom panel shows analogous patterns of language use at home among Inner Islanders. Once more, analysis using the 1995 *Supas* data reveals broadly similar patterns (see Appendix Figure A.6).

Table 2 presents parametric OLS and IV estimates of the main inverted-U relationship. We begin in columns 1 and 2 with a quadratic specification of the Inner-Island ethnic share.[28] Column 1 presents the parametric version of the partially-linear results in Figure 5. In column 2, we approximate a nonlinear first stage by instrumenting for the Inner-Island ethnic share using ventiles of the number of initial transmigrants within 10 km (i.e., 20 dummy variables with equal numbers of villages in each bin). These instruments provide a relatively strong first stage and satisfy overidentifying restrictions.[29] The implied turning point for the inverted U is similar between the OLS and IV estimates. Overall, the similarity between OLS and IV estimates points to the persistent impact of the initial wave of settlers on diversity levels observed up to two decades later in these newly-formed communities. To provide a sense of magnitudes, we estimate analogous specifications in columns 3 and 4 based on terciles of the Inner-Island ethnic share. In the middle tercile, households are 18.7 p.p. more likely to use Indonesian at home than in the bottom and top tercile. This is a large differential relative to the mean of 25 percent.

Reduced form results provide additional support for these findings. Figure 7 reveals an inverted-U relationship between Indonesian use at home and the initial transmigrant assignment, using the same semiparametric approach as in the OLS specifications above. Column 7 of Appendix Table A.1 corroborates this pattern using a parametric tercile specification.

Together, the results thus far point to a causal effect of ethnic diversity on integration at the local level.

[28] This is supported by a Härdle and Mammen (1993) test that fails to reject the null hypothesis of a quadratic shape in Figure 5.

[29] We use generalized method of moments (GMM) to account for the many instruments and cluster standard errors at the district level, of which there are 50. At the bottom of the table, Sanderson and Windmeijer (2016) Wald statistics reject the null of weak instruments on the two endogenous variables. Instrumenting with a smaller number of bins$\geq$10 yields similar albeit slightly noisier results. Based on the Hansen (1982) test, we cannot reject the null hypothesis that the instruments are uncorrelated with the error term and are correctly excluded from the second stage. Coupled with the rejection of the null under the Anderson and Rubin (1949) test (that the coefficients on the endogenous variables jointly equal zero and the overidentifying restrictions are valid), these diagnostics point to a fairly well-specified IV model.

Appendix Table A.1 shows that this core finding is robust to different diversity measures. First, we find that different measures of ethnic fractionalization (*ELF*)—defined based either on the dichotomous Inner-/Outer-Island ethnic categories (column 3) or the full set of ethnic categories (column 4)—also exhibit a strong positive relationship with national language use at home. Finally, columns 5 and 6 exploit plausibly exogenous variation in ethnic diversity among the Inner-Island ethnic population arising from the arbitrary allocation process that sent transmigrants from many different Java/Bali origins to the same transit camps (see Appendix C).[30] The *ELF* index among Inner-Island groups in column 5 and the quadratic Javanese share among Inner-Island groups in column 6 both point to greater home use of the national language in the most diverse communities.

Overall, the findings in this section provide new evidence on the relationship between local diversity and nation building. We now subject these results to a host of robustness and identification checks that not only strengthen the credibility of these findings but also bolster our interpretation of how diversity influences national identity formation and cultural change.

5.3 Addressing Threats to Identification

This section rules out several potential threats to identification. In Table 3 and accompanying results in Appendix A, we provide a series of robustness checks to rule out confounding explanations for national integration in diverse Transmigration areas. As a reference, column 1 of Table 3 reproduces the baseline IV specification for the quadratic Inner-Island ethnic share from Table 2.

Endogenous Initial Assignment. In column 2, we show that the inverted-U relationship is robust to including fixed effects (FE) for the ethnolinguistic homelands native to each village. These FE restrict the identification to be across villages within the same broad ethnolinguistic region reported in *Ethnologue*. This reduces the concern that the inverted U is driven by more diverse settlements being created in regions where the local ethnic group is more culturally similar to and hence tolerant of Inner-Island ethnic groups.[31]

As further evidence for the IV exclusion restriction, Appendix A.II documents the lack of correlation between the IV and predetermined district-level measures of political and economic development not explicitly used for site selection but potentially known to planners at the time. These include home use of Indonesian in the late 1970s as well as other variables potentially correlated with both diversity and national identity in areas near the eventual settlements but prior to their creation. These orthogonality results are unaffected by the inclusion of year-of-settlement fixed effects, which suggests that planners did not systematically adjust the assignment mechanism over time. The IV is also uncorrelated with contemporaneous religious diversity as well as the skill and occupational mix of the older Inner-Island population. This latter result is consistent with column 3 in Table 3, which shows

[30] The OLS results in columns 3 and 4 survive IV specifications where in column 3 the instruments include the same 20 dummy variables as in Table 2, and in column 4 we augment that instrument set to include the *ELF* among Inner-Island-born individuals, thereby capturing the additional diversity induced by the haphazard allocation process. Additionally, Appendix Figure A.7 reveals a nonlinear relationship between overall ethnic diversity and national language use at home. The sharpest increase occurs in villages where there is more than a 50 percent chance of encountering a non-coethnic neighbor, suggesting that ethnic mixing above and beyond the Inner–Outer division generates further integration.

[31] This also mitigates concern that planners assigned greater diversity to settlements in coastal areas of Sumatra and Kalimantan, home to the Malay ethnic groups whose native language is the one upon which Indonesian is based.

that the inverted-U relationship between Indonesian use at home and Inner-Island ethnic share is robust to including an exhaustive set of occupation fixed effects. There is, therefore, little evidence to suggest that policymakers systematically created more diverse settlements in initially more well-suited locations.

Clarifying the Role of Baseline Controls. Given that we rely on non-experimental variation in diversity, it is important that our results do not depend entirely on the inclusion of specific control variables. Column 4 shows robustness to excluding the village-level control variables, $\mathbf{x}_j$, used to select sites and determine carrying capacity. However, note that excluding these variables leads to larger IV estimates (i.e., steeper slopes on both sides of the inverted U), which is intuitive as they capture some of the confounding variation between natural advantages and diversity. Column 5 excludes the individual-level controls for age, education, and gender of the household head, resulting in no change relative to the baseline in column 1. Finally, column 6 excludes all variables, and again leaves the key result intact.

Sorting and Compositional Differences. Naturally, an important concern is that the stronger integration in more diverse communities simply reflects the sorting of more tolerant people. We begin by ruling out the role of schooling and recent immigration as potential mechanisms driving our key results. On the former, as just discussed, column 5 shows that the results are unchanged when we omit the respondent's level of schooling. Furthermore, Appendix Table A.2 shows that the inverted-U relationship is similar for individuals with different education levels. This suggests that prolonged exposure to the national language in schools (see Section 2.2) cannot explain the inclination to use that language at home in diverse communities. Next, we demonstrate a similar robustness when dropping villages that exhibit high levels (above the 90th percentile) of recent immigration between 1995 and 2000 in column 7 of Table 3. Since this post-program immigration measure and our IV are orthogonal (see Figure 4), this result provides suggestive evidence that *ex post* sorting into Transmigration areas is not driving the results.[32]

Furthermore, in column 8 of Table 3, we address the concern that as settlements split into multiple villages (e.g., due to population growth), the new, endogenously created village boundaries would sort existing villagers on their tolerance for national integration. Given the initial lottery assignment of plots within the original settlements, such endogenously-formed villages would need to have relatively more complex boundaries in order to spatially concentrate diverse, tolerant types into a single administrative unit. Following the fractal dimension approach in Alesina et al. (2011), we find that the centrally-planned Transmigration villages have relatively straight borders, suggesting relatively more exogenous boundaries, compared to non-program villages in Indonesia. However, there is some variation in border complexity, and column 8 shows that the baseline inverted-U relationship survives after dropping villages above the 90th percentile of border squiggliness.[33]

[32] In both *Susenas* and *Supas*, the inverted U is even more stark as we expand the settlement catchment area to 20 or 30 km, at which point we still find a significant Inner-Island ethnic population due to the program (see Appendix Figure B.2). Regardless of the catchment area, the turning point is close to the 50 percent threshold. These spatial spillover results are also interesting in light of a concern that APPDT natives were positively selected into the Transmigration program and, being more tolerant, are a source of positive bias in estimating diversity impacts. If all of our effects were driven by the *ex ante* selection of tolerant APPDT natives, then expanding the settlement radius to 20–30 km should have flattened or turned the inverted U upside down. Instead we find the opposite, which is reassuring from an identification standpoint.

[33] The fractal dimension of Transmigration villages is 0.014 on average compared to non-program villages, which have a statistically significantly higher average index of 0.015. For reference, Kenya, a country with straight borders imposed by colonizers

Note also that the symmetry in the inverted-U relationship goes against concerns about endogenous borders. In particular, one worries that villages with low Inner-Island ethnic shares merged with non-Transmigration villages for reasons correlated with tolerance for integration. However, we find similar results in places with low and high Inner-Island ethnic shares, suggesting that the positive effect of local diversity on national integration is not driven by the group configuration of low-diversity villages.

Table 4 provides further evidence that sorting based on tolerance does not drive our results. For this exercise, we use the 1995 *Supas* data, which allows us to observe the timing of immigration and intra-household linguistic heritage. We first show that the inverted-U relationship is largely driven by initial program settlers. Column 1 reports a baseline IV specification analogous to the *Susenas* specification in column 2 of Table 2.[34] Column 2 excludes Outer-Island natives that migrated to Transmigration areas from faraway locations outside the given district. Such long-distance migrants are more likely to have sorted endogenously than the typical APPDT settlers who happen to reside in nearby areas exposed to opportunities to relocate to Transmigration settlements. Column 3 further restricts to those individuals that arrived in the village during the initial years of settlement, thereby excluding post-program immigrants more likely to have sorted in *ex post*.

Meanwhile, columns 4 and 5 of Table 4 rule out the role of pre-existing levels of national language affinity in driving our main results. Column 4 excludes households in which the head and spouse belong to different ethnolinguistic groups. This addresses the concern that intermarried households select-in to diverse communities and are also more likely to use the national language at home. Column 5 takes an even more demanding approach of excluding all households in which the head reports Indonesian as their mother tongue. These households account for around half of Indonesian use at home in the sample, though as discussed in Section 5.5 below, linguistic identity itself, particularly among the younger generation, is affected by diversity within one's community.

Across all robustness checks in Table 4, the baseline inverted-U shape and turning point around an equal mix of Inner and Outer Islanders remains unchanged. The result in column 5 is particularly encouraging as it suggests that the effects of diversity on national language use at home are not driven by those with inherited predisposition to national integration. It is precisely this catalyzing effect of unanticipated exposure to diversity on language use at home that can have intergenerational implications for nation building as discussed in Section 2.2. Together, the evidence from both *Susenas* and *Supas* point to a systematic positive effect of local diversity on nation building as implied by the primary use of Indonesian at home.

5.4 Mechanisms: Identifying Conditions That Facilitate Integration Amidst Diversity

Overall, the positive effects of local diversity on national integration are consistent with contact theory and a large literature on cultural formation, which suggests that exposure to other groups can socialize cooperative preferences in the long run (see Pettigrew and Tropp, 2006, for a survey). As explained in Section 2, policymakers often want to understand how local conditions influence the individual response

(see Michalopoulos and Papaioannou, 2016), has an index of 0.013. The 90th percentile of villages in Table 3 still have very straight borders with a fractal dimension akin to Sudan, which Alesina et al. actually use as an example of straight borders.

[34] Unlike *Susenas*, *Supas* records language questions for all household members. Hence, we include fixed effects for relationship to the household head as well as exhaustive age and years of schooling indicators.

to intergroup contact. Below we shed light on some of the conditions under which diversity increases the relative benefit and/or reduces the relative costs of Indonesian use at home. The wide geographic scope of the program allows us to observe migrants in many different social and economic environments, allowing us to examine community-level mechanisms that may hinder or facilitate the integrative effects of diversity. We draw four lessons that can help inform the design of resettlement policies with respect to (i) the spatial configuration of communities, (ii) the local ethnic mix, (iii) interethnic economic competition, and (iv) the ethnopolitical balance.

First, Table 5 shows how the spatial configuration of individuals mediates the effects of local diversity on integration. Column 1 examines residential segregation by augmenting the baseline quadratic IV specification from Table 2 with the isolation index for Inner- and Outer-Island groups within the village. We instrument the overall isolation index with that for households headed by the older generation of initial settlers, which is more likely to be the outcome of the initial lottery assignment of plots (see Section 5.1) and hence plausibly exogenous.[35] A one standard-deviation increase in isolation of ethnic groups across census blocks leads to roughly 6 p.p. lower use of the national language at home relative to the mean of 25 percent. Moreover, overall village-level diversity remains significant with the inverted U retaining its overall shape and location. These results are consistent with Alesina and Zhuravskaya (2011) who show that ethnic segregation matters for political economy outcomes above and beyond overall ethnic diversity.

We also find that geographical isolation can amplify the relationship between diversity and integration. Columns 2 and 3 show that the inverted-U shape is stronger in more remote places, perhaps because interactions with local neighbors are relatively more frequent. We proxy remoteness using the distance between Transmigration sites and the closest points on national and provincial highways. Unlike feeder roads to the Transmigration settlements, these highway routes were largely established by the Dutch well before the program (Rothenberg, 2013). The results are consistent with the general view that social capital tends to be stronger in more isolated communities with relatively stable membership (Munshi, 2014; Munshi and Rosenzweig, 2016).

Second, we present evidence on the conditions under which diversity leads to national integration rather than local assimilation with the majority. Columns 4 and 5 split the sample by the *ELF* of the majority group and show that the inverted-U relationship is more pronounced in villages with a fragmented majority. For example, consider two Transmigration villages that have a majority of Inner Islanders. In *Sungai Sitolang*, the Inner Islanders are more fragmented with an equal mix of Javanese and Sundanese, compared to *Rokan* which is less fragmented since it is predominantly Javanese. Similarly, consider two villages with a majority of Outer Islanders, *Singapura* (high Outer-Island *ELF*) and *Wonorejo* (low Outer-Island *ELF*). Our results suggest national integration will be stronger in villages where the *ELF* of the majority group is greater (Sungai Sitolang and Singapura), consistent with stronger incentives to coordinate on a common national language when the local majority is fragmented (see Lazear, 1999).

In a similar vein, columns 6 and 7 show that diversity has stronger effects on Indonesian use at home when the Inner–Outer linguistic distance is higher (as reflected in the linguistic similarity index used in Section 5.1).[36] Column 6 demonstrates a strong inverted-U relationship in villages where transmigrants

[35] Treating the isolation index as predetermined and not instrumenting delivers very similar results.

[36] The linguistic distance between Javanese and prominent Outer-Island languages in Transmigration areas (e.g., Minangkabau, Batak, Toraja) can be significant, akin to the linguistic distance between German and French (Lewis et al., 2009).

have languages that are less similar to the native Outer-Island group's language, thereby increasing the cost of assimilation. Meanwhile, the flat shape in column 7 is in line with ethnographic evidence on the relative preference for regional languages rather than Indonesian when both are known among the given ethnic groups (Simpson, 2007a). Given that only one-quarter of the population is speaking Indonesian at home, we expect the stronger results to first appear in places where the Inner–Outer linguistic distance is greatest. Moreover, from the nation builder's perspective, it is useful to understand the economic and social incentives that enhance Indonesian use at home. Together, the results in columns 4–7 highlight how fragmentation and linguistic similarity can lower the cost of assimilation and encourage integration through the national language.

Third, beyond the spatial and social environments above, we illustrate how the incentives created by the local economic environment may shape long-term integration outcomes. We argue in columns 8 and 9 that similarity in agricultural skills between Inner and Outer Islanders can lead to intergroup interactions that are more competitive rather than cooperative in nature. We capture these incentives using a measure of the similarity in agroclimatic characteristics between the migrants' origins in Java/Bali and a given Transmigration settlement. This *agroclimatic similarity index* proxies for transmigrants' familiarity with local growing conditions and is an important determinant of agricultural productivity (Bazzi et al., 2016).[37] Column 8 shows that the inverted-U relationship is stronger in villages whose transmigrants have below median agroclimatic similarity. Interestingly, in villages with above median agroclimatic similarity, transmigrants have similar agricultural skills and potentially posed a greater competition to natives in the broader agricultural and economic marketplace. In fact, we see a U-shaped relationship in column 9, indicative of weaker integration in the most diverse communities. This reversal suggests that in settings with interethnic competition in the local economy, diversity may work against national integration.

Fourth, Table 6 continues a similar line of investigation by examining conditions that may enhance the potential to view Inner-Island transmigrants as a political threat. Columns 1–4 of Table 6 show that the inverted-U relationship between Inner-Island ethnic share and national language use at home is stronger in Transmigration areas where the local Outer-Island ethnic group constitutes a more sizable political force. Columns 1 and 2 split the sample by the size of that group at the subdistrict level, and columns 3 and 4 do so at the district level. In both cases, the inverted U is more pronounced in villages where the local native group has above-median representation at the regional level, and is less likely to view transmigrants as a political threat. Meanwhile, columns 5 and 6 show that these findings do not extend to the province level where the sample split reveals similar inverted-U shapes for both low and high ethnic representation.[38] Although speculative, this can be seen as a validation check insomuch as most political action and public resource contestation in Indonesia takes places at levels of administration below the province.

[37] Importantly, this proxy, detailed further in Appendix D, is orthogonal to the instrument (see Appendix A.II).

[38] Note that even at this level, the median local ethnic group's regional representation remains relatively high at around 20 percent compared to 30 and 40 percent at the district and subdistrict levels.

5.5 Marriage and Intergenerational Transmission of Identity

The mechanisms we identify above suggest conditions where intergroup contact fosters cultural change by shaping identity across generations. Below, we provide direct evidence on intergroup relationships and identity choices among children. Our findings echo the patterns reported in Table 1, which showed that children who grow up in multiethnic parents or speaking Indonesian at home are more likely to be in intermarriages and are more likely to speak Indonesian at home when they form their own households.

Figure 8 presents an inverted-U relationship between the Inner-Island ethnic share and the likelihood of intermarriage between young Inner and Outer Islanders. Since migrants could easily marry within their own ethnic group in villages with an equal mixes, the shape again highlights relatively stronger integration forces in diverse communities. We view these intermarriage patterns as additional evidence of new intergroup relationships. Despite Indonesia's diversity, intermarriage is relatively uncommon with around 10 percent of individuals marrying outside their ethnic group and an even lower rate in rural areas. Such marriages may be a useful vehicle for nation building: Indonesian children in intermarried households exhibit greater tolerance and weaker ethnic attachment later in life (see Table 1).

Next, we examine how diversity shapes national identity among youth by exploring the choice to report Indonesian as a mother tongue. Using the 1995 *Supas* data, we restrict our analysis to the sample of all individuals reported as children of the household head and spouse. As discussed in Section 4.3, we view the decision to report Indonesian as one's mother tongue as a leading indicator of national identity. Individual reporting of mother tongue exhibits considerable within-household variation and is not necessarily the primary language used at home. Among the 12,579 children in the regression sample, 17 percent report Indonesian as their primary language at home, and 16 percent as their mother tongue.[39]

We find a strong inverted-U relationship between the Inner-Island ethnic share and children's attachment to the Indonesian language. This can be seen in Figure 8 and Table 2. Column 1 of Table 2 first confirms that the inverted-U relationship between Inner-Island ethnic share and Indonesian use at home holds for our sample of children as it did for the full population in Appendix Table 4. Column 2 then identifies a similar inverted-U relationship for Indonesian mother tongue.

Interestingly, we find that the role of community-level diversity in shaping national identity among children goes above and beyond parental inputs to the socialization process. The patterns are robust to adding an indicator for whether parents are in an interethnic marriage (column 3) or an indicator for whether parents claim Indonesian as their own mother tongue (column 4), or both (column 5). Moreover, we show in Appendix Table A.3 that diversity has similarly large effects on children's affinity with the Indonesian identity when restricting to a sample of children whose mother reports a primary language at home besides Indonesian. These results lend insight on the extent to which parental choices and local diversity may be substitutable in the identity formation process (Bisin and Verdier, 2000). Across columns, the results indicate that diversity outside the household still influences identity formation after accounting for the familial transmission channel.

Overall, the nexus of results thus far are consistent with nation builders' intentions to use Indonesian

[39]However, the latter is not a strict subset of the former, with around 11 percent of those reporting Indonesian mother tongue not reporting it as the primary language at home, and 15 percent vice versa. Among those reporting Indonesian as their mother tongue, 23 percent have a co-resident mother who does not report Indonesian as her primary language at home.

to promote a shared identity. While parental preferences and behavior play a crucial role in children's attachment to the national identity, our findings here suggest that social conditions also matter. This provides further evidence on the scope for policy (that affects local diversity) to shape identity. To the extent that local diversity helps stimulate national language use at home, ethnic mixing can hasten the development of intergroup relationships and strengthen national identity across generations. By increasing the likelihood that parents socialize their children to use the national language, local diversity has the potential to foster nation building across generations.

6 Did the Transmigration Program Foster Nation Building?

Despite the view among policymakers that the Transmigration program was an effective tool for nation building, popular accounts of local sentiments towards the program were not always positive. Even today, many of these accounts remain colored by particularly egregious cases of failed integration.[40] Here, we shed new light on the long-term nation building impacts of the Transmigration program. Investigating outcomes across *all* Transmigration settlements, we ultimately offer support for a more sanguine view of the program now emerging in the literature. In particular, we show that the program facilitated nation building as reflected in greater use of the national language at home and intermarriage in the Outer Islands.

6.1 Identifying Program Impacts

We identify the impacts of the program using counterfactual villages that resulted from the sudden and large budget cuts in the mid-1980s. The dramatic fall of oil prices in January 1986 led to drastic reductions in the Transmigration budget. The number of resettled transmigrants fell sharply and as a result, many areas that originally planned to receive transmigrants never did (see Appendix C). Land clearing in these planned sites (or RDAs) was deferred indefinitely, and these unfinished sites were eventually settled by spontaneous migrants.

Villages located in these planned but unsettled RDAs provide a counterfactual for what Transmigration villages would have looked like today in the absence of the program. We identify 907 of these counterfactual, almost-treated (control) villages by digitally tracing high-resolution planning maps of RDAs. We assume that if Transmigration settlements had not been part of the program, settlement of new villages would have taken place organically by the endogenous sorting of migrants. Without the almost-treated RDA villages, we would be comparing Transmigration villages with earlier settlements that tended to have better market access and natural advantages. Additionally, focusing on individuals in newly created villages allows us to shed light on how intergroup relationships develop in a community's formative years, when multiple equilibria are possible and social norms are still in flux.

[40] For example, Pisani (2014) details a visit to a particularly unsuccessful Transmigration settlement in the conflict-ridden province of Aceh in the 1990s. She notes that "But even where transmigrants rubbed along well enough with their neighbors, they carried on speaking their mother tongue, they cultivated crops they grew back home, they set up the gamelan gong orchestras that mirrored those of Java or Bali. It was more transplantation than transmigration, hardly a homogenizing force. . . . *Transmigration was a rare failure in Suharto's nation building efforts.*" (pp. 36-7, emphasis added). Similar anecdotes and select case studies abound in the literature on Transmigration program and are part of broader concerns about "sons of the soil" conflict in Indonesia (see Fearon and Laitin, 2011).

Adopting a place-based evaluation strategy, we estimate program impacts based on the following equation for the average treatment on the treated (ATT):

$$y_j = \alpha + \theta \text{ Transmigration}_j + \mathbf{x}_j'\boldsymbol{\beta} + \nu_j, \tag{2}$$

where y_j is an outcome for village j, Transmigration_j is a treatment indicator equal to one for Transmigration villages and zero for RDA villages, $\mathbf{x}_j$ includes the predetermined control variables capturing natural advantages and used by planners in site selection, and ν_j is an error term. The coefficient θ captures the ATT. We estimate θ using a Oaxaca-Blinder reweighting procedure adopted in recent evaluations of place-based policies (Kline and Moretti, 2014). This approach is akin to propensity-score reweighting but can be estimated in a single step (Kline, 2011), with standard errors clustered at the district level. To address the concern that the program targeted under-developed areas, Appendix B.I provides more balance checks and shows how restricting the counterfactual to villages in the almost-treated RDAs and reweighting the sample are important for causal identification.

6.2 Effects on Nation-Building Outcomes

Table 8 presents the ATT results. Panel A shows that the Transmigration program increases adoption of Indonesian at home. Individuals in Transmigration areas are 25 p.p. more likely to report Indonesian as their primary language at home relative to the control area mean of 12 percent. We estimate this individual-level ATT for all people residing within 10 kilometers of treated or control village boundaries based on the 2006 *Susenas* sample. Each cell is a separate ATT estimate where the dependent variable is an indicator for whether the household head reports that the primary language used at home is Indonesian (column 1), native to Inner-Islands ethnicities (column 2), or native to Outer-Island ethnicities (column 3). Columns 2 and 3 show that the increased national language use is driven largely by lower use of native Outer-Island languages.[41]

In Appendix B.III, we show that these results are not explained by important individual- and village-level confounders. The ATT is robust to controlling for demographics and education, occupation, household expenditures per capita, and linguistic similarity between the local indigenous language and Malay (the root of Indonesian). Moreover, the effect size is plausible, roughly akin to the difference in national language use between individuals (i) in urban versus rural areas, or (ii) with middle school versus no education. Reassuringly, despite lower mean use of Indonesian at home in 1995, the *Supas* data reveal similar ATT effect sizes, even after accounting for some of the additional sample restrictions available with this data as noted in Section 5.3.

[41] Indeed, restricting the ATT estimates to Outer Islanders (not shown), we find that those in Transmigration areas are 27 p.p. more likely to report Indonesian use at home relative to those in control areas (10 percent of whom report Indonesian). This suggests that native Outer Islanders are more likely to adopt the national language at home in Transmigration areas where they are more likely to come into contact with Inner-Island ethnic groups. Meanwhile, we do not find similar positive effects on Indonesian use when restricting to Inner Islanders. One potential explanation for this null effect could be that Inner Islanders in control areas tend to be significant minorities (see Appendix Table B.2) for whom adopting the majority, Outer Island group language (rather than Indonesian) offers another means of integration. The evidence presented in Section 5.2 is consistent with this interpretation as is the fact that Inner Islanders in control areas are significantly more likely on average to speak an Outer-Island language than those in treated areas. Of course, another explanation could be that Inner Islanders in both Transmigration and control areas are positively selected on tolerance, given that both were willing to migrate over long distances outside their ethnic homelands.

In addition, the program leads to a higher intermarriage rate between Inner and Outer Islanders, with no effect on the overall marriage rate. Panel B of Table 8 presents estimates of equation (2) for the intermarriage rate among young cohorts.[42] While we find no effects on the marriage rate (row 1), we identify a tripling of intermarriages between Inner- and Outer-Island ethnic groups in treated villages relative to a mean of 2.3 percent in control villages (row 2). The 5 p.p. effect size is large, comparable in magnitude to differences in intermarriage rates between primary- versus junior secondary-educated individuals in Indonesia's ethnically-diverse and cosmopolitan capital, Jakarta. The ATT estimates point to significant changes in matching behavior within the marriage market.[43] Furthermore, applying an adjustment factor for random matching, we still find a significant increase in intermarriage rates above and beyond the implied differences in the diversity of the marriage market across treated and control areas (row 3). In Appendix B.IV, we provide further background on this benchmarking exercise and also rule out additional concerns about selection. In our most demanding analysis, we utilize the full, individual-level Census data on over 1.2 million married individuals in Transmigration and control villages to identify treatment effects using finer comparisons and more granular controls.

Together, these results suggest that integration is stronger in Transmigration settlements than in observably identical control areas. This is intuitive to the extent that endogenous sorting in control areas produces ethnic segregation and less intergroup exposure, thereby limiting the scope for diversity to foster cultural integration via national language use. While perhaps unsurprising through this lens, these patterns are nevertheless inconsistent with popular views of the program, which are overwhelmingly negative. That said, our results are consistent with a recent reappraisal of the program by Barter and Côté (2015) who argue that the state-sponsored Transmigration communities were not associated with the salient conflicts between Inner and Outer Islanders that erupted in the late 1990s. Their revisionist account complements our results and provides a counterpoint to claims that the program was a quintessential example of how state-sponsored migration can stoke "sons of the soil" conflict.

The results in Section 5 suggest that the increase in local diversity is a key mechanism underlying these positive average program impacts on nation building. Taking this interpretation seriously, if the program only generated settlements in the tails of the distribution of Inner-Island ethnic shares, we would likely have inferred that there was a much weaker relationship between diversity and national language use at home. The mixed settlements in the middle of this distribution are much less likely to arise in the absence of the program. Yet, these are precisely the resettlement areas that drive much of the estimated ATT effects on national language use at home and intermarriage.

7 Conclusion

We shed new light on how diversity contributes to an intergenerational process of nation building. We leverage a large-scale resettlement program involving two million voluntary migrants and 900 communities along a continuum of policy-induced local diversity. Importantly, land market imperfections

[42] These age restrictions are as defined in Figure 8 for Transmigration villages. For control villages, we define the restrictions similarly but based on the median year of settlement for Transmigration villages within the same district or province.

[43] The ATT for Inner–Outer intermarriages is similar to the ATT of 0.072 (0.011)*** for intermarriage between any ethnic groups, suggesting that most of the changes can be accounted for by the increased mixing between Inner- and Outer-Island ethnic groups.

limited *ex-post* sorting and tipping that would have otherwise neutralized the initial assignment, thereby allowing us to study how intergroup contact can shape identity and integration over time.

Our findings illustrate how the relationship between diversity and integration is uncertain and malleable, presenting scope for policy makers to manage intergroup relations amidst rising diversity. We find an inverted-U relationship between the Inner-Island ethnic share and Indonesian use at home, intermarriages, as well as the national identity choices of children. This shape is consistent with stronger integration outcomes in places with an equal mix. Interestingly, we also find a U shape in places where migrants' skills are potentially substitutes of the natives. We identify conditions that influence integration in diverse communities, including spatial segregation, linguistic distance, and perceived political competition. These findings contribute important lessons for the design of resettlement policies that should be investigated further in future work.

Beyond Indonesia, the greater integration that we identify is important given that recent work documents potential economic benefits of diversity (e.g., Alesina et al., 2016; Ashraf and Galor, 2013) that may go unrealized if sociocultural concerns preclude efforts to foster greater diversity (e.g., Borjas, 2016). Indeed, our results suggest that the link between ethnic diversity and conflict may be amenable to policy. Our findings underscore the importance of a shared national identity to unite diverse groups and a common national language that can coordinate and connect multiple groups. Future work could examine the welfare consequences of instilling a national identity and weakening group attachments.

From a policy perspective, the changes in marriage and language use that we observe have important implications for nation building given the social spillovers across generations. It is precisely the sort of intergenerational multipliers that we saw in the IFLS results in Section 4.3 that can help sustain high levels of local diversity over time. Although small, the mixed Transmigration communities may in turn matter for aggregate policy outcomes insomuch as local cultural change spills over onto the broader political environment as argued, for example, in the diffusion of democracy from local to national levels (Giuliano and Nunn, 2013). Because Transmigration settlements arose at a critical juncture of institutional development in these frontier areas of the country, it is possible that their impacts on subsequent social and cultural development in formerly frontier areas were quite sizable in the long-run.[44] This potential for persistence is consistent with a growing body of economic research on culture and institutions (see Alesina and Giuliano, 2015) and should be further explored in future work.

[44]The cultural geographer Zelinsky (1973) formalizes this possibility in his "doctrine of first effective settlement," which asserts that "Whenever an empty territory undergoes settlement [...] the specific characteristics of the first group able to effect a viable, self-perpetuating society are of crucial significance for the later social and cultural geography of the area, no matter how tiny the initial band of settlers may have been."

References

Abramitzky, R., L. P. Boustan, and K. Eriksson, "Cultural Assimilation during the Age of Mass Migration," *Working Paper*, 2015.

_ , _ , and _ , "Cultural assimilation during the age of mass migration," Technical Report, National Bureau of Economic Research 2016.

Advani, A. and B. Reich, "Melting pot or salad bowl: The formation of heterogeneous communities," Technical Report, Institute for Fiscal Studies 2015.

Alesina, A. and B. Reich, "Nation Building," *Unpublished Manuscript*, 2015.

_ and E. La Ferrara, "Participation in heterogeneous communities," *The quarterly journal of economics*, 2000, *115* (3), 847–904.

_ and E. LaFerrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, September 2005, *43* (3), 762–800.

_ and E. Zhuravskaya, "Segregation and the Quality of Government in a Cross-Section of Countries," *American Economic Review*, 2011, *101*, 1872–1911.

_ and P. Giuliano, "Culture and Institutions," *Journal of Economic Literature*, 2015, *53* (4), 898–944.

_ , G. Tabellini, and F. Trebbi, "Is Europe an Optimal Political Area?," *Brookings Papers on Economic Activity*, 2017.

_ , J. Harnoss, and H. Rapoport, "Birthplace Diversity and Economic Prosperity," *Journal of Economic Growth*, 2016, *21* (2), 101–138.

_ , W. Easterly, and J. Matuszeski, "Artificial states," *Journal of the European Economic Association*, 2011, *9* (2), 246–277.

Algan, Y., C. Hémet, and D. D. Laitin, "The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation," *Journal of Political Economy*, 2016, *124* (3), 696–733.

Alisjahbana, S. T., *Indonesian language and literature: Two essays*, Yale University, Southeast Asia Studies, 1962.

Allport, G. W., "The nature of prejudice," 1954.

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

Ananta, A., E. N. Arifin, M. S. Hasbullah, N. B. Handayani, and A. Pramono, "Changing ethnic composition: Indonesia, 2000-2010," in "XXVII IUSSP International Population Conference" 2013, pp. 26–31.

Anderson, B., *Imagined Communities*, Verso, 1983.

Anderson, T. W. and H. Rubin, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, 1949, *20* (1), 46–63.

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

Babcock, T., "Transmigration as a regional development strategy," in C. MacAndrews, ed., *Central Government and Local Development in Indonesia*, Oxford: Oxford University Press, 1986.

Bandiera, O., M. Mohnen, I. Rasul, and M. Viarengo, "Nation-Building Through Compulsory Schooling During the Age of Mass Migration," *Economic Journal*, forthcoming.

Bansak, K., J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence, and J. Weinstein, "Improving refugee integration through data-driven algorithmic assignment," *Science*, 2018, *359* (6373), 325–329.

Barbour, S. and C. Carmichael, *Language and nationalism in Europe*, OUP Oxford, 2000.

Barter, S. J. and I. Côté, "Strife of the soil? Unsettling transmigrant conflicts in Indonesia," *Journal of Southeast Asian Studies*, 2015, *46* (1), 60–85.

Bazzi, S., A. Gaduh, A. Rothenberg, and M. Wong, "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, 2016, *106* (9), 2658–2698.

_ , **M. Fiszbein, and M. Gebresilasse**, "Frontier Culture: Historical Roots and Persistence of 'Rugged Individualism' in the United States," *Unpublished Manuscript*, 2017.

Becker, G. S. and K. M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy*, 1988, *96* (4), 675–700.

Beeby, C. E., *Assessment of Indonesian education: a guide in planning* number no. 59. In 'Education research series.', Wellington: New Zealand Council for Educational Research, 1979.

Bell, W., "A probability model for the measurement of ecological segregation," *Social Forces*, 1954, *32* (1), 357–364.

Bertrand, Jacques, *Nationalism and ethnic conflict in Indonesia*, Cambridge, UK; New York: Cambridge University Press, 2004.

Bisin, A. and T. Verdier, "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *The Quarterly Journal of Economics*, 2000, *115* (3), 955–988.

_ **and** _ , "The Economics of Cultural Transmission and Socialization," in Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics, Volume 1, Part A*, Elsevier, 2011, pp. 339–416.

Blouin, A. and S. Mukand, "Erasing Ethnicity? Nation Building, (Mis)Trust and the Salience of Identity in Rwanda," *Unpublished Manuscript*, 2016.

Blumer, Herbert, "Race prejudice as a sense of group position," *Pacific Sociological Review*, 1958, *1* (1), 3–7.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles, "Empathy or antipathy? The impact of diversity," *American Economic Review*, 2006, *96* (5), 1890–1905.

Borjas, G., *We Wanted Workers: Unraveling the Immigration Narrative*, W.W. Norton & Company, 2016.

Bronnenberg, B. J., J.-P. H. Dube, and M. Gentzkow, "The Evolution of Brand Preferences: Evidence from Consumer Migration," *American Economic Review*, May 2012, *102* (6), 2472–2508.

Central Bureau of Statistics, "SUSENAS 2006 JUL - Guidelines of Socio-culture and Education Module Census (id)," 2006.

Charras, M. and M. Pain, *Spontaneous settlements in Indonesia: agricultural pioneers in Southern Sumatra*, Jakarta and Bondy, France: Departemen Transmigrasi ; Institut francais de recherche scientifique pour le devéloppement en coopération : Centre national de la recherche scientifique, 1993.

Clauss, W., H.-D. Evers, and S. Gerke, "The Formation of A Peasant Society: Javanese Transmigrants in East Kalimantan," *Indonesia*, 1988, (46), 79–90.

Clingingsmith, D., A. I. Khwaja, and M. Kremer, "Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering," *Quarterly Journal of Economics*, 2009, *124* (3), 1133–1170.

Clots-Figueras, I. and P. Masella, "Education, Language and Identity," *The Economic Journal*, 2013, *123* (570), F332–F357.

Colby, S. L. and J. Ortman, "Projections of the Size and Composition of the U.S. Population: 2014 to 2060, Current Population Reports, P25-1143," Technical Report, U.S. Census Bureau 2014.

Coleman, David, "Projections of the Ethnic Minority Populations of the United Kingdom 20062056," *Population and Development Review*, 2010, *36* (3), 441–486.

Dahlberg, M., K. Edmark, and H. Lundqvist, "Ethnic Diversity and Preferences for Redistribution," *Journal of Political Economy*, 2012, *120* (1), 41–76.

Dell, M. and P. Querubin, "Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies," *Quarterly Journal of Economics*, forthcoming.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg, "Culture, ethnicity, and diversity," *American Economic Review*, 2017, *107* (9), 2479–2513.

_ , **J. Gomes, and I. Ortuño-Ortín**, "The geography of linguistic diversity and the provision of public goods," *Unpublished Manuscript*, 2016.

Duflo, E., "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 2001, *91* (4), 795–813.

Duggan, C., *The Force of Destiny: A History of Italy Since 1976*, Penguin Books, London, 2007.

Edin, P.A., P. Fredriksson, and O. Åslund, "Ethnic Enclaves and the Economic Success of Immigrants–Evidence from a Natural Experiment," *Quarterly Journal of Economics*, 2003, *118* (1), 329–357.

Esteban, J. and D. Ray, "Linking conflict to inequality and polarization," *American Economic Review*, 2011a, *101* (4), 1345–1374.

Esteban, J.M. and D. Ray, "On the measurement of polarization," *Econometrica*, 1994, pp. 819–851.

Fan, J. and I. Gijbels, *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, Vol. 66, CRC Press, 1996.

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

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

_ **and D. D. Laitin**, "Sons of the soil, migrants, and civil war," *World Development*, 2011, *39*, 199–211.

Feith, H., *The decline of constitutional democracy in Indonesia*, 1st equinox ed ed., Jakarta: Equinox Pub, 1962/2007.

Fernandez, R., "The Economics of Cultural Transmission and Socialization," in Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics, Volume 1, Part A*, 2011, pp. 339–416.

Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst, and D. Wiberg, "Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)," 2008.

Fouka, V., "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I," *Stanford Center for International Development Working Paper*, 2016, (591).

Frey, W. H., *Diversity Explosion: How New Racial Demographics are Remaking America*, Brookings Institution Press, 2014.

Giuliano, P. and N. Nunn, "The transmission of democracy: from the village to the nation-state," *American Economic Review*, 2013, *103* (3), 86–92.

_ **and** _ , "Understanding cultural persistence and change," *Harvard Universtity, mimeo*, 2017.

Gordon, M. M., *Assimilation in American life: The role of race, religion, and national origins*, New York: Oxford University Press, 1964.

Hansen, L. P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50* (4), 1029–1054.

Hanson, G. and C. McIntosh, "Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run," *The Journal of Economic Perspectives*, 2016, *30* (4), 57–81.

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

Härdle, W. and E. Mammen, "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 1993, *21* (4), 1926–1947.

Hoey, Brian A., "Nationalism in Indonesia: Building Imagined and Intentional Communities Through Transmigration," *Ethnology*, 2003, *42* (2), 109–126.

Hoshour, C. A., "Resettlement and the Politicization of Ethnicity in Indonesia," *Bijdragen tot de Taal-Land-en Volkenkunde*, 1997, (4de Afl), 557–576.

Huntington, S. P., *Who are we?: The challenges to America's national identity*, Simon and Schuster, 2004.

Janvry, A. De, K. Emerick, M. Gonzalez-Navarro, and E. Sadoulet, "Delinking Land Rights from Land Use: Certification and Migration in Mexico," *Unpublished Manuscript*, 2012.

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

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

_ **and E. Moretti**, "Local Economic Development, Agglomeration Economies and the Big Push: 100

Years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 2014, *129*, 275–331.

Kramsch, C., *Language and Culture*, Oxford University Press, 1998.

Laitin, D. and R. Ramachandran, "Linguistic Diversity, Official Language Choice and Nation Building: Theory and Evidence," *Unpublished Manuscript*, 2015.

Lazear, E., "Culture and Language," *Journal of Political Economy*, 1999, *107*, s95–s126.

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

Lewis, M. P., G. F. Simons, and C. D. Fennig, "Ethnologue: Languages of the world, 16th edition," *SIL International*, 2009.

Lind, J. T. and H. Mehlum, "With or without U? the Appropriate Test for a U-Shaped Relationship," *Oxford Bulletin of Economics and Statistics*, 2010, *72* (1), 109–118.

Lowe, M., "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *Unpublished Manuscript*, 2018.

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

Mangunrai, H., "Evaluation of the development of transmigration in South Sulawesi: a study on the integration between the transmigrants and natives in the transmigration area of Luwu South Sulawesi," *SEAPRAP Research Reports*, 1977, (27).

Martinez-Bravo, M., "The Local Political Economy Effects of School Construction in Indonesia," *American Economic Journal: Applied Economics*, 2017, *9* (2), 256–289.

Matsuura, K. and C. J. Wilmott, "Terrestrial Precipitation: 1900-2010 Gridded Monthly Time Series (V 3.02)," 2012.

Matsuura, Kenji and Cort J. Wilmott, "Terrestrial Air Temperature: 1900-2010 Gridded Monthly Time Series (V 3.01)," 2012.

McPherson, M., L. Smith-Lovin, and J. M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, August 2001, *27* (1), 415–444.

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

_ **and E. Papaioannou**, "The long-run effects of the scramble for Africa," *American Economic Review*, 2016, *106* (7), 1802–1848.

Miguel, E., "Tribe or Nation?: Nation Building and Public Goods in Kenya versus Tanzania," *World Politics*, 2004, *56* (3), 327–362.

Miles, W. F. and D. A. Rochefort, "Nationalism versus ethnic identity in sub-Saharan Africa," *American Political Science Review*, 1991, *85* (2), 393–403.

Miller, R., "The Development of European Identity/Identities: Unfinished Business," *A Policy Review*, 2012.

Mulder, N., *Inside Indonesian society: cultural change in Java*, Amsterdam: Pepin Press, 1996.

Munshi, K., "Community networks and the process of development," *The Journal of Economic Perspectives*, 2014, *28* (4), 49–76.

_ **and M. Rosenzweig**, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 2016, *106* (1), 46–98.

Okunogbe, O., "Does Exposure to Other Ethnic Groups Promote National Integration? Evidence from Nigeria," *Unpublished Manuscript*, 2015.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and validation," *Journal of Business and Economic Statistics*, forthcoming.

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

Paauw, S., "One land, one nation, one language: An analysis of Indonesia's national language policy," *University of Rochester Working Papers in the Language Sciences*, 2009, *5* (1), 2–16.

Pettigrew, T. F. and L. R. Tropp, "A meta-analytic test of intergroup contact theory.," *Journal of Personality and Social Psychology*, 2006, *90* (5), 751–783.

Pew Research Center, "What It Takes to Truly Be 'One of Us'," *Washington: Pew Research Center*, February 2017.

Pisani, E., *Indonesia, Etc.: Exploring the Improbable Nation*, WW Norton & Company, 2014.

Putnam, R. D., "E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture," *Scandinavian Political Studies*, 2007, *30* (2), 137–174.

Rao, G., "Familiarity does not breed contempt: Diversity, discrimination and generosity in Delhi schools," *Unpublished Manuscript*, 2018.

Ray, D. and J. Esteban, "Conflict and Development," *Annual Review of Economics*, 2017, *9* (1).

Ricklefs, M. C., *A history of modern Indonesia since c. 1200*, 4th ed ed., Stanford, Calif: Stanford University Press, 2008.

Rigg, J., *Southeast Asia (Routledge Revivals): A Region in Transition*, Routledge, 2013.

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

Rothenberg, A., "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways," 2013. Working Paper.

Sanderson, E. and F. Windmeijer, "A weak instrument F-test in linear IV models with multiple endogenous variables," *Journal of Econometrics*, 2016, *190* (2), 212–221.

Sappington, J. M., K. Longshore, and D. Thompson, "Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study using Bighorn Sheep in the Mojave Desert," *Journal of Wildlife Management*, 2007, *71* (5), 1419–1426.

Schelling, T. C., "Dynamic models of segregation," *Journal of Mathematical Sociology*, 1971, *1* (2), 143–186.

Simpson, A., "Indonesia," in A. Simpson, ed., *Language and National Identity in Asia*, Oxford University Press, 2007.

_ , *Language and National Identity in Asia*, Oxford University Press, 2007.

_ , *Language and national identity in Africa*, Oxford University Press, 2008.

Sneddon, J.N., *The Indonesian Language: Its History and Role in Modern Society*, University of New South Wales Press, 2003.

Strauss, J., F. Witoelar, and B. Sikoki, "The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report," RAND Working Paper WR-1143/1-NIA/NICHD, RAND March 2016.

Surbakti, S., R. L. Praptoprijoko, and S. Darmesto, "Indonesia's 2000 Population Census: A Recent National Statistics Activity," Technical Report, United Nations Economic and Social Commission on Asia and Pacific 2000.

Suryadinata, Leo, "Government Policy and National Integration in Indonesia," *Asian Journal of Social Science*, January 1988, *16* (1), 111–131.

Tanasaldy, T., *Regime change and ethnic politics in Indonesia; Dayak politics of West Kalimantan*, Brill, 2012.

Thornton, D. L., "Javanization of Indonesian politics," 1972.

Weber, E., *Peasants into Frenchmen: Modernization of Rural France, 1870-1914*, Stanford University Press, 1976.

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.

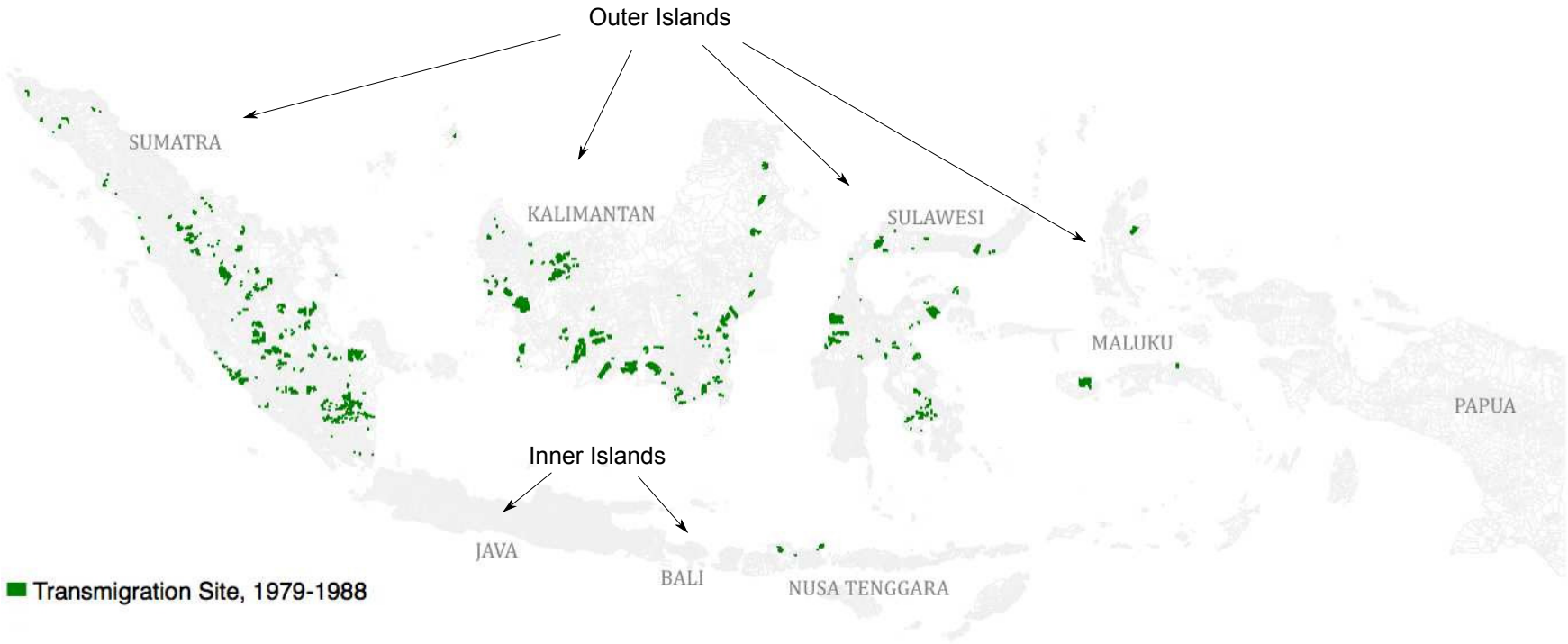
_ , *Indonesia: The Transmigration Program in Perspective* 1988.

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

Zelinsky, W., *The Cultural Geography of the United States*, Pearson Education, Limited, 1973.

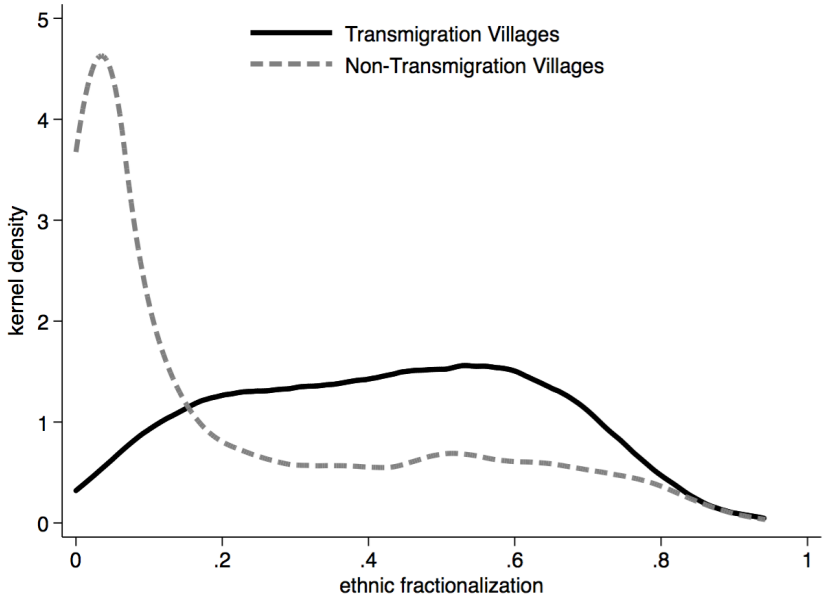
Figures

Figure 1: Map of Transmigration Villages



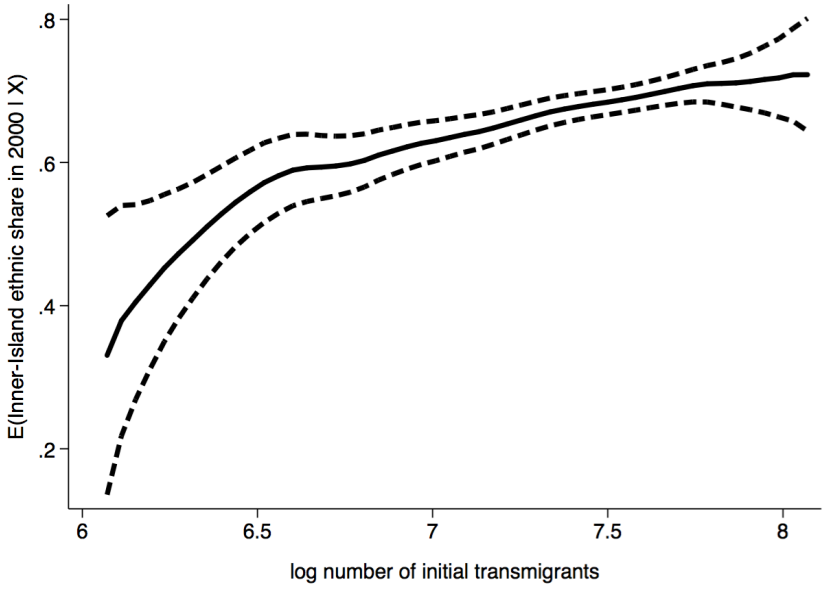
Notes: Each colored location on the map corresponds to a Transmigration village settled between 1979 and 1988. The white areas outlined in grey are other villages.

Figure 2: Transmigration Generated a Persistent Continuum of Diverse Communities



Notes: This figure plots the kernel density of ethnic fractionalization in 2000 for Transmigration villages and non-Transmigration villages greater than 30 kilometers from the boundaries of Transmigration villages. For both densities, we employ an Epanechnikov kernels and rule-of-thumb bandwidth.

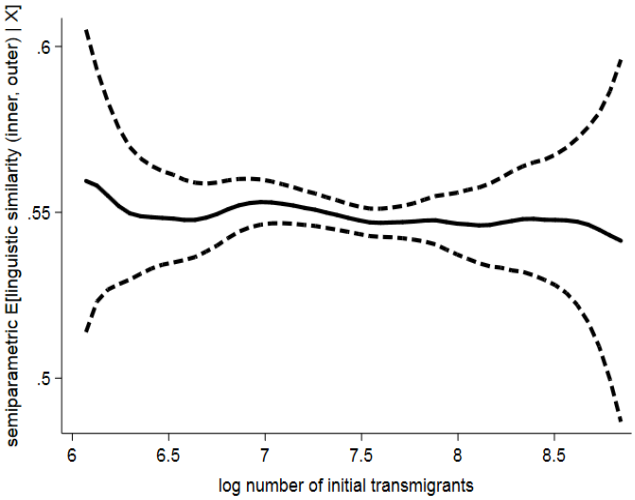
Figure 3: Initial Transmigrant Assignment and Long-Run Inner-Island Ethnic Share



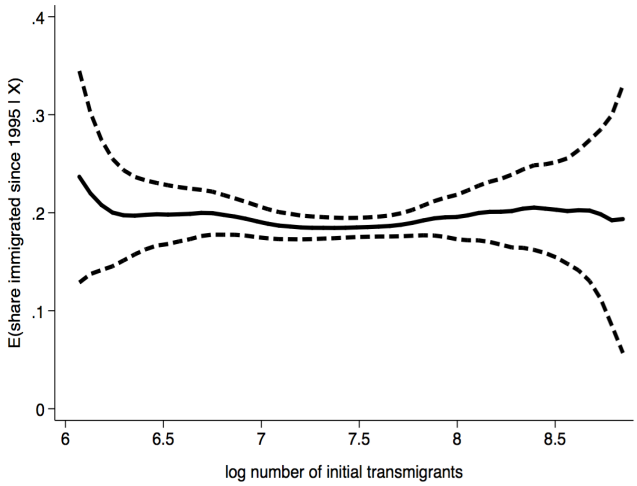
Notes: This figure reports a semiparametric Robinson (1988) regression and 95 confidence interval of the Inner-Island ethnic share in 2000 (based on the Population Census) on the log of the transmigrant population from Java/Bali placed in that village in the initial year of settlement. The local linear regression is conditional on island fixed effects and the vector $\mathbf{x}$ of predetermined site selection variables described in the paper, and it is estimated based on an Epanechnikov kernel, Fan and Gijbels (1996) rule-of-thumb bandwidth, and trimming of the top 5th and bottom 1st percentile for presentational purposes. This provides a semiparametric view of the first stage of our main IV specification.

Figure 4: Initial Transmigrant Assignment Uncorrelated with Proxies for Sorting

(a) Linguistic Similarity with Natives (*ex ante*)

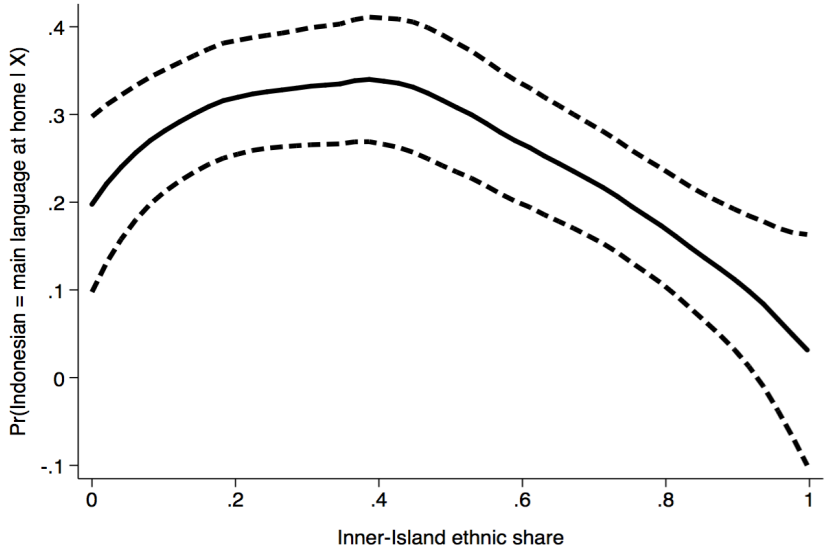


(b) Post-Program Immigration Share (*ex post*)



Notes: This figure reports a semiparametric Robinson (1988) regression and 95 confidence intervals of the Inner-Island ethnic share in 2000 (based on the Population Census) on (a) the linguistic similarity between the Inner-Island ethnic population and the indigenous Outer-Island group according to the *Ethnologue* and World Language Mapping System, and (b) the share of the population that immigrated to the village between 1995 and 2000. We use a local linear regression with island fixed effects and the vector $\mathbf{x}$ of predetermined site selection variables, an Epanechnikov kernel, Fan and Gijbels (1996) rule-of-thumb bandwidth, and trimming of the top and bottom percentiles for presentational purposes.

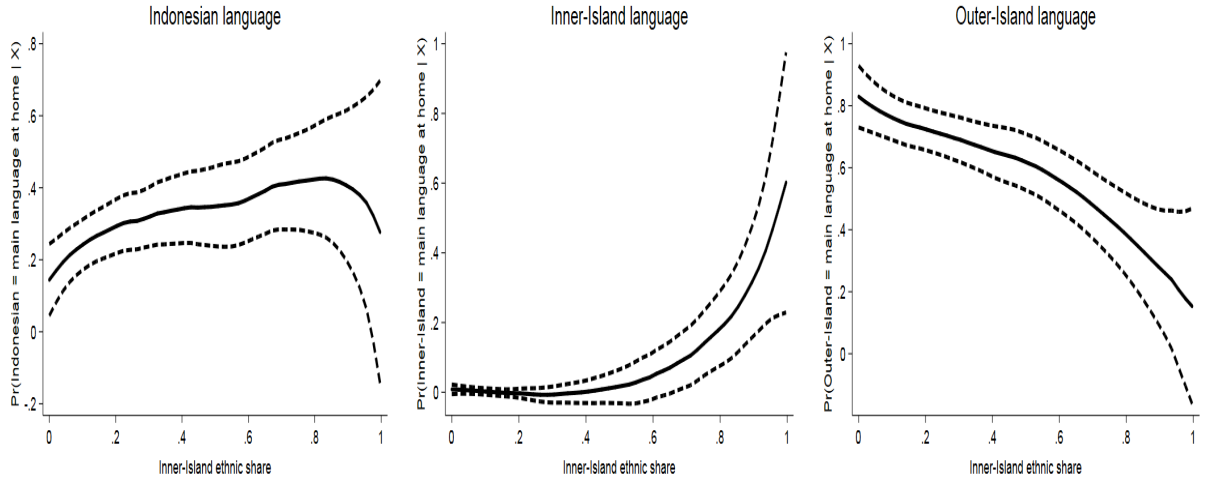
Figure 5: Ethnic Diversity and National Language Use at Home



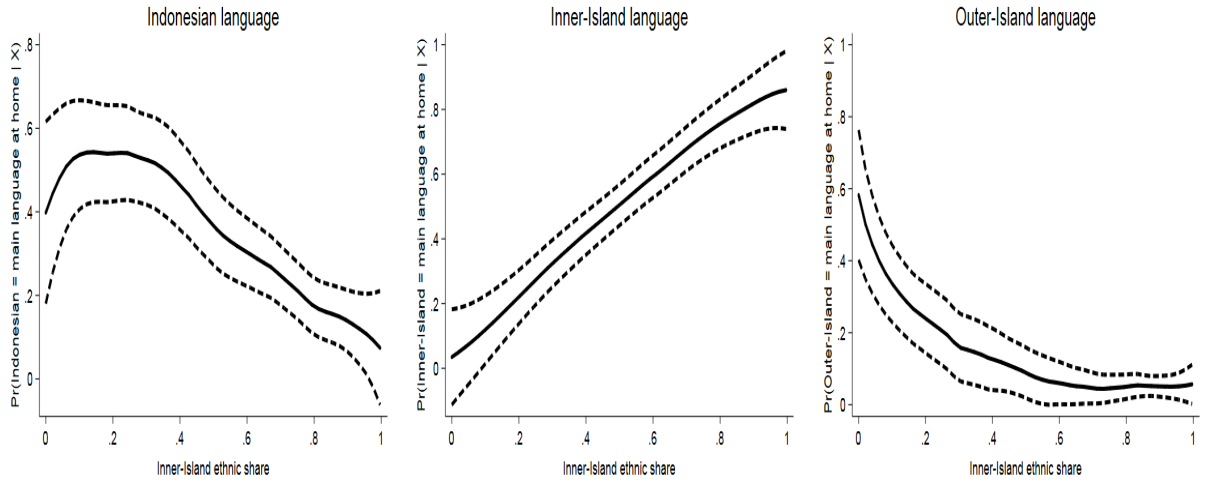
Notes: This figure reports a semiparametric Robinson (1988) OLS estimate and 95 percent confidence interval for the effect of ethnic diversity (Inner-Island ethnic share) on the likelihood of reporting the national language, *Bahasa Indonesia*, as the primary language used at home as reported by household heads in the 2006 *Susenas* data. The sample includes all individuals within 10 km of Transmigration villages including those villages. These estimates condition on island fixed effects, the $\mathbf{x}$ vector of predetermined village-level controls, and individual-level controls for age, age squared, gender, and education. The estimates are based on an Epanechnikov kernel and Fan and Gijbels (1996) rule-of-thumb bandwidth. Standard errors are clustered by village, of which there are 135. The Härdle and Mammen (1993) test rejects the null of a linear shape with a p-value of 0.002 and fails to reject the null of a quadratic parametric shape with a p-value of 0.21.

Figure 6: Ethnic versus National Language Use, by Group

(a) Outer-Island Ethnic Groups

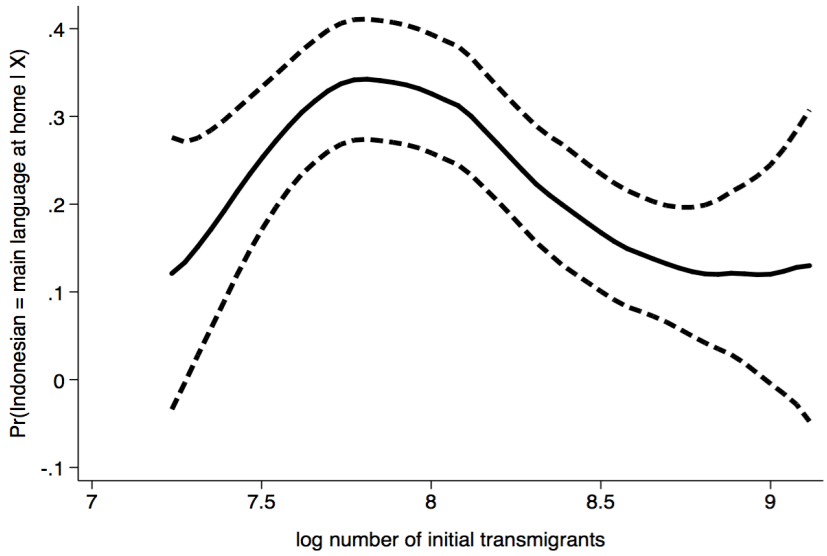


(b) Inner-Island Ethnic Groups



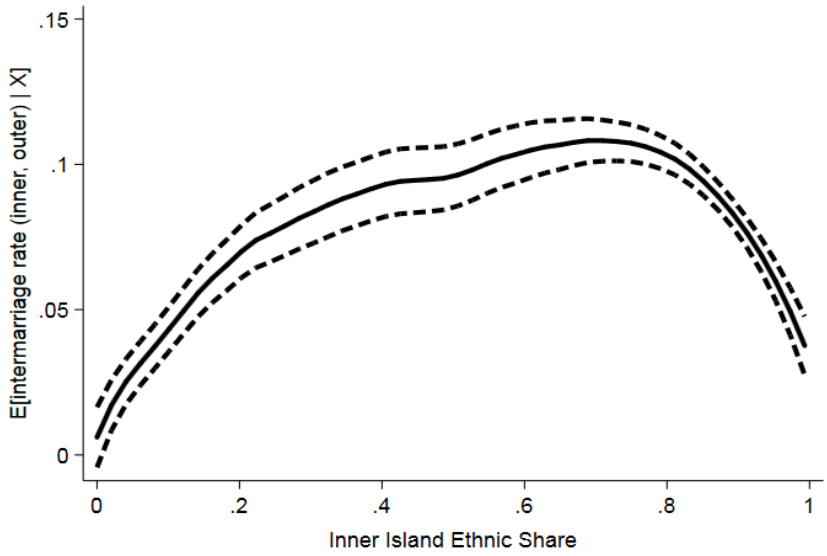
Notes: These figures report Inner- and Outer-Island group-specific semiparametric Robinson (1988) estimates of the effect of ethnic diversity (Inner-Island ethnic share) on the likelihood of reporting the given language (national, Inner-Island, Outer-Island) as the primary daily one used in the household as reported by the household in 2006 *Susenas* data for villages within 10 km of Transmigration villages.. Panel (a) is estimated over all individuals belonging to Outer-Island ethnic groups, and Panel (b) is estimated over all those belonging to Inner-Island ethnic groups. The specifications are analogous to those in Figure 5 for the full population and are based on the partially linear model that conditions on our usual $\mathbf{x}$ vector of control variables and uses an Epanechnikov kernel and Fan and Gijbels (1996) rule-of-thumb bandwidth. Standard errors are clustered at the village level. The Härdle and Mammen (1993) test fails to reject the null of a linear relationship with national language use with p-values of 0.12 and 0.11 in panels (a) and (b), respectively. For the other group's language, threshold regression tests identify inflection points at 0.487 Inner-Island ethnic share in the second figure in panel (a) and 0.360 Inner-Island ethnic share in the third figure in panel (b) with significantly different slopes above and below those thresholds

Figure 7: Reduced Form: Initial Transmigrant Assignment and National Language



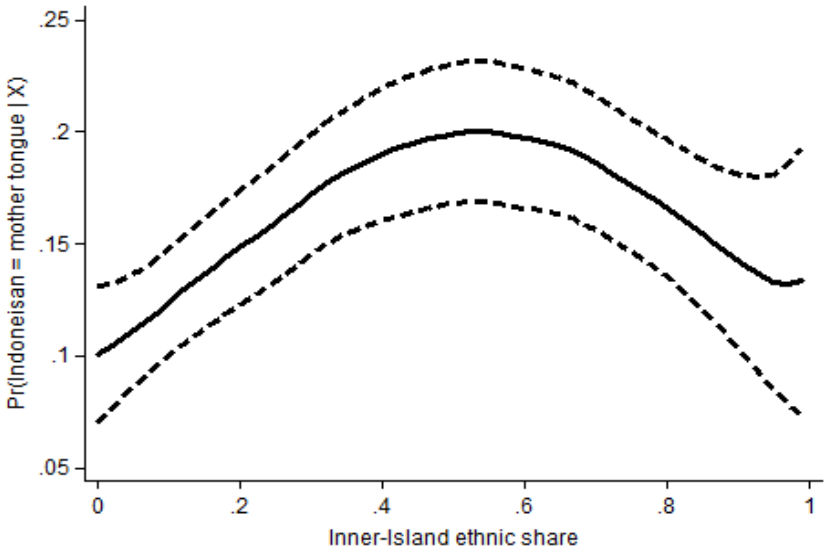
Notes: This figure reports a semiparametric (Robinson, 1988) estimate of the reduced form effects of the initial number of transmigrants (the instrument in Table 2) on the likelihood of reporting the national language as the primary one used at home. The estimate is otherwise based on the same specification as Figure 5. See the notes to that figure for further details.

Figure 8: Ethnic Diversity and Intermarriage



Notes: This figure reports a semiparametric Robinson (1988) OLS estimate and 95 percent confidence interval for the effect of ethnic diversity (Inner-Island ethnic share) on the overall share of young cohort marriages that occur between Inner and Outer Islanders. We define the younger generation as those individuals who were under 16 years of age during the period of settlement in the 1980s, but results are unchanged using other cutoffs. The regressions include 832 Transmigration villages and condition on island fixed effects and the $\mathbf{x}$ vector of predetermined village-level controls. The estimates are based on an Epanechnikov kernel and Fan and Gijbels (1996) rule-of-thumb bandwidth. Standard errors are clustered by district.

Figure 9: Ethnic Diversity and Indonesian Mother Tongue among Children



Notes: This figure reports a semiparametric Robinson (1988) OLS estimate and 95 percent confidence interval for the effect of ethnic diversity (Inner-Island ethnic share) on the Indonesian identity as reflected in reported mother tongue among children of household heads in the *Supas* data. The regressions condition on island fixed effects and the **x** vector of predetermined village-level controls plus individual-level controls for age and years of education fixed effects. Full details on the sample are provided in the notes to Table 7. The estimates are based on an Epanechnikov kernel and Fan and Gijbels (1996) rule-of-thumb bandwidth. Standard errors are clustered by district.

Tables

Table 1: National Identity Formation through Language Use at Home

	Dependent Variable as Adult in 2014:			
	Speaks Indonesian at Home	Changes Ethnicity from 1997	In Interethnic Marriage	Trust Other Ethnic Groups (z-score)
	(1)	**(2)**	**(3)**	**(4)**
	Panel A: Baseline			
Indonesian was Primary Language at Home as Child in 1997	0.156	0.062	0.053	0.148
	(0.022)***	(0.019)***	(0.023)**	(0.054)***
	Panel B: Adding Parental Intermarriage			
Indonesian was Primary Language at Home as Child in 1997	0.151	0.045	0.046	0.131
	(0.022)***	(0.019)**	(0.023)**	(0.054)**
Parents from Different Ethnic Groups	0.053	0.177	0.092	0.160
	(0.021)**	(0.030)***	(0.031)***	(0.055)***
Number of Individuals	8,623	6,594	5,628	8,236
Dependent Variable Mean	0.369	0.114	0.103	0.00
Age, Gender, Education Fixed Effects	Yes	Yes	Yes	Yes
Village Fixed Effects	Yes	Yes	Yes	Yes

Notes: This table reports estimates of the correlation between parental daily Indonesian language use at home as a child in 1997 and the given column's dependent variable for individuals in the 2014 round of the *Indonesia Family Life Survey*. Panel B augments the baseline specification with a control for whether the child's parents hail from different ethnic groups. The sample is restricted to all individuals greater than 15 years old who live in a different household in 2014 compared to 1997. The dependent variables include in column (1) an indicator for whether the individual used the Indonesian language at home on a regular basis in 2014, (2) an indicator for whether the individual switched his/her reported ethnicity between 1997 and 2014, (3) an indicator for whether a married individual is in an interethnic marriage in 2014, (4) an index normalized to have mean zero and standard deviation one based on ordered response on a 4 point scale to the question "Do you trust people from other ethnic groups less than you trust your people from own group?". Note that the language use at home variable is distinct from the *Susenas* measure used elsewhere in the paper, which requires that household heads only list a single, primary language at home as opposed to listing all languages used at home. All specifications include the fixed effects listed at the bottom of the table where the age FE are for each individual age. Standard errors are clustered at the village level of which there are around 1,300 across columns. */**/*** denotes significance at the 10/5/1 percent level.

Table 2: Ethnic Diversity and National Language Use At Home

	Dependent Variable: Indonesian is Main Language at Home			
	OLS	IV	OLS	IV
	(1)	**(2)**	**(3)**	**(4)**
Inner-Island ethnic share	0.665	0.845		
	(0.284)**	(0.379)**		
Inner-Island ethnic share squared	-0.854	-0.927		
	(0.312)***	(0.382)**		
Inner-Island ethnic share, bottom tercile			0.066	-0.042
			(0.058)	(0.067)
Inner-Island ethnic share, middle tercile			0.203	0.187
			(0.059)***	(0.110)*
Turning point	0.390	0.456		
p-value	[0.012]**	[0.015]**		
Number of Individuals	2,126	2,126	2,126	2,126
Sanderson-Windmeijer E_1 p-value		0.000		0.000
Sanderson-Windmeijer E_2 p-value		0.000		0.000
Hansen J test p-value		0.161		0.206

Notes: This table reports estimates of the effect of ethnic diversity—captured by the Inner-Island ethnic share—on national language use at home in 135 villages in the 2006 *Susenas* survey within 10 kilometers of Transmigration villages, including those villages. The mean of the dependent variable is 0.25 across all columns. The instrumental variables for the endogenous variables—E_1 for the linear or bottom tercile Inner-Island ethnic share and E_2 for the quadratic or middle tercile Inner-Island share—in columns 2 and 4 are dummies for 20 equally sized bins (i.e., ventiles) of the number of initial transmigrants settled within 10 km of the given village. Each regression controls for (i) predetermined village characteristics, $\mathbf{x}_v$, used in selecting village locations and determining carrying capacity as detailed in the text, and (ii) individual-level controls, $\mathbf{x}_i$, for the household head's age, age squared, dummies for education level and gender. The IV regressions are estimated by generalized method of moments to account for the many instruments. The null hypotheses of (i) the Sampson-Windmeijer test is that the instruments for the given endogenous variable are weak, and (ii) the Hansen J test is that the instruments are uncorrelated with the error term and correctly excluded from the second stage. The p-values for the inverted-U turning points in columns 1 and 2 are based on the exact test of Lind and Mehlum (2010). Standard errors in all columns are clustered by district, of which there are 50. */**/*** denotes significance at the 10/5/1 percent significance levels.

Table 3: Robustness Checks on Ethnic Diversity and National Language Use At Home

	Baseline x Controls			Excluding			Dropping Villages w/	
	Column 2 Table 2	+ Ethnic Homeland FE	+ Occupation FE	Village Controls	Individual Controls	All x Controls	High Recent Immigration	Squiggly Borders
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Inner-Island ethnic share	0.845	0.765	1.103	1.166	0.829	1.257	1.396	0.862
	(0.379)**	(0.373)**	(0.202)***	(0.542)**	(0.350)**	(0.554)**	(0.635)**	(0.369)**
Inner-Island ethnic share squared	-0.927	-0.815	-1.409	-1.417	-0.927	-1.590	-1.386	-0.958
	(0.382)**	(0.384)**	(0.268)***	(0.503)***	(0.353)**	(0.494)***	(0.611)**	(0.407)**
Turning point	0.456	0.391	0.469	0.412	0.447	0.395	0.504	0.450
p-value	[0.015]**	[0.000]***	[0.024]**	[0.018]**	[0.011]**	[0.014]**	[0.017]**	[0.016]**
Number of Individuals	2,126	2,078	2,126	2,126	2,126	2,126	1,823	1,902
Sanderson-Windmeijer Linear p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sanderson-Windmeijer Quadratic p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hansen J test p-value	0.161	0.279	0.167	0.130	0.132	0.111	0.344	0.134

Notes: This table reports robustness checks on the baseline IV result in column 2 of Table 2, reproduced in column 1 here for reference. The mean of the dependent variable is 0.25 across all columns. Column 2 adds dummy variables for the indigenous ethnolinguistic homeland in the given village as reported in the *Ethnologue* and World Language Mapping Survey data, which is missing for three villages. Column 3 adds 18 exhaustive dummy variables for occupation of the household head. Column 4 removes the village-level controls, $\mathbf{x}_v$, column 5 removes the individual-level controls, $\mathbf{x}_i$, and column 6 removes all controls. Column 7 drops the top 10th percentile of villages in terms of the share of the population that migrated to the village between 1995 and 2000. Column 8 drops the top 10th percentile of villages in terms of the fractal dimension of the village boundary as a proxy for the potential endogeneity of borders. Standard errors in all columns are clustered by district, of which there are 50. */**/*** denotes significance at the 10/5/1 percent significance levels.

Table 4: Further Robustness Checks on Sorting Using Auxiliary Survey Data

	Dependent Variable: Indonesian is Main Language at Home				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Sample Restrictions	None	Excl. Faraway APDDT Immigrants	Only Initial Period Immigrants	HH Head Not Intermarried	HH Heads Not Indonesian Mother Tongue
Inner-Island ethnic share	1.519	1.328	1.331	1.586	0.659
	(0.395)***	(0.394)***	(0.423)***	(0.378)***	(0.192)***
Inner-Island ethnic share squared	-1.523	-1.334	-1.366	-1.594	-0.601
	(0.431)***	(0.430)***	(0.460)***	(0.423)***	(0.222)***
Turning point	0.499	0.498	0.487	0.498	0.548
p-value	[0.001]***	[0.003]***	[0.004]***	[0.001]***	[0.021]**
Number of Individuals	28,532	24,595	17,175	25,509	25,313
Dependent Variable Mean	0.159	0.139	0.130	0.126	0.081
Sanderson-Windmeijer Linear p-value	0.073	0.067	0.076	0.068	0.067
Sanderson-Windmeijer Quadratic p-value	0.091	0.088	0.041	0.110	0.064
Anderson-Rubin Weak IV Robust p-value	0.004	0.004	0.020	0.000	0.005

Notes: This table reports estimates of the relationship between the Inner-Island ethnic share and national language use at home in the auxiliary 1995 *Supas* survey data. Because the data are prior to the 2000 Population Census, we construct a measure of the Inner-Island ethnic share based solely on the 2000 population that did not live elsewhere in 1995. As in the main *Susenas*-based regressions, we include all villages within 10 km of a Transmigration village, and the instrumental variables are ventiles of the number of initial transmigrants within 10 km of the given village. The regressions condition on the $\mathbf{x}$ vector of predetermined village-level controls plus individual-level controls for age fixed effects, years of education fixed effects, gender dummy, and relationship to household head fixed effects. Column 1 is the baseline with no sample restrictions. Column 2 drops all non-Inner-Island ethnics born outside the district of the given village, i.e., those native Outer Islanders that migrated from far distances. Column 3 drops all individuals who migrated to the village more than 3 years after the initial year of settlement. Column 4 drops all individuals residing in households where the head is married to someone with a different mother tongue. Column 5 drops all individuals who reside in households where the head and spouse report Indonesian as the mother tongue. The p-values for the inverted-U turning points are based on the exact test of Lind and Mehlum (2010). Standard errors in all columns are clustered by district. */**/*** denotes significance at the 10/5/1 percent significance levels.

Table 5: Identifying Conditions That Facilitate Integration

Mechanisms	**Spatial**			**Social**				**Economic**	
		distance to major historical roads		majority group ethnic fractionalization		linguistic similarity w/ native homeland		agroclimatic similarity	
		low	high	low	high	low	high	low	high
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Inner-Island ethnic share	1.421	0.238	1.366	0.350	1.690	0.774	0.067	1.543	-0.755
	(0.552)**	(0.794)	(0.197)***	(0.469)	(0.544)***	(0.337)**	(0.142)	(0.687)**	(0.146)***
Inner-Island ethnic share squared	-1.559	-0.605	-1.584	-0.571	-1.367	-0.796	0.032	-1.373	0.597
	(0.583)**	(0.797)	(0.186)***	(0.555)	(0.459)***	(0.381)**	(0.149)	(0.736)*	(0.171)***
isolation index (Inner, Outer)	-0.056								
	(0.032)*								
Turning point	0.456	0.197	0.431	0.307	0.618	0.486		0.562	0.633 (U)
p-value	[0.007]***	[0.383]	[0.000]***	[0.230]	[0.010]**	[0.034]**		[0.073]*	[0.037]**
Number of Individuals	2,126	1,070	1,056	1,072	1,054	1,137	814	1,023	992
Sanderson-Windmeijer Linear p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sanderson-Windmeijer Quadratic p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sanderson-Windmeijer Isolation p-value	0.000								
Hansen J test p-value	0.152	0.144	0.221	0.214	0.068	0.501	0.565	0.415	0.578

Notes: This table explores mechanisms underlying the baseline IV result on national language use at home from column 2 in Table 2. Column 1 adds a variable capturing ethnic residential segregation across Census blocks within villages. This isolation index is normalized to have mean zero and standard deviation one. This specification also dummies out the four villages with zero Inner-Island ethnic share and define their isolation index to equal zero. This overall isolation index is instrumented using the isolation index among the older generation of initial Columns 2–9 estimate separate regressions for villages with the given variable below (low) and above (high) the median in even and odd columns, respectively. Columns 2–3 split by proximity to the nearest historical (i.e., early 20th century) road. Columns 4–5 split by the agroclimatic similarity among the Java/Bali-born migrants, capturing the similarity in agroclimatic conditions between origin and destination (see Appendix D). Columns 6–7 split by the linguistic similarity between the Inner-Island ethnics' languages and the language of the local indigenous group (see notes to Figure 4 on this variable). Columns 8–9 split by the ethnic fractionalization among the Inner or Outer Islanders depending on which of the two broad groups comprises the majority in the village. See the notes to Table 2 for details on the reported instrument and turning point diagnostics. Note that the turning point is for a U rather than inverted-U shape in column 5 and falls outside the unit interval in 7 and hence is not computed. Standard errors in all columns are clustered by district. */**/*** denotes significance at the 10/5/1 percent significance levels.

Table 6: Regional Ethnopolitical Balance and Integration

Sample Restriction	Median Sample Splitting by Size of Largest Outer-Island Group. . .					
	Within Subdistrict		Within District		Within Province	
	low	high	low	high	low	high
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Inner-Island ethnic share	-0.446	1.172	-0.593	1.306	0.869	0.850
	(0.376)	(0.327)***	(0.273)**	(0.228)***	(0.440)*	(0.198)***
Inner-Island ethnic share squared	-0.103	-1.341	-0.087	-1.233	-1.213	-0.846
	(0.382)	(0.329)***	(0.291)	(0.324)***	(0.479)**	(0.268)***
Turning point		0.437		0.530	0.358	0.502
p-value		[0.000]***		[0.007]***	[0.029]**	[0.014]**
Number of Individuals	1,072	1,054	1,071	1,055	1,071	1,055
Sanderson-Windmeijer Linear p-value	0.000	0.000	0.000	0.000	0.000	0.000
Sanderson-Windmeijer Quadratic p-value	0.000	0.000	0.000	0.000	0.000	0.000
Hansen J test p-val	0.158	0.223	0.411	0.481	0.567	0.294

Notes: This table explores an additional political economy mechanism underlying the baseline IV result from column 2 in Table 2. Each pair of columns splits the sample by whether the given village's largest Outer-Island ethnic group constitutes a share of the Outer-Island population in the given region—province, district, or subdistrict—that is above (high) or below (low) the median in the sample. For example, the village of Sanuanggamo has 54 percent Inner-Island ethnic share and 39 percent Tolaki, an Outer-Island ethnic group native to Southeast Sulawesi, in a district where Tolaki comprise 65 percent of the population. In this village, the largest Outer-Island group has an above median representation at the district level in column 4. Such measures are computed for all administrative levels. Note, though, that ethnicity is a key mobilizing force at the district and subdistrict levels where contestable resources are often at the discretion of officials elected or appointed through majority voting schemes. The province is not a relevant political unit in the post-1998 era. See the notes to Table 2 for details on the reported instrument and turning point diagnostics. Note that the turning point falls outside the unit interval in 1 and 3 and hence is not computed. Standard errors in all columns are clustered by district. */**/*** denotes significance at the 10/5/1 percent significance levels.

Table 7: Ethnic Diversity and Intergenerational Transmission of National Identity

	Dependent Variable: Indonesian is Child's ...				
	Main Language at Home	Reported Mother Tongue			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Inner-Island ethnic share	1.709	2.269	2.023	1.015	0.844
	(0.489)***	(0.529)***	(0.530)***	(0.332)***	(0.324)**
Inner-Island ethnic share squared	-1.713	-2.316	-2.043	-0.952	-0.764
	(0.534)***	(0.559)***	(0.564)***	(0.350)***	(0.345)**
parents intermarried			0.197		0.147
			(0.038)***		(0.026)***
father has Indonesian mother tongue				0.390	0.384
				(0.026)***	(0.025)***
mother has Indonesian mother tongue				0.466	0.460
				(0.032)***	(0.032)***
Turning point	0.499	0.490	0.495	0.533	0.553
p-value	[0.003]***	[0.000]***	[0.001]***	[0.011]**	[0.037]**
Number of Individuals	12,579	12,579	12,579	12,579	12,579
Sanderson-Windmeijer Linear p-value	0.087	0.087	0.090	0.125	0.127
Sanderson-Windmeijer Quadratic p-value	0.082	0.082	0.085	0.109	0.111
Anderson-Rubin Weak IV Robust p-value	0.004	0.001	0.001	0.012	0.005
Hansen J test p-value	0.535	0.517	0.513	0.544	0.469

Notes: This table reports the effects of village-level and intrahousehold ethnolinguistic diversity on children's ethnolinguistic identity captured in column 1 by an indicator for whether the child reports *Bahasa* Indonesia as the main language used at home, and in columns 2–5 by an indicator for whether the child reports *Bahasa* Indonesia as their mother tongue. The mean in column 1 is 0.166 and in columns 2–5 is 0.158, but note that the latter group is not a strict subset of the former. Moreover, some children report an Indonesian mother tongue while their co-resident mother reports primarily speaking another language. We consider this more stringent test of national identity in Appendix Table A.3. We define children based on their relationship to the household head, and although this includes some co-resident adult children, 96 percent of the sample is less than age 25. We dummy out households where the spouse is not resident. These results are based on the 1995 *Supas* survey sample used in Table 4. As in the *Susenas*-based regressions, the instrumental variables are ventiles of the number of initial transmigrants within 10 km of the given village. The Inner-Island ethnic share is based on the 2000 Population Census as in other tables and figures but here restricted to those individuals residing in the village in 1995, which we compute using place of residence five years ago. The intermarried parents variable equals one if the household head and spouse reporting different ethnolinguistic identities (mother tongues). Each parent's national language affinity is then allowed to have a different effect on their child's reported mother tongue. Standard errors in all columns are clustered by district. */**/*** denotes significance at the 10/5/1 percent significance levels.

Table 8: Effects of the Transmigration Program on Nation Building

	ATT Estimate (1)	Control Mean (2)
Dependent Variable	**Panel A**: Primary Language at Home	
Indonesian Language	0.250	0.122
	(0.126)**	
Inner-Island Language	-0.002	0.073
	(0.068)	
Outer-Island Language	-0.248	0.805
	(0.162)	
	Panel B: Intermarriage	
marriage rate	0.019	0.830
	(0.013)	
intermarriage rate (inner, outer)	0.050	0.023
	(0.006)***	
adjusted intermarriage rate (inner, outer)	0.177	0.253
	(0.035)***	

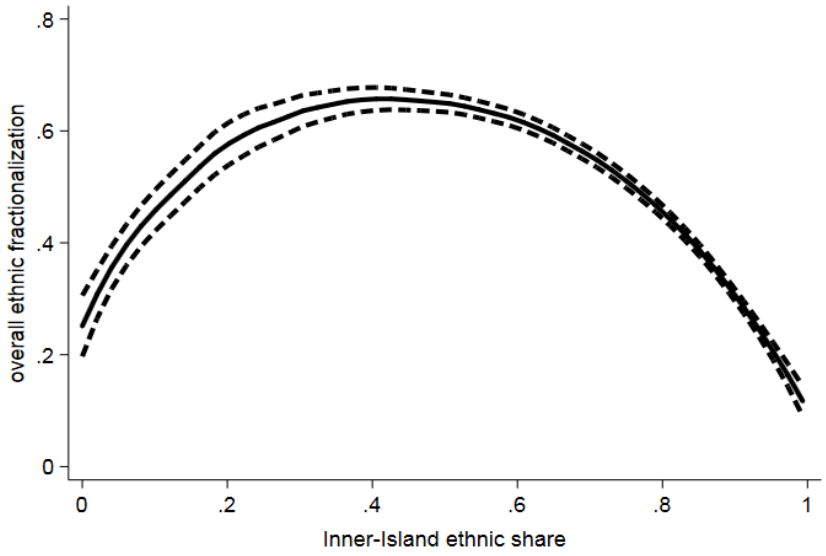
Notes: Each cell in column 1 reports the coefficient from a regression of the given dependent variable listed in the first column on an indicator for whether the village is a Transmigration settlement. Column 2 reports the mean outcome for control villages. Panel A is based on the 2006 *Susenas* survey, and Panel B on the 2000 Population Census. The sample size in panel A is 2,878 individuals comprising 134 villages within 10 km of a Transmigration village and 47 villages within 10 km of a control village, panel B is 1,500 villages comprising 832 Transmigration and 668 control villages. The marriage outcomes are restricted to individuals below an age that implies the marriage would likely have taken place after the initial period of settlement. The adjusted intermarriage rate divides the actual intermarriage by the rate determined by random matching. We estimate the ATT using a control function specification based on a Blinder-Oaxaca decomposition akin to a reweighting approach as derived in Kline (2011). All specifications include island fixed effects and the $\mathbf{x}$ vector of predetermined village-level controls. Standard errors are clustered by district across all specifications. */**/*** denotes significance at the 10/5/1 percent significance levels.

Online Appendix (Not For Publication)

A Empirical Strategy and Additional Robustness Checks

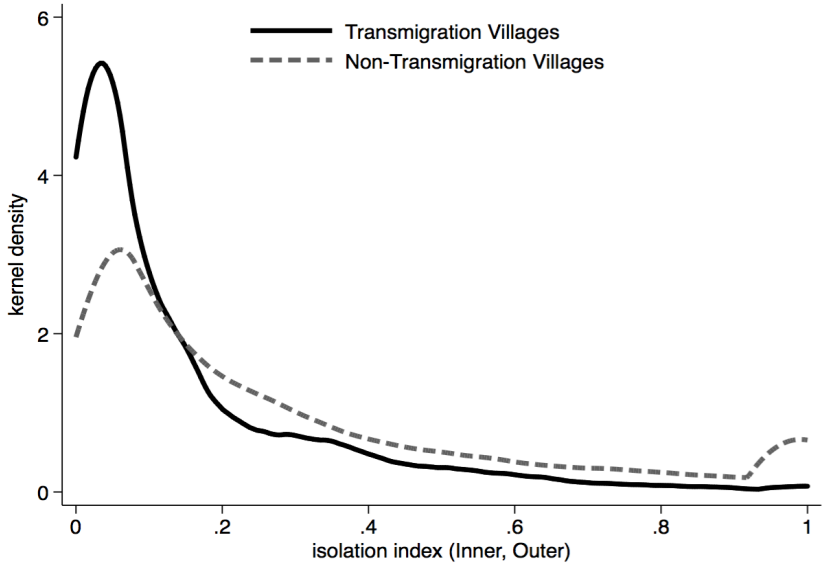
I Sources of Ethnic Diversity

Figure A.1: Inner-Island Ethnic Share and Ethnic Fractionalization



Notes: This figure plots the nonparametric, local linear relationship between the Inner-Island ethnic share and ethnic fractionalization in Transmigration villages in 2000. The estimates are based on an Epanechnikov kernels and rule-of-thumb bandwidth.

Figure A.2: Transmigration Generates Persistently Desegregated Communities



Notes: This figure plots the kernel density of the Bell (1954) residential isolation index capturing block-level segregation between Inner- and Outer-Island ethnic groups in 2000 for Transmigration villages and non-Transmigration villages greater than 30 kilometers from the boundaries of Transmigration villages. For both densities, we employ an Epanechnikov kernels and rule-of-thumb bandwidth.

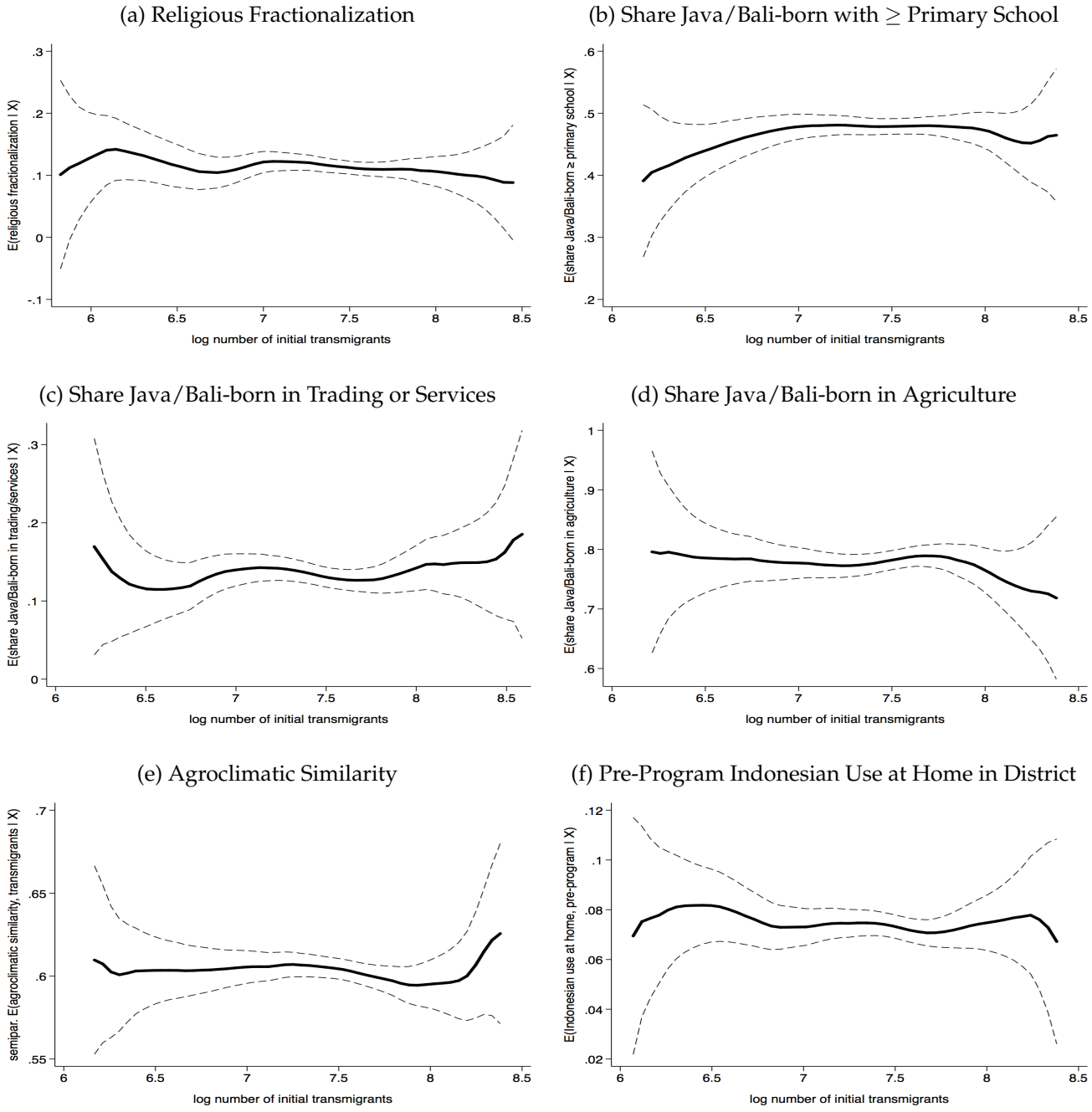
II Probing Instrument Validity

This section describes results mentioned in Section 5 supporting the excludability of the initial number of transmigrants as an instrument for ethnic diversity in 2000. We show that there is not a systematic inverted-U relationship between the initial transmigrant assignment and a wide array of omitted variables potentially correlated with ethnic diversity and integration. These results help rule out confoundedness insomuch as the lack of an inverted U stands in contrast to the strong reduced form, inverted-U relationship in Figure 7.

First, as seen in panels (a)-(d) of Appendix Figure A.3, the instrument does not predict other measures of population diversity such as religious fractionalization, skill levels, or occupational mixes among the initial transmigrants. This suggests that larger settler groups were not mechanically more likely to have greater diversity along other dimensions besides ethnicity. Second, the instrument is uncorrelated with other predetermined proxies for political and economic development not captured in the $\mathbf{x}$ vector used for site selection (see Appendix Figure A.4). These proxies include measures of potential agricultural yields, malaria suitability in 1978, the district-level share of votes going to the Golkar party of President Suharto in the 1977 legislative elections, and a host of district-level characteristics of the population residing within these areas (but not in the immediate settlements) as of 1978, including information on wealth, infrastructure access, schooling, and sector of work.

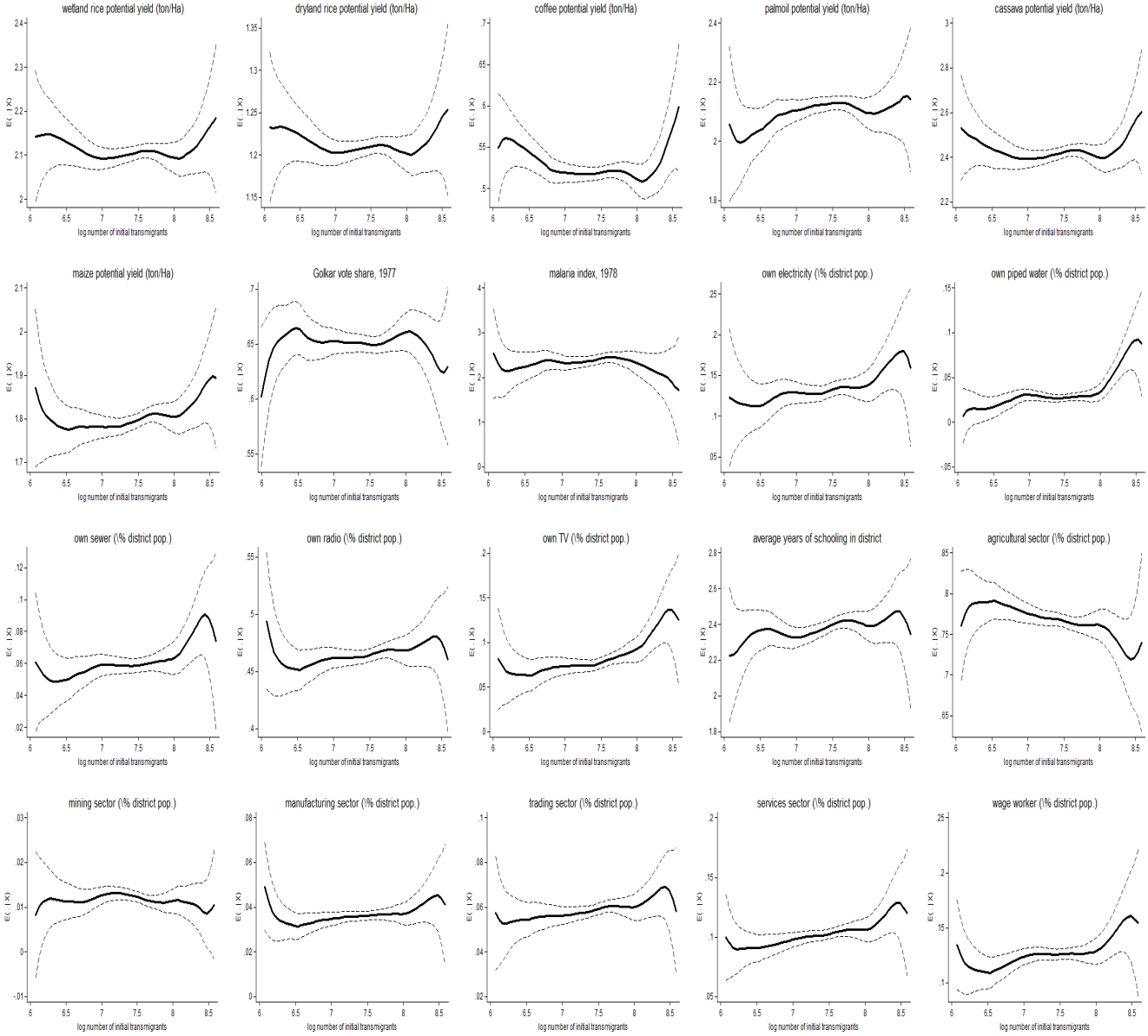
Additionally, panel (e) Appendix Figure A.3 shows a flat relationship with agroclimatic similarity (see Bazzi et al., 2016), suggesting that the assignment rule was not systematically correlated with the transferability of skills among the pool of potential transmigrants at a given point in time. Finally, panel (f) Appendix Figure A.3 shows a similar flat relationship with respect to national language use at home in the late 1970s in areas near the eventual Transmigration settlements but prior to their creation. The lack of an inverted-U relationship here rules out the concern that planners created more diverse settlements in districts with already high levels of national affinity.

Figure A.3: Probing the Validity of the Initial Transmigrant Assignment



Notes: This figure reports semiparametric Robinson (1988) estimates of the relationship between the number of initial transmigrants instrument and other potential confounders of our main diversity measure, the Inner-Island ethnic share. These graphs serve to rule out first order concerns that the instrument for the Inner Island ethnic share (see Figure 3) is correlated with other measures of diversity and population characteristics associated with the initial immigrant influx. We capture in panel (a) religious diversity in the village, (b) the share of transmigrants (born in Java/Bali) with at least primary school, (c) the share of transmigrants working in trading or services, (d) the share of transmigrants in agriculture, (e) the agroclimatic similarity between transmigrants' origin and the given destination, which is a strong proxy for their economic well-being (see Bazzi et al., 2016), and (f) the share of the district that spoke the national language at home in 1978 based on the population residing in the given village's district at the time according to the 1980 Population Census. The semiparametric regression is based on a local linear regression that conditions on island fixed effects and the vector $\mathbf{x}$ of predetermined site selection variables, an Epanechnikov kernel and Fan and Gijbels (1996) rule-of-thumb bandwidth. The top and bottom p percentiles of the x-axis are trimmed for presentational purposes where p varies across figures but is in the 0–4 range.

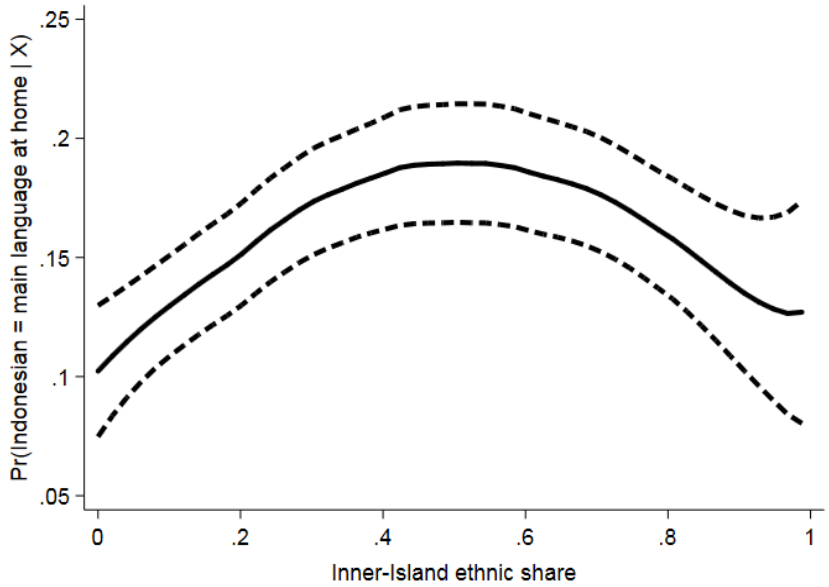
Figure A.4: Initial Transmigrant Assignment Uncorrelated with Predetermined Development Proxies



Notes: These figures report additional semiparametric regression tests relating the instrument to other predetermined measures of political and economic development. The specifications are otherwise akin to those in the prior figure. Potential yields are obtained from FAO-GAEZ. The Golkar vote share is the share of the population in the given district that voted for President Suharto's Golkar party in the 1977 legislative elections. The malaria suitability index is based on work by Gordon McCord, who generously provided us with the data. The variables beginning with "own electricity" 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.

III Robustness to 1995 *Supas* Data

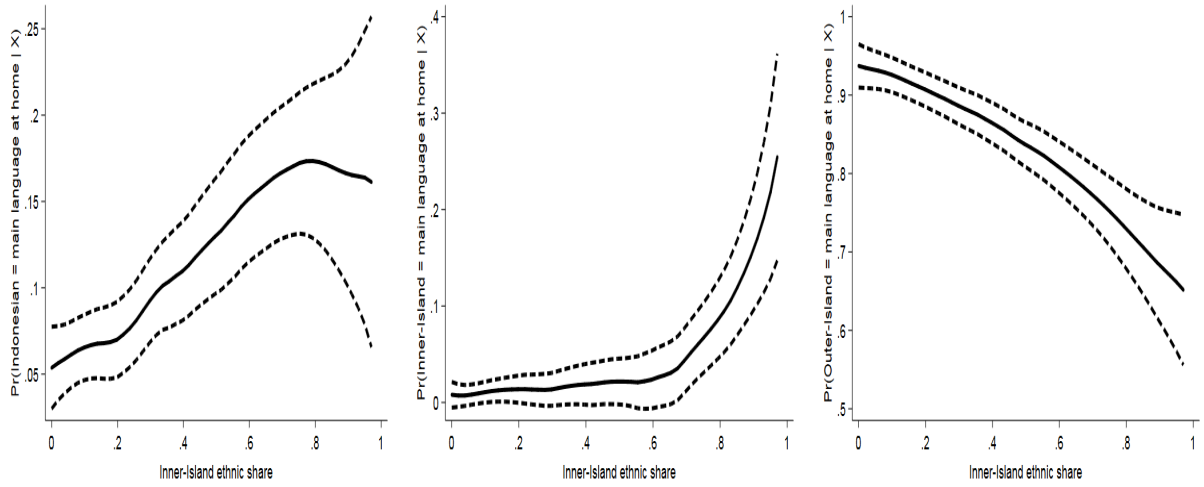
Figure A.5: Ethnic Diversity and National Language Use at Home, 1995 *Supas* Data



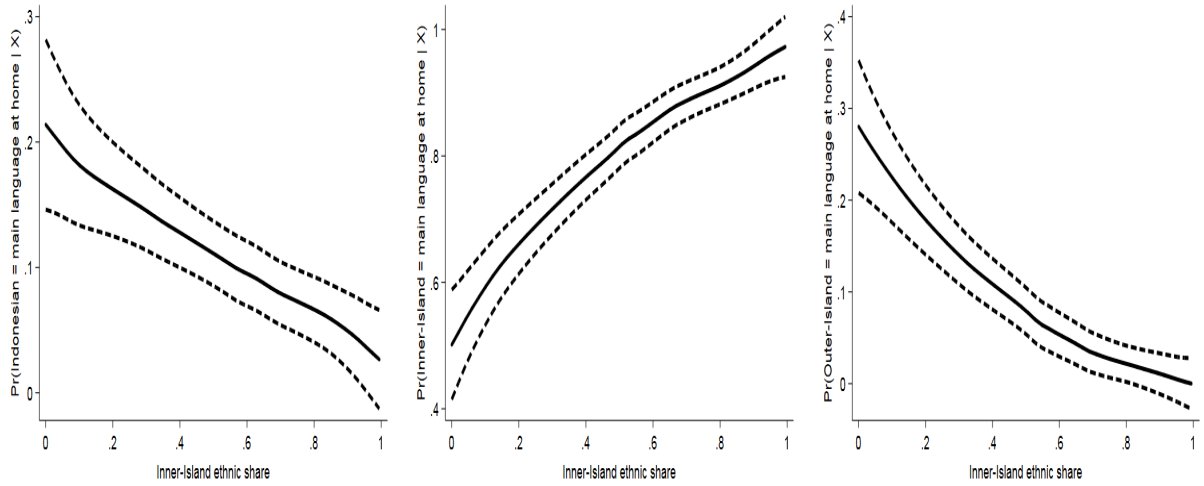
Notes: This figure reports a semiparametric estimate and 95 percent confidence interval for the impact of ethnic diversity (Inner-Island ethnic share) on the likelihood of reporting the national language as the primary language used at home in 1995 *Supas* survey data where the sample includes all individuals within 10 km of Transmigration villages including those villages. The result here is the semiparametric version of the OLS analogue to the IV estimate in column 1 of Table 4. See the notes to that table for further details.

Figure A.6: Ethnic versus National Language Use, by Group, 1995 *Supas* Data

(a) Outer-Island Ethnic Groups



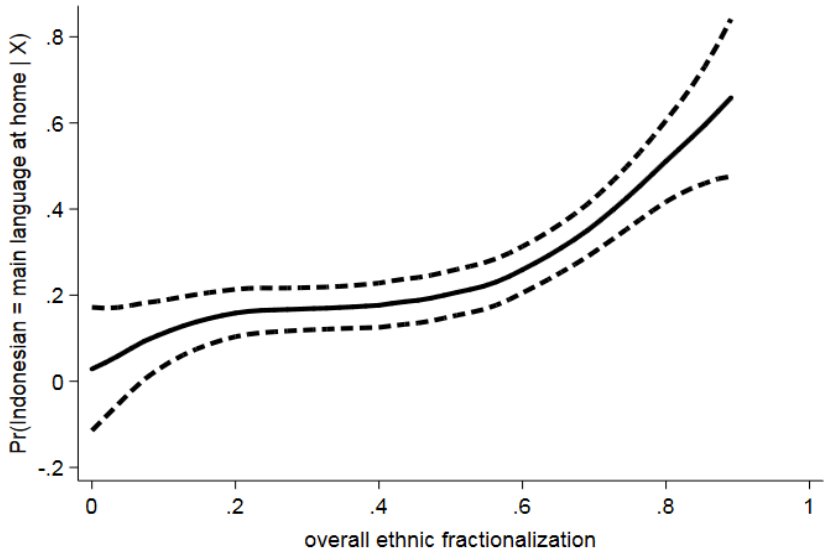
(b) Inner-Island Ethnic Groups



Notes: These figures use the auxiliary *Supas* survey data from 1995 and report an analogous set of semiparametric estimates as in Figure 6. The specifications are otherwise identical to the one in Appendix Figure A.5, based on the Robinson (1988) partially linear model that conditions on our usual **x** vector of control variables and uses an Epanechnikov kernel and Fan and Gijbels (1996) rule-of-thumb bandwidth. Standard errors are clustered at the village level.

IV Robustness to Other Diversity Measures

Figure A.7: Ethnic Fractionalization and National Language Use



Notes: This figure reports a semiparametric (Robinson, 1988) estimate of the effects of overall ethnic fractionalization on the likelihood of reporting the national language as the primary one used at home. The estimate is otherwise based on the same specification as Figure 5. See the notes to that figure for further details.

Table A.1: Other Diversity Metrics and National Language Use

	Dependent Variable: Indonesian is Main Language at Home						
	Baseline		Alternative Diversity Specifications				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Inner-Island ethnic share	0.665						
	(0.284)**						
Inner-Island ethnic share squared	-0.854						
	(0.312)***						
Inner-Island ethnic share, bottom tercile		0.066					
		(0.058)					
Inner-Island ethnic share, middle tercile		0.203					
		(0.059)***					
ethnic fractionalization (Inner, Outer)			0.318				
			(0.142)**				
overall ethnic fractionalization				0.462			
				(0.105)***			
ethnic fractionalization, all Inner-Island					0.165		
					(0.122)		
Javanese share of Inner-Island ethnics						0.777	
						(0.395)*	
Javanese share of Inner-Island ethnics squared						-0.638	
						(0.312)**	
number of initial transmigrants, bottom tercile							0.065
							(0.055)
number of initial transmigrants, middle tercile							0.168
							(0.060)***
Number of Individuals	2,126	2,126	2,126	2,126	2,047	2,047	2,126

Notes: This table reports OLS regressions of the relationship between various measures of ethnic diversity and primary use of the national language at home (mean of 0.25 across columns) beginning with the baseline specifications from Table 2 reproduced in columns 1 and 2. Column 3 is a measure of ethnic fractionalization based only on the two broader groups of Inner- and Outer-Island ethnicities. Column 4 is the fully generalized measure of ethnic fractionalization with all self-reported groups. Column 5 is the ethnic fractionalization among the Inner-Island ethnicities. Column 6 is a quadratic term for the share of the largest Inner-Island ethnicity, Javanese, among Inner Islanders. Column 7 is a version of the reduced form of our IV specification where we split up the instrument into terciles, reporting effects relative to the top third of villages in terms of the number of initial transmigrants. The semiparametric version of this specification is in Figure 7. Standard errors in all columns are clustered by district. */**/*** denotes significant at the 10/5/1 percent significance levels.

V Robustness Across Different Subsamples

Table A.2: Robustness for National Language Use Across Education Level

	Baseline		Sample Splitting by Education Level			
			≤ Primary Education		> Primary Education	
	OLS	IV	OLS	IV	OLS	IV
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Inner-Island ethnic share	0.665	0.845	0.549	0.891	0.813	1.344
	(0.284)**	(0.379)**	(0.289)*	(0.326)***	(0.338)**	(0.534)**
Inner-Island ethnic share squared	-0.854	-0.927	-0.775	-1.041	-0.868	-1.283
	(0.312)***	(0.382)**	(0.311)**	(0.329)***	(0.394)**	(0.589)**
Turning point	0.390	0.456	0.354	0.428	0.468	0.524
p-value	[0.012]**	[0.015]**	[0.032]**	[0.004]***	[0.028]**	[0.037]**
Number of Individuals	2,126	2,126	1,229	1,229	897	897
Dependent Variable Mean	0.249	0.249	0.163	0.163	0.368	0.368
Sanderson-Windmeijer Linear p-value		0.000		0.000		0.000
Sanderson-Windmeijer Quadratic p-value		0.000		0.000		0.000
Anderson-Rubin Weak IV Robust p-value		0.000		0.000		0.000
Hansen J test p-val		0.161		0.139		0.243

Notes: This table reports OLS and IV results from 2, reproduced in columns 1–2, splitting the sample by whether the respondent household head has less than or equal to a primary education in columns 3–4 or more than a primary education in column 5–6. The p-values for the inverted-U turning points are based on the exact test of Lind and Mehlum (2010). Standard errors in all columns are clustered by district. */**/*** denotes significant at the 10/5/1 percent significance levels.

Table A.3: Robustness to Stricter Definition of National Identity

	Dependent Variable: Indonesian is Child's . . .				
	Mother Tongue	. . . and Mother Does Not Report Indonesian as Main Language at Home			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Inner-Island ethnic share ethnic share	2.269	0.749	0.735	0.624	0.614
	(0.529)***	(0.269)***	(0.268)***	(0.234)***	(0.236)**
Inner-Island ethnic share squared	-2.316	-0.781	-0.765	-0.653	-0.642
	(0.559)***	(0.283)***	(0.283)***	(0.247)**	(0.250)**
parents intermarried			0.002		0.001
			(0.015)		(0.015)
father has Indonesian mother tongue				0.228	0.228
				(0.032)***	(0.032)***
mother has Indonesian mother tongue				-0.120	-0.120
				(0.065)*	(0.065)*
Turning point	0.490	0.480	0.480	0.478	0.478
p-value	[0.000]***	[0.005]***	[0.005]***	[0.006]***	[0.007]***
Number of Individuals	12,579	12,579	12,579	12,579	12,579
Sanderson-Windmeijer Linear p-value	0.087	0.087	0.090	0.125	0.127
Sanderson-Windmeijer Quadratic p-value	0.082	0.082	0.085	0.109	0.111
Anderson-Rubin Weak IV Robust p-value	0.001	0.065	0.069	0.073	0.071
Hansen J test p-valu	0.517	0.795	0.783	0.709	0.701

Notes: This table reproduces estimates in Table 7 for a more stringent definition of national linguistic identity captured by the child reporting Indonesian as his/her mother tongue but their co-resident mother reporting a different language as their primary one use at home. The original dependent variable equal to one if the child reports Indonesian as their mother tongue is reproduced in column 1 for reference. The specification is otherwise identical to Table 7. Standard errors in all columns are clustered by district. */**/*** denotes significant at the 10/5/1 percent significance levels.

B Program Impact: Empirical Strategy and Additional Results

I Developing the ATT Identification Strategy

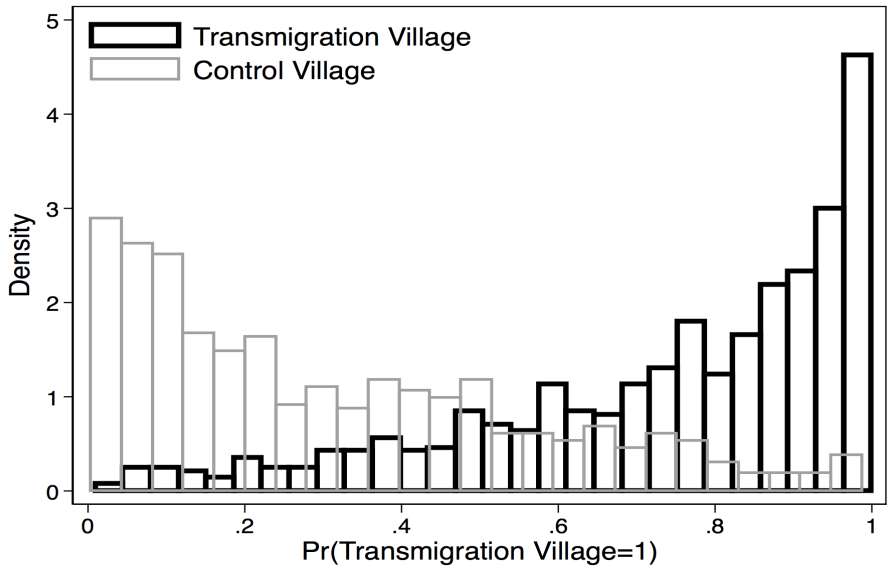
We describe below the reweighting procedure to estimate the ATT in 6. The objective of the procedure is to rebalance the treatment and control villages in terms of predetermined (observable) characteristics. We first predict the probability of being a Transmigration village:

$$\mathbb{P}(\text{Transmigration}_j = 1) = \Lambda(\mathbf{x}_j'\widehat{\boldsymbol{\zeta}}), \tag{B.1}$$

where $\Lambda(\cdot)$ is a logit function. Consistent with the bivariate comparisons in Appendix Table B.1, the conditional estimates of $\boldsymbol{\zeta}$ (which is available upon request) are indicative of sequential site selection among eligible settlement areas. The covariates explain over one-third of the variation in site selection; with province fixed effects, which we use in robustness checks, they explain over half. The estimated probabilities, $\widehat{P}_j$, exhibit considerable overlap across treated and control villages (see Bazzi et al., 2016). We then reweight control village j by its odds of being a Transmigration site, $\widehat{\kappa} = \widehat{P}_j/(1-\widehat{P}_j)$.

This reweighting procedure effectively removes the site selection differentials and brings us closer to a causal interpretation of the ATT in equation (2). As seen in Appendix Table B.1, this helps to rebalance the sample as if planners in 1978 randomly chose Transmigration sites from the initial potential settlements, with selection probabilities based on observables.

Figure B.1: Estimated Propensity Scores



Notes: This figure, also in Bazzi et al. (2016), plots the distribution of estimated probabilities of site selection based on the propensity score regression.

II Demographic Impacts and Spillovers

The reweighting procedure above helps ensure that the counterfactual composition of villages would have been the same on average if not for the Transmigration program facilitating the long-distance migration of Inner Islanders. Appendix Table B.2 documents these persistent demographic impacts of resettlement. While there are no effects on population, Transmigration villages have a 54 p.p. higher Inner-Island ethnic share than control villages where, on average, only 6 percent of individuals claim Inner Islands ethnicity. Furthermore, the overall *ELF* index increases by 0.13 relative to a control group

Table B.1: Baseline Village Characteristics by Treatment and Control Status

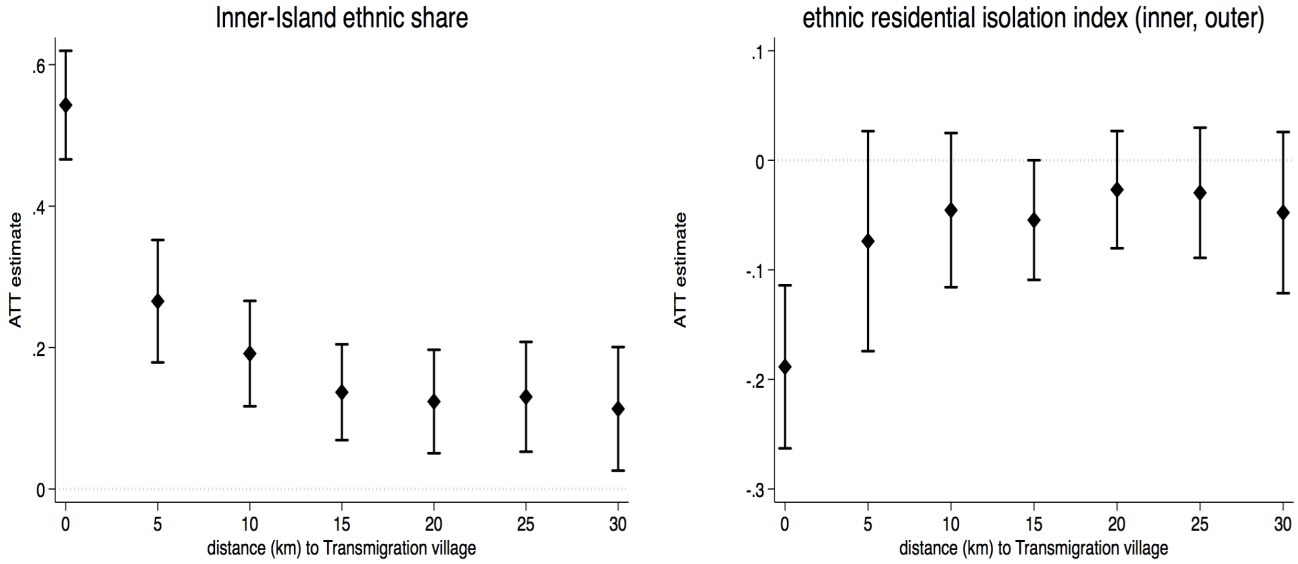
	Transmigration Treated (T)	Non-Treated (NT)	RDA Control (C)	(t-stat) Statistical Difference		
	mean (sd)	mean (sd)	mean (sd)	$\mu_T - \mu_{NT}$	$\mu_T - \mu_C$	Adjusted $\mu_T - \mu_C$
log village area, Ha	7.53 (1.0)	6.99 (1.5)	8.24 (1.3)	5.32***	-4.13***	-0.21
% w/ slope between 0-2%	37.63 (28.3)	30.19 (30.6)	16.96 (17.7)	2.69***	5.88***	0.92
% w/ slope between 2-8%	48.25 (25.4)	39.27 (24.7)	48.21 (21.7)	3.89***	0.01	-0.12
% w/ slope between 8-30%	9.87 (16.5)	20.32 (22.6)	24.73 (19.4)	-6.35***	-5.35***	0.56
Vector Ruggedness Measure	0.31 (0.1)	0.28 (0.2)	0.31 (0.1)	2.81***	0.21	-0.58
log altitude, m^2	3.29 (2.9)	3.77 (2.7)	4.83 (2.2)	-1.91*	-5.08***	1.54
Organic Carbon (%)	4.77 (6.2)	3.53 (4.7)	3.06 (5.2)	2.69***	2.77***	-0.33
Topsoil Sodicity (ESP) %	1.57 (0.4)	1.50 (0.5)	1.63 (0.5)	1.72*	-0.86	1.81*
Topsoil pH (-log(H+))	5.07 (0.4)	5.33 (0.7)	5.35 (0.6)	-5.22***	-2.26**	0.56
Coarse texture soils (%)	0.10 (0.2)	0.06 (0.2)	0.09 (0.2)	3.44***	0.30	-1.12
Medium texture soils (%)	0.70 (0.2)	0.71 (0.2)	0.65 (0.2)	-0.49	1.98*	-0.85
Very poor or poor drainage (%)	0.39 (0.4)	0.30 (0.3)	0.20 (0.3)	3.15***	5.50***	0.87
Imperfect drainage soils (%)	0.06 (0.2)	0.12 (0.3)	0.21 (0.3)	-2.73***	-3.24***	-0.86
Avg. rainfall, 1948-1978	225.26 (35.1)	215.29 (41.4)	237.66 (35.8)	2.39**	-1.70*	0.31
Avg. temp (Celcius), 1948-1978	26.26 (1.7)	25.36 (2.7)	25.75 (1.8)	4.74***	1.77*	0.15
Minimum Log Distance to Villages on Java or Bali	6.69 (0.5)	6.91 (0.6)	6.91 (0.3)	-2.66***	-1.97*	-0.73
Log Distance to Nearest Major Road	0.08 (0.1)	0.07 (0.1)	0.10 (0.1)	1.48	-1.09	-0.29
Log Distance to Nearest Coast	10.56 (1.1)	9.96 (1.5)	10.84 (0.9)	4.32***	-1.65	1.07
Log Distance to Nearest River	8.09 (0.8)	7.95 (1.1)	8.22 (0.8)	2.06**	-1.54	0.45
Log Distance to Subdistrict Capital	2.43 (1.5)	1.73 (1.6)	1.97 (1.8)	7.54***	3.35***	1.18
Log Distance to District Capital	4.12 (1.0)	3.46 (1.4)	4.10 (1.1)	7.33***	0.13	2.40**
linguistic similarity with Java/Bali languages	0.58 (0.1)	0.60 (0.1)	0.60 (0.1)	-2.36**	-1.61	-1.09

Notes: This table reports the sample means (standard deviations, sd) for the predetermined village-level characteristics that comprise our main covariate vector $\mathbf{x}$, which planners used to inform site selection. We consider three groups of villages: Transmigration villages settled in the period 1979–1988 or treated sites (T), non-Transmigration villages or non-treated sites (NT), and Recommended Development Areas (RDA) or control sites (C) that were suggested as resettlement areas but never received the program due to sudden budgetary cutbacks. Note that C villages are a subset of NT villages. The t-statistics reported in the latter three columns are recovered from the coefficient on the treatment variable in a regression of the given characteristic on the treatment indicator. The final column reweights the control observations by their inverse probability of being a Transmigration village recovered from a first step estimate of the propensity score. Standard errors are clustered at the district level. */**/*** denotes significance at the 10/5/1 percent level.

mean of 0.23, whereas there are null effects on religious diversity. The program also led to a nearly 50 percent reduction in the Bell (1954) isolation index, some of which can be explained by the random allocation of plots to farmers when villages were initially settled.

We show in Appendix Figure B.2 that the impacts of the Transmigration program on local ethnic diversity spill over to neighboring villages outside the immediate settlement boundaries. In particular, we estimate a version of the village-level ATT in Table 8 in which we set $Transmigration_j = 1$ for all villages that are within five kilometer discs (0–5, 5–10, ...) from the centroid of Transmigration villages, which are excluded, and $Transmigration_j = 0$ for all villages that are within the given distance from the centroid of the control villages, which are also included. The graphs point to significant populations of Inner-Islanders up to 10 kilometers beyond the Transmigration village boundaries, leveling off albeit remaining positive thereafter. We find similar patterns for ethnic residential segregation, although the effects go to zero by the 10 kilometer radius, which is consistent with the initial residential lottery in Transmigration areas facilitating persistent local mixing within census blocks. These spillover effects can be explained by several forces including, among others, the expansion of the original settlement area. Regardless of the causes, these multiplier effects are important from a policy perspective in that they suggest resettlement programs can have meaningful social impacts beyond the population within the immediate settlement areas.

Figure B.2: Diversity Spillovers from the Transmigration Program



Notes: These figures report the 95 percent confidence intervals around the ATT estimates for the given dependent variable estimated at different spatial lags from the centroid of the Transmigration village. The "0 km" distance is the specification from Table B.2. The remaining estimates are based on 5 km discs radiating outward from the Transmigration villages. For example, the estimate at 5 km is based on all villages outside the Transmigration village boundary but within 5 km of that boundary, and the estimate at 10 km is based on all villages greater than 5 and less than 10 km. The control group at the given distance includes all villages less than that distance from the control village boundary and including the control village itself. See the notes to Tables 8 and B.2 for further details.

Table B.2: Demographic Impacts of the Transmigration Program

	ATT Estimate (1)	Control Mean (2)
log population	0.068	7.215
	(0.088)	
Inner-Island ethnic share	0.540	0.061
	(0.038)***	
residential isolation index (inner, outer)	-0.184	0.340
	(0.037)***	
ethnic fractionalization	0.126	0.238
	(0.031)***	
religious fractionalization	0.011	0.186
	(0.022)	

Notes: This table reports ATT estimates of the impact of the Transmigration program on demographic outcomes in the Outer Islands using the same specification as in Table 8. Standard errors are clustered by district. */**/*** denotes significance at the 10/5/1 percent level.

III Robustness Checks on National Language ATT Results

Table B.3 shows that the ATT results for language use at home are robust to accounting for several individual- and village-level confounders. First, controlling for basic demographics and dummies for years of schooling (row 2) leaves the ATT estimates unchanged as does restricting to those less than the median age of 40 (row 3). These specifications rule out concerns about cohort composition effects and exposure to *Bahasa Indonesia* in primary school. Second, although individuals in certain occupations are more inclined to use Indonesian on a daily basis, the main ATTs are robust to controlling for 17 exhaustive occupation dummies (row 4), and to controlling for household expenditures per capita (row 5). Although perhaps jointly determined with language decisions (and due to the Transmigration program), these controls help rule out potential channels through which the program hastened the diffusion of the national language at home. Additionally, the effects cannot be explained by differential ease of speaking Indonesian as seen in rows 6 and 7, which control for linguistic similarity between the local indigenous language and Malay (the root of Indonesian).

IV Robustness Checks on Intermarriage ATT Results

Accounting for the Transmigrant Supply Effects. We begin by showing that the changes in intermarriages are not entirely due to supply effects. We use two reduced form approaches to adjust for aggregate supply effects at the village level.[1] First, we divide the actual intermarriage rate by the average intermarriage rate from 10,000 simulations of random matching among the young, married population. This measure allows us to identify how much intermarriage we observe relative to what would be expected solely from the policy-induced shock to the supply of potential non-coethnic partners. We show the results in the last row of Table 8. In the typical control village, the actual intermarriage rate is only 25.3 percent of the (random) potential intermarriage rate. The ATT implies that this ratio increases by 17.7 p.p. for Transmigration villages.

Additionally, the village-level specification also allows us to control flexibly for aggregate supply effects using linear, quadratic, or cubic terms of the random intermarriage rate as "bad controls" in the ATT equation. The conclusions are similar; the nonlinear controls reduce the ATT to around 0.03 (instead of 0.05), but the effects remain statistically significant at the 1 percent level. These reduced form adjustments imply that around two–thirds of the ATT effect for intermarriage among the younger generation can be explained by a change in marriage preferences.

Robustness Checks: Selection and Spillover Effects. We show here that the main village-level results on intermarriage rates in Table 8 survive several important robustness checks. In Appendix Table B.4, Panel (A) includes province fixed effects, (B) includes fixed effects for the indigenous ethnolinguistic group in each village, and (C) includes a control for the number of INPRES primary schools per student at the district level in the 1970s as used in Duflo (2001) as well as a set of contemporaneous controls for public goods observed in 2003 used by Martinez-Bravo (2017).[2] The latter public goods controls are potentially endogenous but nevertheless instructive as the large intermarriage effects are robust to their inclusion. More generally, the stability of the ATT across panels suggests that other place-based unobservables potentially correlated with the Transmigration program are not driving the intermarriage effects.

The apparent lack of selection-on-unobservables is consistent with more formal tests of coefficient

[1] This exercise treats the village as the marriage market. If we used the district as the marriage market, we would likely have smaller supply adjustments. This is because supply effects due to the program are concentrated at the village level, and quite muted at the district level. Therefore, supply adjustments at the village level are more conservative. It is also important to note that the regression results in Tables 8 and B.2 are robust to including all villages within 5 kilometers of Transmigration village centroids. This helps rule out concerns about treatment misclassification due to noise in the underlying shapefiles.

[2] The public goods include the number of schools, different types of health clinics, doctors, midwives, and basic transport and sanitation infrastructure.

Table B.3: Robustness Checks on the ATT for National Language Use at Home

	Dependent Variable: Pr(. . . =main language at home)		
	Indonesian	Inner Island	Outer Island
	(1)	**(2)**	**(3)**
1. Baseline ATT	0.250	-0.002	-0.248
	(0.126)**	(0.068)	(0.162)
Number of Individuals	2,878	2,878	2,878
Control Group Mean	0.122	0.073	0.805
2. Conditional on age, gender, education	0.256	-0.005	-0.252
	(0.118)**	(0.070)	(0.159)
3. Individuals aged ≤ 40	0.249	0.011	-0.260
	(0.125)**	(0.070)	(0.155)*
4. Conditional on age, gender, education, occupation	0.245	-0.002	-0.243
	(0.127)*	(0.068)	(0.162)
5. Conditional on log household expenditures/capita	0.250	0.000	-0.250
	(0.127)**	(0.068)	(0.164)
6. Conditional on Malay indigenous language	0.248	-0.003	-0.244
	(0.131)*	(0.070)	(0.165)
7. Conditional on indigenous language distance to Malay	0.243	-0.011	-0.231
	(0.123)**	(0.071)	(0.158)

Notes: Each cell corresponds to a separate individual-level, linear probability regression for the ATT of living in a village within 10 km of a Transmigration site (including those sites). The control group includes all individuals within 10 km of a control/RDA site. The dependent variables, which are mutually exhaustive, are indicators for whether or not the individual household head reports in column (1) the national language *Bahasa Indonesia*, (2) an Inner Island language, or (3) an Outer Island language as the primary daily language used at home. All regressions are based on the Blinder-Oaxaca decomposition approach of Kline (2011). The data come from the 2006 *Susenas* household survey. Row (2) controls for household head gender, age, age squared, and an exhaustive set of indicators for education level; (3) restricts to individuals less than 40 years old; (4) augments the row 2 specification with an exhaustive set of indicators for occupation; (5) controls for log household expenditures per capita; (6) includes an indicator for whether the villages lies within the ethnolinguistic homeland of the Malay (the root language of Indonesian); (7) controls for distance of the given ethnolinguistic homeland to Malay using the metric suggested in Fearon (2003). Standard errors are clustered by district. */**/*** denotes significant at the 10/5/1 percent significance levels.

stability using using the Oster (forthcoming) procedure.[3] Specifically, we compute measures of $\delta = \left(\frac{\theta_c}{(\theta_u - \theta_c)}\right) \times \left(\frac{R_c^2 - R_u^2}{\beta \times R_c^2}\right)$ for the intermarriage and adjusted intermarriage outcomes where θ_u is the ATT in baseline regression from Table 8 and θ_c is the ATT with additional covariates in Appendix Table B.4, R_u^2 and R_c^2 are the corresponding R^2 measures, and β is a scaling parameter that we set equal to 0.3 following common practice, but insights are similar for larger values. Oster (forthcoming) argues that measures of $|\delta| > 2$ are consistent with quite limited selection-on-unobservables. Looking across specifications, we generally find $\delta > 10$ for intermarriage and for adjusted intermarriage, which is reassuring given that the R^2 moves quite a bit with the different sets of controls.[4]

[3] This procedure builds on Altonji et al. (2005) by accounting for the possibility that the additional covariates beyond the baseline do not add much explanatory power to the regressions.

[4] Concretely, we take the estimated θ and R^2 from Tables 8 and B.4 and find that $\delta = 16.9, \texttt{null}, 12.7, 36.6$ ($\delta = -287.8, 34.2, -52.1, 1.4$) across panels A, B, C, and D for intermarriage (adjusted intermarriage) where the `null` value

Individual-Level, Cohort-Specific ATT. We provide here further background on the individual-level ATT estimates in Figure B.3. The basic strategy for identifying exposure effects is akin to approaches commonly used to study the effects of exogenously assigned immigrants (e.g., Dahlberg et al., 2012; Edin et al., 2003). We group individuals into five-year age bins and report 95 percent confidence intervals based on robust standard errors clustered by districts.

Figure B.3 consistently shows greater intermarriage rates for younger cohorts in Transmigration villages relative to similar cohorts in control villages who were not exposed to the migrant shock. The large, differential ATT effects for the younger cohorts provide reassurance that these are new marriages initiated after the program. The youngest cohort (age 16–20) was less than five years old during program implementation in the 1980s and in some cases not even born yet. These patterns are remarkably stable across all panels with increasingly demanding fixed effects.

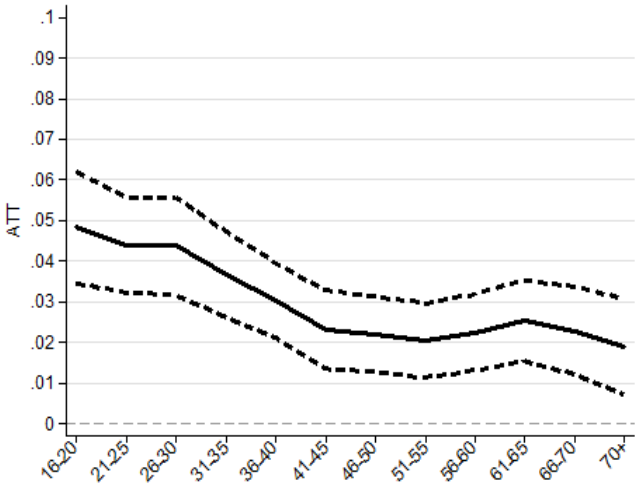
We take additional steps to rule out concerns that the ATT estimates are confounded by selection biases. First, we address the potential threat that the greater integration outcomes in Transmigration villages are driven by tolerant, intermarried individuals who selected into the program *ex ante* or endogenously migrated to Transmigration villages *ex post*. Importantly, the differential cohort effects in Figure B.3(d) remain unchanged when including a further set of exhaustive fixed effects for years of schooling, occupation, and migration status.[5]

is due to the fact that θ is identical despite the change in R^2. Note that we use the column 2 estimates as the Kline (2011) estimate in column 3 does not admit a conventional R^2, but the point estimates of the ATT are similar in column 2.

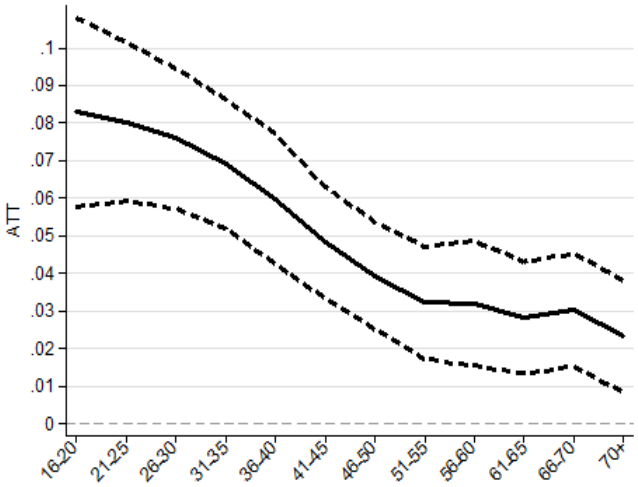
[5]In results that are available upon request, we show that the cohort-specific trends are mostly unchanged when we split the sample by gender, education, or by occupation. We find similar differential, cohort-specific trends for interethnic marriage rates including all ethnic groups rather than just Inner versus Outer Island ethnicities. We can show further that there are null treatment effects across the entire age distribution for marriages in which both spouses are born in Java/Bali, consistent with the discussion above that we are isolating marriages that formed after the program.

Figure B.3: Individual-Level ATT for Intermarriage (Inner, Outer)

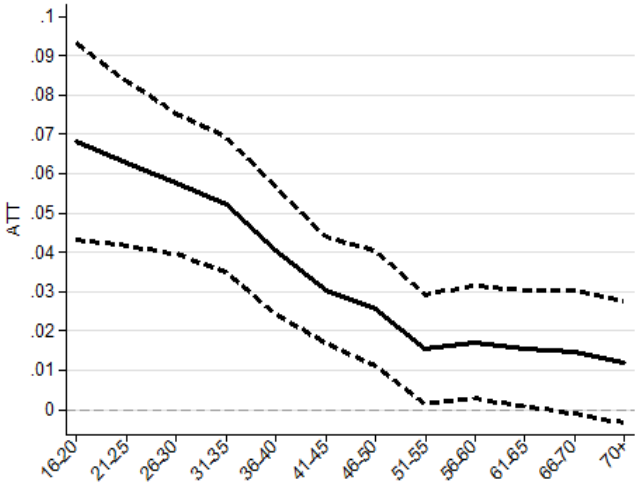
(a) Baseline



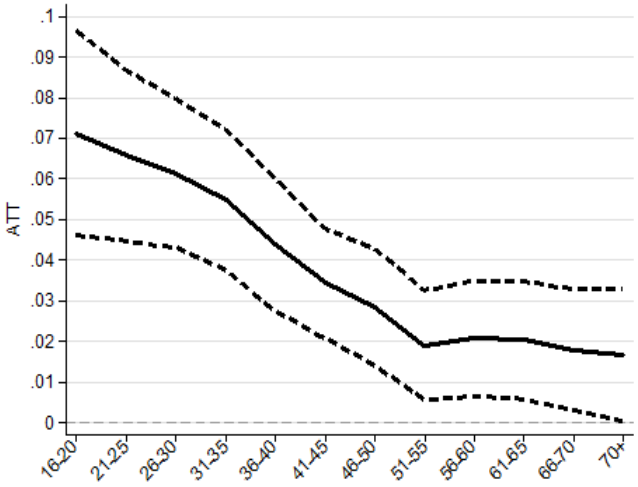
(b) Restricting to Native-Born Outer-Islands Ethnicity



(c) . . . Local to District of Birth and 5 Years Prior



(d) . . . Controlling for Ethnicity Fixed Effects



Notes: These figures plot individual-level, age-specific ATT estimates where the control group is restricted to similarly-aged individuals in RDA villages. Each age-specific ATT identifies the differential likelihood of interethnic marriage between Inner and Outer-Island natives for individuals living in Transmigration villages compared to those living in control villages as reported in the 2000 Population Census. All specifications include island fixed effects and the $\mathbf{x}$ vector of predetermined village-level controls. The specifications in panel (b)-(d) include district of birth fixed effects (FE), panel (b) restricts to individuals belonging to Outer-Islands ethnic groups born in the Outer-Islands, panel (c) further restricts to those residing in the same district as they were born and lived in 1995, and panel (d) further adds fixed effects for the 900 different ethnicities. Sample sizes are (a) 1,215,730, (b) 564,185, (c) and (d) 500,770. These latter refinements hone in on those Outer-Islands natives differentially exposed to the Transmigration program in their locality. The dashed lines are 95 percent confidence intervals based on clustering at the district level, of which there are 94.

Table B.4: Robustness Checks on Intermarriage Impacts of the Transmigration Program

	ATT Estimate (1)	Control Mean (2)
	Baseline Specification	
	Panel A: + Province FE	
marriage rate	-0.001	0.829
	(0.022)	
intermarriage rate (inner, outer)	0.061	0.023
	(0.010)***	
adjusted intermarriage rate (inner, outer)	0.127	0.253
	(0.074)*	
	Panel B: + Ethnolinguistic Homeland FE	
marriage rate	-0.009	0.829
	(0.015)	
intermarriage rate (inner, outer)	0.031	0.023
	(0.008)***	
adjusted intermarriage rate (inner, outer)	0.149	0.253
	(0.050)***	
	Panel C: + (Bad) Controls for Public Goods	
marriage rate	0.029	0.829
	(0.016)*	
intermarriage rate (inner, outer)	0.043	0.023
	(0.007)***	
adjusted intermarriage rate (inner, outer)	0.099	0.253
	(0.044)**	

Notes: This table augments the specifications in Panel B of Table 8 with the covariates listed in the panel headings. See the notes to that table for further details on other aspects of the estimating equations. Panel A includes province fixed effects. Panel B includes fixed effects for the indigenous ethnolinguistic group spanning the given village. Panel C includes all of the public goods investigated in Martinez-Bravo (2017) as well as the number of INPRES primary schools per student as used in Duflo (2001). Standard errors are clustered at the district level. */**/*** denotes significance at the 10/5/1 percent level.

C Transmigration Program: Design and Implementation

The Transmigration program was developed to address to the perceived overpopulation in Indonesia's Inner Islands of Java and Bali. The program subsidized the relocation of agricultural households from rural Java/Bali (transmigrants) to newly created rural settlements in the Outer Islands. First conceived in the Dutch colonial period and revived after independence, the program received a major overhaul in Indonesia's third and fourth Five-Year Development periods (or *Pelita* III and IV) from 1979–1988 under Suharto. In the five years of *Pelita* III (1979–1984), the program resettled 1.2 million people — in contrast to less than 600,000 people under the Dutch colonizers and post-independence under Sukarno and the early Suharto years (1945-1968) (Hardjono, 1988; Kebschull, 1986).[1]

Following the success in *Pelita* III, the government planned for an ambitious expansion in *Pelita* IV (1984–88), targeting 3.75 million people for relocation. However, this plan was derailed by the sudden fall in oil prices in the mid-1980s. From 1984 to 1985, the government resettled around 307,000 people, but as oil prices fell dramatically in 1986, declining revenues forced the government to cut the Transmigration budget by 44 percent in that fiscal year. Consequently, many areas that were designated as potential Transmigration sites (also known as the "recommended development areas" or RDAs) were not developed, and instead, the government focused on developing and maintaining existing Transmigration sites.

There are a number of incarnations of the Transmigration program over the years (Fearnside, 1997). Our analysis focuses on the main form of the program that was implemented in *Pelita* III and IV, known as the general Transmigration program (*Transmigrasi Umum*). The following description, partly adopted from Bazzi et al. (2016), highlights the main features of this general transmigration program on program provisions, individual eligibility and assignment to transmigration locations, and how officials determine the location and carrying capacity of potential settlements.

Program Provisions. The program provided participating households with land and housing (that are allocated by lottery), as well as free transportation to their new homes in newly opened settlements. Prior to their arrival, program officials prepared these sites for immediate agricultural use, and connected them to the road network. They also built houses for transmigrants and prepared 2-hectare plots of agricultural land that are allocated to the settlers by lottery. Settlers received property rights to the homes and land, and they were given provisions for the first few growing seasons, including seeds, tools, and food (Otten, 1986). The government was also supposed to provide temporary agricultural extension services, social infrastructure (e.g, schools, mosques, and health facilities), and support for the development of cooperatives and other social institutions.

Individual Eligibility and Assignment. Participation in the program is voluntary, but participants must meet a set of eligibility criteria. 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 limited schooling (Kebschull, 1986).[2]

The process by which transmigrants were assigned to destination settlements was neither rigorous nor systematic. Many reports indicate that the process suffered from time, information, and institutional

[1] The total program budget in the third and fourth Five-Year Development Plan was approximately $6.6 billion (in 2000 USD) or roughly $3,330 per person moved, which is similar to other large-scale programs implemented in developing countries at the time (see World Bank, 1982, 1984).

[2] Government-sponsored migrants in Indonesia are more comparable to non-migrants 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 their origin district in Java/Bali (based on the 2000 Population Census). 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. While transmigrants surely had unobservable traits similar to other frontier settlers throughout history, they were often among the poorest and least educated members of their home villages. This mitigates some of the first-order confounds of tolerance, but we revisit these concerns about selection when interpreting key results.

constraints that led to haphazard implementation. Coordination problems between government agencies made it infeasible to systematically match transmigrants to settlements. Ultimately, the allocation of transmigrants was largely driven by the coincidental timing of transmigrants' arrival to transit camps in Java/Bali and the opening of settlements in the Outer Islands (Hardjono, 1988). Participants could not choose their destinations and were often ill-informed about the conditions they would face in the new settlements (Levang, 1995; Kebschull, 1986).

An aspect of the program that helped introduce ethnic diversity across Transmigration villages was the allotment of the transmigration quote to local Outer Islanders. To reduce the potential for conflict, promote integration, and encourage the transfer of agricultural knowledge, official guidelines stipulated that a share of each settlement would be allocated to local inhabitants, defined as individuals hailing from the settlement province (Levang, 1995). In 1979, this share—officially known as *Alokasi Pemukiman bagi Penduduk Daerah Transmigrasi* (APPDT)—was recommended to be 10 percent, and it was increased to 20–25 percent in 1982. In practice, as discussed in the paper, these shares departed from these *de jure* stipulations.

Site Selection. Ministry of Transmigration (MOT) officials worked closely with international consultants to select agriculturally viable settlement areas according to several criteria in a three-phase process. 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 development potential. Areas with slopes greater than 8 percent of the net area were deemed unsuitable for the program. Finally, once suitable (or "recommended") areas were found, detailed surveys of topography, hydrography, and soil quality were conducted. These surveys informed final decisions about the location and carrying capacity of settlements.

D Data Appendix

Our study employs multiple datasets that combine information on realized Transmigration placements and planned Transmigration sites with information on demographic, geographic, and socio-economic characteristics at the village and individual levels. Table D.1 summarizes the different datasets used for the analysis. We describe each of these datasets in the following sections.

Table D.1: Summary of Datasets

Dataset	Description	Obs. Unit
Transmigrant placement		
Transmigration census	Location of Transmigration sites; the number of households and individuals, and years placed in each site.	Transmigration site
Reppprot Maps	1:250,000 Regional Physical Planning Program for Transmigration (Reppprot) maps that include "recommended development area" or RDAs.	
Demographic Variables		
Population Census, 2000	Full dataset: Relationship to household head, ethnicity, highest level of schooling, sectoral employment, birth information (year and month, district), district of residence in 1995.	Individual
Socio-Economic Variables		
Susenas, 2006	Language use at home	Individual
Supas, 1995	Relationship to household head, language (i.e., mother tongue and language used at home), birth information (year and month, district)	Individual
Indonesia Family Life Survey (IFLS) 1997, 2014	Language use at home (1997, 2014); own, mother's, and father's ethnicity; relative trust of non-coethnics.	Individual
Agroclimatic Variables		
GIS Map - Dept. Public Works	Village area, distance to coast, roads and others.	
Harmonized World Soil Database	Elevation, ruggedness, soil quality (organic carbon, topsoil characteristics, texture, drainage).	30-arc-second grid
Terrestrial Precipitation and Temperature Data	Rainfall (Matsuura and Wilmott, 2012a) and temperature (Matsuura and Wilmott, 2012b), 1948-1978.	Monthly, interpolated to 0.5×0.5 degree grid

I Transmigration Census and Maps

We employ the Ministry of Manpower and Transmigration's census of Transmigration sites established between 1952 and 1998 to obtain details about the placements of transmigrants. The census identifies the physical locations and names of realized transmigration sites, years of establishment, and the number of transmigrants — that includes the number of individuals and households — at the time of the initial settlement. Our within-program analysis focuses on the 911 transmigration villages established in Indonesia's Third and Fourth Five-Year Development Periods (1979–1988) in the Outer islands, excluding

Papua.[1]

Moreover, to identify the "almost-treated" villages for our average treatment effect analysis, we use the set of maps of locations of planned (but unrealized) Transmigration sites produced by the ministry. Using GIS software, we overlay these planned sites — or "recommended development areas" (RDAs)— onto the 2000 and 2010 digital maps of Indonesia's administrative boundaries (down to the village level) produced by Indonesia's Central Statistical Agency, *BPS Statistics* (hereafter, BPS). We digitally trace these RDAs and define as "almost-treated" (controls) those 907 non-Transmigration villages observed in 2000 that share any area with the RDAs from the 1980s. As a baseline, we exclude control villages that are within 10 kilometers of Transmigration settlements to minimize potential bias from spillovers. Our conclusions are similar using other cutoffs. This leaves us with 832 treated villages and 668 control villages after accounting for missing data. We identified a total of 907 "RDA villages", defined as villages that shared any area with the RDA polygons.

II Demographic and Socioeconomic Variables

We link several census-, administrative- and survey-based data sources to Transmigration, RDA, and other villages.

Population Census Data, 2000. The 2000 Population Census contains information on Indonesia's demographic characteristics, and was produced by BPS. It contains a rich set of variables on demographic characteristics, including sex, birth information (month-year and district), marital status, education, and district of residence in 1995 as well as religion and ethnicity. It was designed to collect demographic information through a complete enumeration of the all individuals in every household with 100 percent coverage. However, in its implementation, the provincial offices of BPS had to estimate the data for some of the areas due to to communal violence following the 1998 political transition.[2]

An important and unique feature of the 2000 Population Census (compared to existing Indonesian censuses) makes it particularly suitable for our analysis. Unlike previous and subsequent censuses, it is the only Indonesian census that allows researchers to identify individual ethnicity at a very fine, census-block level. It is the first census in post-independence Indonesia that reports individual ethnicity. Furthermore, while the subsequent 2010 Population Census reported individual ethnicity, BPS removed the subdistrict and village administrative codes in the public version of the 2010 Census microdata with the ethnicity information.

The census was designed to enumerate all households and all individuals living in those households. Moreover, it also asks whether an individual is a spouse of the head of the household. We are therefore able to use it to construct the characteristics of household heads and their spouse(s) (including whether the couple is ethnically intermarried) at the village level.

***Survei Sosial Ekonomi Nasional (Susenas)*, 2006.** *Susenas* is a nationally representative household survey that collects detailed data on Indonesian households. In addition to the core household module, *Susenas* collects rotating, supplementary modules on different topics. The 2006 Socio-Cultural Module of *Susenas* includes a question on the primary language used daily by the household head. Despite the relatively small sample of households within each village, the *Susenas* data provides enough variation to identify key patterns of language use even as we face limitations in statistical power for some purposes.[3]

[1] In the online appendix of Bazzi et al. (2016), we describe in detail how we constructed this dataset from the original Transmigration census.

[2] The areas where data were estimated instead of enumerated are in the provinces of Aceh, Maluku, Papua, and Central Sulawesi (Surbakti et al., 2000)

[3] Although individual language use is reported in the 2010 Census, this information (like the ethnicity information) is not linked with village-level administrative codes as required for our analysis.

***Survei Penduduk Antar Sensus (Supas)*, 1995.** *Supas* is the decadal intercensal population survey conducted on the mid-point between two population censuses. Like the census, *Supas* collects detailed population characteristics information. *Supas* 1995 includes questions on demographic characteristics and migration, as well as information on individual language use. It is notable that with respect to language use, *Supas* 1995 distinguishes between mother tongue and the language used at home. However, unlike the population census, *Supas* only covers a representative sample of the Indonesian population. Moreover, unlike *Susenas 2006*, *Supas* does not have information on individuals' (self-reported) ethnicity.

Note that given their random sampling strategies, *Supas* and *Susenas* cover different subsets of Transmigration villages.

Indonesia Family Life Survey (IFLS). IFLS is a longitudinal household dataset that was collected between 1993 and 2015. Five waves of data collection had been conducted in 1993, 1997, 2000, 2007, and 2014. Over the span of more than two decades, IFLS tracked all individuals from the 7,224 households in the first wave with a very low attrition rate (of less than 10 percent) between IFLS1 and IFLS5 (Strauss et al., 2016). In particular, it tracks individuals who left their original (IFLS1) households, either due to the formation of new households or emigration out of their villages.

IFLS has an incredibly rich set of variables. Included among the rich set of IFLS variables are (reported) ethnicity, the ethnicity of an individual's mother and father, language spoken at home, and discriminative attitudes (with respect to ethnicity). These variables are collected for all members of the surveyed households members. We use these variables to identify individuals who were brought up in households that use Indonesian at home, as well as those whose parents are of mixed ethnicity.

III Spatial, Topographical, and Agroclimatic Variables

We include geographical characteristic and climatic variables to construct the controls for natural endowments. These include measures of: (i) topography (land area, elevation, slope, ruggedness, and altitude), (ii) pre-program market access (distance to (sub)district capitals, roads, rivers, and the sea coast), and (iii) soil quality such as texture, drainage, sodicity, acidity, and carbon content. Many of these variables are explicitly listed in program manuals from 1978 in the MOT archives that provided guidance for site selection. We construct these variables from a variety of sources. Below, we briefly discuss the construction and sources of these variables. The online appendix of Bazzi et al. (2016) provide more details of the variable construction procedures.

Distances and Map Projection. Data for the shapefiles for Indonesia's rivers, roads, major cities, and coast lines were all provided by Indonesia's Department of Public Works (*Departemen Pekerjaan Umum*). Using GIS, we constructed the distance from each village polygon in the dataset to the coast, the nearest river, the nearest road, and major cities using the Euclidean distance tools from ArcView.

Slope, Aspect, and Elevation Data. We construct the topographical variables using raster data from the *Harmonized World Soil Database* (HWSD), Version 2.0 (Fischer et al., 2008).[4] We use the raster data to compute the average elevation, slope, and aspect over the entire polygon for each village. For the slope variables, we the average share of each village corresponding to each slope class (0-2 percent, 2-4 percent, etc.) using ArcView.

Ruggedness. A 30 arc-second ruggedness raster was computed for Indonesia according to the methodology described by Sappington et al. (2007), and village-level ruggedness was recorded as the average raster value. The authors propose a Vector Ruggedness Measure (VRM), which captures the distance or dispersion between a vector orthogonal to a topographical plane and the orthogonal vectors in a neighborhood of surrounding elevation planes.

[4]Data from the HWSD project are publicly available at http://www.iiasa.ac.at/Research/LUC/luc07/External-World-soil-database/HTML/index.html?sb=1

Soil Quality Covariates. HWSD provides detailed information on different soil types across the world. For Indonesia, the data come from the FAO-UNESCO Soil Map of the World (FAO 1971-1981). We created for each village the following measures of soil types: percentage of land covered by coarse, medium, and fine soils, percentage of land covered by soils with poor or excessive drainage, average organic carbon percentage, average soil salinity, average soil sodicity, and average topsoil pH.

Rainfall and Temperature, 1948–1978. The database of Matsuura and Wilmott (2012b,a) at the Department of Geography, University of Delaware compiles monthly temperature and rainfall data across the globe. The monthly data for Southeast Asia come from the Global Historical Climatology Network v2 (GHCN2) database, which were interpolated to estimate monthly precipitation and temperature to a 0.5×0.5 degree (or 55 km) resolution grid (Matsuura and Wilmott, 2012b,a). For the districts in our dataset, we averaged the numbers provide by the database for the period of 1948–1978 to obtain the predetermined measures of rainfall and temperature.

IV Linguistic Distance: World Language Mapping System (WLMS) and Ethnologue

Finally, we use the *World Language Mapping System* (WLMS) to construct our measure of linguistic distance. WLMS uses the sixteenth edition of the *Ethnologue* database and maps each of 6,909 living languages recorded in the database to its relevant location. There are more than 700 ethnolinguistic groups in its entries for Indonesia, including eight ethnolinguistic groups indigenous to Java/Bali.[5] We then map these ethnolinguistic groups to those recorded in the 2000 Population Census.

V Constructing Key Variables

Below, we provide more details on the construction of the key variables in our main analysis. We discuss the two main nation-building outcomes: intermarriage rates and language used at home.

Intermarriage Rates. In our study, we define intermarriage as a marriage between a transmigrant and non-transmigrant ethnicity. The data come from the full 2000 Population Census micro-dataset. We use individual-level information on ethnicity and relationship to the household head to construct the intermarriage rates. The main sampling frame is based on the household heads and their spouse(s).

We are unable to to identify secondary couples in extended households since the data only provide information on relationship to the household head. However, among all married individuals in our study area, only around 5 percent is neither a household head nor a spouse, suggesting that this measurement error is relatively limited. Moreover, we develop an algorithm to identify some subset of these other marriages, and doing so leaves all of our results unchanged. Hence, we maintain the cleaner household head measure as our baseline.

We aim to isolate changes in marriage patterns for cohorts who were likely to have been affected by the program (had they lived in a Transmigration settlement). We therefore investigate intermarriage effects across the age distribution and focus part of our analysis on young couples. For treated villages, a young couple is one in which the ages of both spouses were less than 15 (the legal age of marriage) in the village's year of settlement. For control villages, we define "the year of settlement" as the median year of settlement of treated villages in the same district or province. By 2000, the average age for these young individuals is around 25, and 84 percent of them are married. Our results are robust to alternative definitions of young cohorts.

Language Use at Home. The data to construct the variable for the language used at home come from the Socio-Cultural Module of *Susenas* 2006 dataset. As discussed above, the module contains a question on the primary language used at home. In the case where households employ multiple languages at home,

[5]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.

enumerators were encouraged to ask for the language used most frequently (Central Bureau of Statistics, 2006, p.8). We use responses to this question to create dummy variables of whether the primary language used at home is Indonesian, a transmigrant-ethnic language, and a non transmigrant-ethnic language.

Linguistic Similarity. We use the *Ethnologue* database and the WLMS map to construct the linguistic similarity measure. The database contains the linguistic tree, which shows how different languages and dialects are related among the different language families. This linguistic tree allows us 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. Our *linguistic similarity* for village j is given by:

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

where $\pi_{\ell j}$ is the share of immigrants in village j from ethnolinguistic group ℓ in Java/Bali (from the 2000 Population 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). For each village j, we deem the native language to be the WLMS linguistic-homeland polygon with maximum coverage of village area.

Agroclimatic Similarity. The agroclimatic similarity measure captures the similarity in the agroclimatic environments between migrant origins and destinations. As in Bazzi et al. (2016), we construct this variable using the spatial, topographical, and agroclimatic variables described in AppendixD. III above. All land attributes are either time-invariant or measured before the villages we study were created, and hence do not reflect settler activities.

The agroclimatic similarity between an individual's origin i and her destination j is defined as:

$$agroclimatic\ similarity_{ij} \equiv \mathcal{A}_{ij} = (-1) \times \mathrm{d}\left(\mathbf{x}_i, \mathbf{x}_j\right) \tag{D.2}$$

where $\mathrm{d}\left(\mathbf{x}_i, \mathbf{x}_j\right)$ is the agroclimatic distance between locations i and j, using a metric defined on the space of agroclimatic characteristics. We observe origins at the district-level and hence construct the index based on measures of $\mathbf{x}$ in the destinations at that same spatial frequency. We use the sum of absolute deviations as the distance metric, converting each characteristic to z-scores before taking the absolute difference between origins and destinations. 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. We multiply by (-1) so that larger differences correspond to lower values of agroclimatic similarity.

An agroclimatic similarity index for location j is then calculated by aggregating the individual $\mathcal{A}_{ij}$ across i using population weights:

$$agroclimatic\ similarity_j \equiv \mathcal{A}_j = (-1) \times \sum_{i=1}^{I} \pi_{ij}\, \mathrm{d}\left(\mathbf{x}_i, \mathbf{x}_j\right), \tag{D.3}$$

where π_{ij} is the share of migrants residing in Transmigration village j who were born in district i (calculated using the 2000 Population Census microdata). We use π_{ij} terms based on all individuals born in Java/Bali.

Table D.2: Summary Statistics

	Mean	Std. Deviation
(a) Demographics and Residence		
total population	2,010	(1,306)
Inner-Island born population share	0.38	(0.20)
Inner-Island ethnic share	0.67	(0.31)
ethnic fractionalization	0.42	(0.21)
number of ethnic groups	23.65	(20.18)
ethnic residential isolation index (inner, outer)	0.15	(0.18)
(b) Primary Language at Home		
All Individuals		
National Language	0.25	(0.43)
Outer-Island Language	0.46	(0.50)
Inner-Island Language	0.29	(0.46)
Inner-Island Natives		
National Language	0.26	(0.44)
Outer-Island Language	0.08	(0.26)
Inner-Island Language	0.66	(0.47)
Outer-Island Natives		
National Language	0.24	(0.43)
Outer-Island Language	0.74	(0.44)
Inner-Island Language	0.02	(0.14)
(c) Marriage among Young Cohort		
marriage rate	0.85	(0.14)
intermarriage rate (inner, outer)	0.08	(0.07)
adjusted intermarriage rate (inner, outer)	0.39	(0.37)

Notes: Panels (a) and (c) report summary statistics for the 832 Transmigration villages in our full sample. These results are based on the universal Population Census from 2000. Panel (b) reports summary statistics for language use at home at the individual-level for all villages within 10 kilometers of Transmigration villages. These results are based on the nationally representative 2006 *Susenas* household survey, which reports primary language used at home as reported by the household head. We group the different languages into those spoken by native Inner Island and native Outer Island ethnic groups and report summary statistics separately for all individuals, those identifying with ethnic groups native to the Inner Islands, and those identifying with ethnic groups native to the Outer Islands. The marriage variables are defined for those within age groups plausibly marriageable after the initial years of settlement in the 1980s. These are village-specific age cutoffs that broadly cut off ages around 30–40 in the year 2000. The intermarriage rate captures the fraction of all marriages that are between Inner- and Outer-Island ethnic groups. The adjusted intermarriage rate divides that actual rate by the implied intermarriage rate given by simulated random matching between the married population.