# Unity in Diversity?
# Ethnicity, Migration, and Nation Building in Indonesia*

Samuel Bazzi† *Boston Univ., CEPR and BREAD*

Arya Gaduh‡ *Univ. of Arkansas*

Alexander Rothenberg§ *RAND Corporation*

Maisy Wong¶ *Univ. of Pennsylvania*

February 2017

**Abstract**

Nation building is a major policy challenge in many countries. As population mobility increases, governments are increasingly interested in promoting social integration and a national identity among groups from diverse cultural backgrounds. We use a large resettlement program in Indonesia—an expansive archipelago home to several hundred ethnolinguistic groups—as a natural experiment to investigate the relationship between diversity and nation building. Between 1979 and 1988, the Transmigration Program created several hundred new communities with varying degrees of ethnic diversity by relocating two million voluntary migrants from the Inner Islands of Java and Bali to rural settlements in the Outer Islands. These migrants could not choose their destinations, and by exploiting certain features of the planning and implementation process, our empirical strategy addresses key challenges in identifying the causes of long-run integration. Using rich microdata on marriage, language use at home, and intergenerational ethnic identity choices, we find substantial changes in social identity behavior consistent with deeper ethnic integration amidst rising diversity. For choices with social externalities, we identify thresholds in relative group sizes that have important policy implications. We also shed new light on the role of a neutral national language in helping mediate the cultural costs and economic benefits associated with greater diversity. Overall, our findings provide a unique lens into the otherwise slow intergenerational process of breaking down interethnic cleavages and converging towards new forms of shared identity.

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

**Keywords**: Identity, Language, Intergenerational Transmission, Diversity, Migration, Nation Building

*Please email us for the latest version. We thank Oriana Bandiera, Toman Barsbai, Giorgio Chiovelli, Ben Olken, Imran Rasul and seminar participants at Boston University, Georgetown University, Harvard University, University of Denver, University of Southern California, the Barcelona Graduate School Summer Forum, the Ninth International Migration and Development Conference, the 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. 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.

# 1 Introduction

Uniting people from diverse cultures is a founding principle of many nation states.[1] However, as global migration flows increase and cause major demographic shifts within and between countries, maintaining national unity becomes challenging.[2] Some argue that exposure to diverse, new cultures evokes negative sentiments against outsiders and may incite conflict, particularly in the short run (Borjas, 2016; Enos, 2014; Fearon and Laitin, 2011). However, other research posits that these negative sentiments may dissipate as intergroup relationships develop over time and diverse communities become integrated (Allport, 1954; Anderson, 1983; Putnam, 2007). We know little about these longer run integration processes because diverse communities are often unstable due to segregation and tipping forces (Schelling, 1971), and the diverse ones that persist tend to be confounded by endogenous sorting.

We use a large-scale policy experiment in Indonesia to shed new light on the long-run process of integrating people from diverse backgrounds. Between 1979 and 1988, the Transmigration program assigned two million voluntary migrants (hereafter, transmigrants) from the Inner Islands of Java and Bali to nearly 900 rural settlements across the diverse Outer Islands. Endowed with the same local institutions, the new settlements included a mix of Inner and Outer Islanders. Although the primary objective of the Transmigration program was to alleviate perceived over-population pressures in the Inner Islands, the government also envisioned resettlement as part of a broader nation building effort, to encourage ethnic integration between geographically isolated and culturally disparate Inner and Outer Island groups. This broad Inner–Outer divide is one of Indonesia's main intergroup cleavages, which runs deep historically and parallels the majority versus minority distinction in many countries.[3]

Indonesia's Transmigration program provides a unique opportunity to investigate how diversity shapes the process of integration and nation building. The unprecedented scale of the program created new communities spanning the archipelago and resulted in a continuum of policy-induced ethnic diversity. Transmigrants could not choose their destinations and received farm plots and new housing units assigned by lottery upon arrival in the new communities. Planners determined the local mix of Inner and Outer Island groups within settlements depending on land availability, which was a function of both the quality of arable land and the size of the indigenous population in nearby areas. Imperfect land markets also limited the scope for *ex post* sorting. The combination of initial policy assignment and limited outmigration allows us to identify how interethnic relationships and preferences changed, one to two decades after the initial demographic shock. Using comprehensive microdata on marriage, language, and intergenerational ethnic identity choices, we find substantial changes in social identity behavior consistent with deeper ethnic integration amidst rising diversity.

Our research design uses three identification strategies to understand whether and how ethnic inte-

[1] For example, the motto for the European Union is "United in Diversity," and "Unity in Diversity" is the motto for South Africa and Indonesia. Also, *E pluribus unum* (out of many, one) is a motto on The Great Seal of the United States.

[2] Recent estimates suggest there are around 230 million international migrants and 760 million internal migrants (Bell and Charles-Edwards, 2013). Based on recent trends, migration pressures are growing among minorities within rich countries (see Frey, 2006, on the United States) and in newer migration corridors from poor to rich countries (Hanson and McIntosh, 2016).

[3] Indonesia is home to more than 700 ethnolinguistic groups. Transmigrants came from all eight groups native to Java/Bali, but the majority hailed from Indonesia's dominant ethnic group, the Javanese, which comprise around 40 percent of the national population. The Outer Islands are considerably more diverse, home to the vast majority of Indonesia's unique ethnic groups, each with its own native language, most of which are mutually unintelligible. Most groups are small, but around 20 have more than 1 million self-identified members. See Section 2.1 for further background.

gration happens. First, we compare Transmigration villages (treated) to placebo villages (control) that were planned but never received the program to estimate *whether* resettlement improved interethnic relations between Inner and Outer Islanders. These almost treated villages—also in frontier rural areas—exist because of abrupt budget cutbacks following the sharp drop in global oil prices in the mid-1980s. At the time, control areas were also largely unpopulated, and their spontaneous settlement by migrants who endogenously sorted into these areas over the last two decades provides a credible counterfactual for what would have happened to Transmigration sites in the absence of the program. By comparing villages with similar endowments and by roughly holding the age of the settlement constant, we mitigate concerns that treated villages are negatively selected places relative to earlier settlements that have more favorable natural advantages and transportation access. Our estimation strategy combines features of the site selection process with an Oaxaca-Blinder type reweighting of control areas (akin to Busso et al., 2013; Kline and Moretti, 2014) to remove this potential downward bias.

Our second identification strategy uses the same treated versus control comparison but focuses on individual-level outcomes from the universe of 2000 Population Census data. We compare native Outer Islanders living in Transmigration areas who were exposed to program-induced Inner Islanders with natives who live in control areas and were exposed to endogenous immigrants. This allows us to estimate individual-level exposure effects caused by the program, both for the initial generation and also for children who grew up in these diverse communities.

Our final identification strategy exploits program-induced variation in diversity within treated areas to quantify *how* the degree of diversity in a community influences the incentive to integrate with other groups. We leverage knowledge of the planning process to identify plausibly exogenous variation across Transmigration villages in the share of Inner versus Outer Islanders. The initial placement of transmigrants across several hundred settlements and relatively limited mobility allow us to flexibly estimate the nonlinear relationship between different levels of diversity and identity choices with social externalities. We provide a battery of robustness checks to isolate policy-induced variation in ethnic diversity and use counterfactual exercises to address confounding due to *ex post* sorting.

Our findings suggest that Transmigration led to large and persistent demographic changes in the Outer Islands. While program villages are no more populous on average, they are more diverse and less residentially segregated than control areas. Ethnic fractionalization—the probability that any two village residents belong to different ethnicities—increases by nearly 50 percent. Most of that differential is due to the initial influx and natural rate of increase of transmigrants with the share of Inner Islands ethnicities increasing from 6 percent in control villages to nearly 60 percent in settlement villages. Moreover, Inner and Outer Islanders are more residentially integrated across census blocks than in control areas. The persistence of this local diversity for over a decade suggests limited sorting or tipping behavior that would have otherwise neutralized the initial shock.

In addition to stimulating local diversity, the Transmigration program strengthened interethnic integration through the marriage market. While the program did not change overall marriage rates, it did lead to changes in the composition of new marriages. Interethnic marriage rates for young cohorts (who married after the program) more than triples from a low control group mean of 2.3 percent. This average treatment effect on the treated (ATT) villages is large and equivalent to the difference in interethnic marriage rates between primary and junior secondary educated individuals in Indonesia's ethnically diverse

and cosmopolitan capital, Jakarta.

Individual-level ATT estimates indicate that Outer Island natives in Transmigration villages responded relatively more positively to diversity, relative to Outer Island natives in control villages. A difference-in-difference analysis indicates that the ATT appears to be driven by sharp increases in interethnic marriages among young natives in Transmigration villages relative to the older cohorts. By contrast, the trend is smooth across cohorts for control villages. Given that observed mobility and schooling patterns are not differential for old versus young cohorts, this mitigates concerns that our results are driven by differential *ex post* sorting in treated versus control areas. Additionally, the ATT effects do not appear to be driven by differences in public goods, complementary nation building activities such as the primary school expansion (Duflo, 2001), or other institutional changes potentially due to the program.[4]

Further results suggest that the increase in interethnic marriage is not only due to exposure to transmigrants (*supply*) but also plausibly to changes in preferences (*demand*). We find an inverted U shape with the ATT effect maximized in villages with close to equal mixes of Inner and Outer Islanders. This is not due to exposure effects alone. In fact, individuals need not develop intergroup relationships in places with 50/50 mixes as they tend to have enough in-group members. Other research suggests that oppositional identities and racist attitudes may develop in precisely these types of ethnically mixed communities (e.g., Bisin et al., 2011). We find that interethnic marriages increase above and beyond what would be expected if individuals randomly matched in the more ethnically diverse marriage markets. Reduced form results suggest that around one–third of the differential increase in interethnic marriage in settlement areas is due to a change in preferences for non-coethnic partners.[5]

These changes in marriage behavior point to shifts in identity that we further corroborate by identifying more intense diffusion of the national language in treated areas. We use auxiliary household survey data from 2006 on the primary daily language spoken at home. Since nearly everyone is able to speak *Bahasa Indonesia* (hereafter Indonesian), the program did not expand the supply of national language speakers.[6] Moreover, this outcome measures daily usage at home, consistent with parents using language as a vehicle to inculcate social values. Indonesian has its roots not in the dominant language of the largest ethnic group (Javanese) but rather in Malay, a historical lingua franca used across the archipelago. In diverse countries like Indonesia where each group speaks its own native language, a neutral national language can allow individuals from distinct groups to interact without having to sacrifice their native (ethno)linguistic identity.

Individual-level estimates reveal a significant increase in Indonesian use at home, particularly in the most diverse Transmigration villages. We find that individuals in treated areas are 25 percentage points more likely to report Indonesian as primary language at home relative to 12 percent in control areas. This large effect is robust to controlling for age, schooling, and other confounders and is similar in magnitude to the gap in language at home between rural and urban areas or between those with less than primary and those with junior secondary education. Moreover, like intermarriage, the ATT effect is largest in Transmigration villages where Inner and Outer Islanders are in roughly equal proportion. Meanwhile,

[4]Furthermore, beyond the observed residential and marital integration, we also find that residents of Transmigration areas have more favorable sentiments towards other ethnic groups in the village according to a survey of social attitudes in 2012.

[5]In ongoing work, we take a complementary, structural approach to quantifying how much of this increase is due to changes in supply (of non-coethnics) versus changes in demand (or preferences) for out-group partners (see Appendix B).

[6]Today, over 90 percent of the population is able to speak Indonesian, but only 20 percent report speaking it at home.

in villages with imbalanced group sizes, individuals are more likely to retain their own native language, or to adopt the majority language in the case of minorities. The nonlinear relationships we identify are consistent with social externalities and multiple equilibria in language use, which we rationalize by extending the Lazear (1999) two group model of language adoption to a setting with a national language.

Finally, we shed light on how diversity affects the intergenerational transmission of identity. In particular, we exploit the fact that children of interethnic marriages only report (or have reported by their parents) a single ethnicity. In settlements where transmigrants are a dominant majority (as induced by the initial assignment), children in mixed marriages are more likely to adopt the majority Inner Islands ethnicity. The opposite holds in settlements where transmigrants are a minority. Meanwhile, we find considerably more neutral transmission in villages with more balanced shares of Inner and Outer Islanders. These nonlinearities hold even when restricting to households where the father is of transmigrant identity, which is important given that most ethnic groups are patrilineal. Nevertheless, the fact that we do not observe tipping towards the majority identity once the 50/50 threshold is crossed suggests that ethnic identity is not very persistent even in these traditional, patrilineal rural cultures.

Overall, the deeper long-run interethnic integration in Transmigration villages is striking in light of prevailing concerns that such large-scale resettlement would stoke ethnic conflict. Early observers feared Inner Islanders, and the dominant Javanese in particular, would be viewed as outsiders competing for resources with native "sons of soil" (Fearnside, 1997; Otten, 1986). Only recently has the narrative of conflict been challenged by historical ethnographers (Barter and Côté, 2015), and indeed our findings are consistent with this reappraisal.

Looking across the continuum of diverse new communities, we offer several potential explanations as to why abrupt ethnic-based immigration shocks did not lead to prolonged social disintegration. First, minorities in these settings likely face low returns to establishing an oppositional identity. In particular, the newly salient Indonesian identity offers an alternative source of pro-social identification that does not imply succumbing to the majority culture. Additionally, the opportunity cost of opposition is high given that many minorities can reap significant monetary gains from interactions with the majority group based on their comparative advantage as intermediaries in the local economy. To be sure, as their own ethnic group share falls, minorities are significantly more likely to be working in trader-related occupations, with a clear turning point around the 50 percent threshold. Second, in communities with major groups in roughly equal proportion, the social and economic tradeoffs of intergroup interactions are less stark. Despite having large enough groups to sustain residential enclaves and marriage segregation, mixed communities exhibit perhaps the strongest interethnic integration and diffusion of the national identity.

**Contributions and Related Literature.** Together, our findings on marriage, language, and identity choices suggest that although many aspects of identity are resilient to change, some are produced through social interactions and hence amenable to policy. The Transmigration program provides a unique laboratory to deepen our understanding of how diversity affects identity choices and the formation of intergroup relationships. Our natural experiment contributes to several areas of the political economy literature discussed here.

First, we offer new insights into how ethnic diversity affects nation building. Although a large litera-

ture identifies adverse socioeconomic consequences of diversity and ethnic segregation in particular (see Alesina and LaFerrara, 2005; Alesina and Zhuravskaya, 2011; Esteban et al., 2012), little is known about how diverse countries might effectively deal with such divisions. Okunogbe (2015) exploits the random assignment of youth volunteers across regions of Nigeria to show that increased exposure to other ethnic groups boosts own group pride but also strengthens appreciation of national identity. Our results provide longer-run evidence on the process of integration across generations identifying marriage and language as important vehicles for intergenerational transmission of national identity. Other work finds that ethnic residential desegregation leads to mixed social outcomes in France, where increases in diversity prevent collective action around public goods provided by individuals who interact relatively frequently with outsiders (Algan et al., forthcoming). Like this study, we leverage plausibly exogenous variation in very localized diversity, allowing us to identify policy-relevant integration outcomes obscured at higher levels of geographic aggregation. Together, our results suggest that ethnic mixing can offer additional modes of identity formation and hence be an important vehicle for nation building by (i) crowding out more traditional forms of vertical socialization (from conservative older generations) (Bisin and Verdier, 2000), and (ii) complementing other nation building efforts such as mass education.[7]

Second, by studying newly formed communities with different initial levels of diversity, we are able to provide a unique lens into the otherwise slow process of breaking down interethnic cleavages and converging towards new forms of shared identity. Desmet et al. (2012, 2015b) draw a distinction between deep and shallow group ethnolinguistic cleavages. The former are rooted in immovable cultural preferences between very distant groups whereas shallow cleavages are more superficial and can be overcome if the economic returns to cooperation are high enough and there is a vehicle for coordinating interactions. Our results suggest that interethnic marriage and a national language can be such vehicles. The latter is related to recent work by Laitin and Ramachandran (2015) who argue that the choice of an appropriate national language—typically, a lingua franca rather than a dominant ethnic group's language—can undo the well-established negative correlation between national ethnolinguistic diversity and socioeconomic development.[8] Their cross-country regressions provide supportive correlational evidence. Our within-country identification strategy allows us not only to hold constant confounding national institutions but also to clarify how the local mix of ethnolinguistic groups shapes incentives for cross-group interactions and intergenerational transmission.

Finally, our findings contribute policy-based evidence to a small but growing literature examining identity and integration in diverse societies (Akerlof and Kranton, 2000; Bisin et al., 2008, 2016; Manning and Roy, 2010). Our findings on child identity in mixed marriages builds on recent work showing how

[7]We build upon recent work exploring other vehicles for nation building. Bandiera et al. (2015) argue that voters adopted compulsory schooling to integrate new immigrants in the melting pot of the United States during the age of mass migration. However, Lleras-Muney and Shertzer (2015) show that English-language schooling policies had no impact on social integration. Clots-Figueras and Masella (2013) show that exposure to Catalan in school increases identification Catalan causes in Spain. Dell and Querubin (2016) identify unintended consequences of United States military intervention and bombing in Vietnam on nation building outcomes. Meanwhile, Bazzi and Gudgeon (2016) show for Indonesia that well-designed administrative boundaries and inclusive local government institutions can mitigate conflict in ethnically polarized areas. And Glennerster et al. (2013) show in Sierra Leonne that strong local leaders can also serve to coordinate interethnic cooperation from above. Blouin and Mukand (206) similarly demonstrate how strong propaganda and sanctioning by the central government in post-genocide Rwanda led to a weakening of interethnic biases.

[8]Their July 2015 working paper highlights the choice of *Bahasa Indonesia* as the quintessential example of the optimal language in their model. Alesina and Reich (2015) develop a model of nation building in which the national language serves as a key technology for coordinating identity formation and cooperation. Our empirical results are consistent with their propositions.

identity formation depends upon the relative size of one's own group (Abramitzky et al., 2015; Algan et al., 2013; Carneiro et al., 2015; Jia and Persson, 2015). However, our focus on the social externalities associated with national language use is novel.[9] Language policies have been introduced in many settings to encourage diverse ethnic groups to rely less upon local languages and instead coordinate on the use of a single, national language for communication (see, e.g., Miguel, 2004, on Tanzania). The successful diffusion of the national language in Indonesia stands in sharp contrast to other large and diverse developing countries such as India and the Philippines (see Sneddon, 2003). Our findings suggest that greater internal migration and interethnic mixing in marriage and residence—both of which may be more constrained in countries like India—can hasten national language diffusion and contribute to nation building.[10]

The rest of the paper proceeds in six sections. Section 2 provides background on nation building efforts, the Transmigration program and ethnolinguistic diversity in Indonesia. Section 3 details our main data sources. Section 4 develops the empirical strategy for identifying causal impacts of resettlement and diversity. Section 5 presents the main impact estimates for village-level outcomes. Section 6 shows how the diffusion of the national language is an important vehicle for nation building amidst migration-induced changes in diversity, and Section 7 concludes with a discussion of policy implications and ongoing work.

## 2 Background: Diversity and the Transmigration Program

In this section, we first discuss ethnolinguistic diversity in Indonesia and the history of *Bahasa Indonesia*, the national language, and its intended role in nation building. We then describe the Transmigration program and its goal of fostering interethnic cooperation. Finally, we highlight various aspects of the program that are important for our identification strategy and analysis.

### 2.1 Ethnolinguistic Diversity and the National Language

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. According to the 2010 Population Census, Indonesia is home to more than 1,200 self-identified ethnic groups living on roughly 6,000 islands. By far, the Javanese are the largest ethnic group, constituting 40.1 percent of the population, followed by the Sundanese (15.5 percent) (Ananta et al., 2013). Both of these ethnic groups originate from the Inner Island of Java. Each of the next thirteen largest ethnic groups comprise between 1.2 and 3.7 percent of the country's population.

Despite this vast diversity, most Indonesians live in ethnically homogeneous communities. Using the universal 2000 Population Census, we construct the widely used index of ethnic fractionalization, $ELF$, for each of the country's more than 60,000 urban and rural villages (Easterly and Levine, 1997).[11]

[9] In related work, Advani and Reich (2015) identify diversity tipping points in English language adoption and interethnic marriage within U.S. counties during the period of mass immigration in the early 20th century. There are three key differences with our study: (i) our focus on policy, (ii) our ability to rule out first-order confounding due to endogenous sorting across locations, and (iii) the distinction between a national language and the majority group language.

[10] Of course, having a national language and interethnic integration are neither necessary nor sufficient for successful nation building. Rwanda and Somalia provide salient counterexamples. We discuss further policy implications in the conclusion.

[11] 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.

The median village in Indonesia has an $ELF$ of 0.05, which means that in half of Indonesia's villages, there is no more than a five percent chance that two individuals randomly chosen from the same village belong to different ethnicities. In the Outer Islands (i.e., outside of Java and Bali), the $ELF$ of the median village is 0.14, suggesting similarly homogeneous communities. By comparison, the $ELF$ calculated for the entire country is around 0.7.

Along with the deep ethnic diversity comes significant linguistic diversity. Of the nearly 700 languages currently spoken in Indonesia, almost all belong to the Austronesian language family, but many are very different from one another, and the differences are particularly large across the Inner–Outer divide.[12] This diversity presented policymakers, from the earliest stage of the country's independence struggle, with a problem of nation building.[13] Indonesia's founding leaders confronted this challenge by introducing a national language, *Bahasa Indonesia* or Indonesian—a modified version of Malay, which originated along the eastern coast of Sumatra (and peninsular Malaysia) but has been used as a trading language in the archipelago for centuries.

This choice of Malay as a national language exemplifies the importance of language in promoting national unity. This choice was made in the Second Indonesian Youth Congress on 28 October 1928, almost two decades before Indonesian independence. Prior to its recognition as the national language, only 5 percent of the population living under Dutch colonial authority spoke Malay as the native language, whereas nearly 40 percent of the population spoke Javanese. By (unanimously) choosing a minority language, the delegates of the 1928 Youth Congress avoided the resistance of non-Javanese ethnic groups and signaled their commitment to political unity.[14]

Article 36 of Indonesia's 1945 Constitution, ratified following its independence, cemented the status of Indonesian as the national language. Subsequent policies established Indonesian as the language for official communication in government, administration, and the army, and incorporated the language in the national curriculum (Nababan, 1991; Wright, 2016). Together with the expansion of access to education, this policy helped spread the adoption of the national language. Given the challenge arising from its vast diversity, Indonesia's national language policy is often considered an exemplary success.[15] Knowledge of Indonesian steadily increased from 50 percent in 1981 to 88 percent in 2010 (Wright, 2016).

In diverse multilingual settings like Indonesia, a national language not only helps facilitate social and economic interactions across groups but may also play an instrumental role in identity formation, which can have important implications for nation building. As one early observer argued, "the more [the Indonesian people] learned to express themselves in Indonesian, the more conscious they became of the ties which linked them" (Alisjahbana, 1962). Today, many view Indonesian as "a symbol of national unity identification" rather than simply an official language used in politics or business (Sneddon, 2003). However, despite its official status and widespread use in formal communication, *Bahasa Indonesia* has

[12] For example, the linguistic distance between Javanese and prominent Outer Islands languages in Transmigration areas (e.g., Minangkabau, Batak, Toraja) is akin to the linguistic distance between German and French (Lewis et al., 2009).

[13] The problem of nation building is particularly salient because prior to its independence, Indonesia consisted of separate empires with little shared history other than the experience of Dutch colonialism.

[14] As Sneddon (2003) points out, this choice avoided the type of opposition from minority ethnic groups that was found, for example, in India during attempts to establish Hindi as the national language.

[15] 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."

not been universally adopted. More than 80 percent of the population still reports using native languages as their primary daily medium of communication (based on the 2010 Census).

### 2.2 Transmigration and Nation Building

Indonesia's Transmigration program was designed to address the perceived problem of uneven population distribution. In the late 1970s, nearly two-thirds of the population lived in Java/Bali, which comprised seven percent of the country's land. The program subsidized the relocation of agricultural households from rural Java/Bali to newly created rural settlements in the Outer Islands. Historically, Transmigration began during the Dutch colonial period and was revived after independence under the Sukarno and Suharto administrations. However, it received a major overhaul in the third and fourth Five-Year Development periods (or *Pelita*) from 1979–1988 under the Suharto administration, and expanding national food production was a major goal during this period (see Bazzi et al., 2016).

Beyond agricultural development, planners hoped that the program would promote nation building by fostering interactions between the country's diverse ethnic groups (Hoey, 2003; Kebschull, 1986; MacAndrews, 1978; World Bank, 1988). In speeches and policy documents, government officials allude to the program's role in enhancing interethnic cooperation. For instance, in 1985, Martono, 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 ..." (Fearnside, 1997).[16]

However, some critics, especially those living in the Outer Islands, perceived this resettlement effort as part of a "Javanization" agenda. The relative influence of the Javanese—both in terms of popular culture and politics—was a salient issue in the post-independence period (see, e.g., Ananta et al., 2004; Hoshour, 1997; Mangunrai, 1977; Mulder, 1996; Thornton, 1972). At issue was whether the Transmigration program would be a vehicle for cultural imperialism over the Outer Island cultures (see e.g., Aspinall, 2008; Charras et al., eds, 1993; Levang, 1995). Their concerns reflect the unease among minority ethnic groups experiencing rapid immigration of majority, "sons of the soil" ethnic groups from the political and economic center of the country.

### 2.3 Transmigration Program Design and Implementation

Two key features of the design and implementation of the Transmigration program will be central to our empirical analysis. First, due to a large shock to oil prices in the mid-1980s, program funds were cut drastically, and a large number of villages that had been selected to receive transmigrants ultimately did not. Second, because of the scale and pace of implementation, individual transmigrants could not select their destinations, and coordination issues lead to an as-if-random placement of transmigrants to destinations, where they were encouraged to interact with local native groups. We describe each of

[16] Discussing the program objective, World Bank (1988, p. 5) noted: "[Transmigration] has been seen by national leaders as a tool for national integration ... The ideal of cultural pluralism is recognized in the constitution and in Indonesia's motto 'Unity in Diversity', and virtually all Indonesian ethnic groups preserve their mother tongue and many of their customs and traditions ... Migration of outer island residents, mainly to Javanese cities, and of inner island residents, mainly to rural areas in the outer islands, is seen by national leaders, from both Java and the other islands, as a means of promoting cultural contact and building national unity."

these features here.

**Budget Shock and Almost Treated Villages.** Encouraged by the apparent success of the program during the third development planning period (*Pelita* III, 1979-1983), when nearly 1.5 million people were resettled, the government set an ambitious target of relocating 750,000 households (3.75 million people) in its plan for *Pelita* IV (1984–88). In the first two years, the government resettled around 307,000 people, but in in January 1986, oil prices fell dramatically, and declining revenues forced the government to cut the Transmigration budget. The overall development budget was cut from Rp 10.7 trillion in FY 1985–86 to Rp 8.3 trillion in FY 1986–87, a reduction of 22 percent in nominal terms. The Transmigration program was hit especially hard as the budget was reduced by 44 percent (Rp 578 billion) in FY 85/86. Figure 1 illustrates the impact of oil prices on the resettlement program, showing how large fluctuations in the annual number of transmigrants placed coincided with large fluctuations in the world oil price.

In response, the Ministry of Transmigration (MOT) reduced its FY 1986–87 targets for settlement on sites already under preparation from 100,000 families to 36,000 fully sponsored families. In May 1986, with a continuing decline in oil prices, the MOT budget fell again to roughly Rp 208 billion ($\approx$ 58 percent of FY 84/85 expenditures). Under these circumstances, priority was given to maintaining and upgrading existing sites, to the completion of projects already underway. Crucially, all land clearing in areas slotted for final phase analysis was deferred indefinitely, except on a few externally assisted projects expected to settle no more than 3,000–5,000 families per year.

According to planning manuals, sites were selected to become Transmigration settlements using a three stage 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 indigenous settlements. The second phase involved aerial reconnaissance to identify "recommended development areas" (RDA) based on climate, hydrology, present land use, forest status, agricultural potential, and nearby indigenous populations. The final phase involved local surveys of topography, hydrography, and soil quality that helped determine the total number of transmigrants to be allocated to the settlement. In practice, given the scale of the program and associated logistical constraints, the planning process was much less in depth than intended and often involved on an ad hoc, "plan-as-you-proceed" approach (World Bank, 1988).

As detailed in Section 4.1, these planned but unfinished RDA villages provides a credible counterfactual for what Transmigration villages would have looked like today in the absence of the program. The villages covered by RDA maps were eventually spontaneously settled, and because they were largely unpopulated at the time of the program, these villages are also relatively new. Focusing on newly created villages allows us to shed light on how social interactions develop in a community's formative years, when multiple equilibria are possible and social norms are still developing. Moreover, as described below, we observe many of the features governing the selection process (natural endowments) and that determine agricultural suitability (and hence, local economic conditions).

**Selection of Transmigrants and Initial Placement.** Transmigrants volunteered for the program, but to participate, they 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).[17] Transmigrants received free transport to the new settlements, housing, and farm plots assigned by lottery.

As described at length in Bazzi et al. (2016), the process by which individual Transmigrants were assigned to destination settlements was neither rigorous nor systematic. Many reports indicate that the process suffered from time, information, and institutional constraints and led to a haphazard implementation. During *Pelita III*, when the program rapidly scaled up in resources, the Directorate General of Transmigration (DGT) was encouraged to delegate tasks to separate government agencies to speed up the settlement process, but coordinate problems between these agencies made it infeasible to systematically match transmigrants to settlements. The actual allocation of transmigrants to destinations seems to be largely driven by the coincidental timing of transmigrants' arrival to transit camps and the opening of new settlements in the Outer Islands (Hardjono, 1988). Importantly, participants had no choice over their destinations, and were often ill-prepared for the agricultural conditions in their new settlements (Levang, 1995; Kebschull, 1986).

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 from around the settlement areas. In 1979, this share—officially known as *Alokasi Pemukiman bagi Penduduk Daerah Transmigrasi* (APPDT)—was recommended to be 10 percent, and it was officially increased to 20–25 percent in 1982. In practice, these thresholds varied considerably across settlements (Clauss et al., 1988; Rigg, 2013), and planners determined the number of transmigrants to be sent to a given settlement based in large part on the size of the nearby Outer Islands population.

The combination of the haphazard assignment of Transmigrants to destination villages and variation in the APPDT shares resulted in plausibly exogenous variation in ethnic diversity across settlements, at least insofar as how they were initially established. Below, we use this variation to identify the effects of diversity, addressing important concerns about selection in the program and *ex post* sorting that could confound estimates.

## 3 Data: New Settlements and Nation Building Outcomes

We describe here the construction of key variables and underlying data sources.

### 3.1 Identifying Treated and Control Villages

The main source of data on Transmigration villages is the 1998 Transmigration Census, produced by the MOT. We digitize this Census to identify 911 Transmigration villages (outside of Papua) established

[17] As discussed in Bazzi et al. (2016), transmigrants are similar to the target beneficiaries of programs designed to resettle populations displaced by climate change (IPCC, 2014). In particular, these 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 not unlike 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.

from 1979 to 1988. These villages, which comprise our core treatment group, received on average 1,872 transmigrant individuals in the initial year of settlement. However, some villages received as few as 350 transmigrants while others received as many as 8,500.

We identify control villages using the MOT's high resolution maps of recommended development areas (RDAs) constructed after large-scale mapping of agricultural viability during the site selection process described in Section 2.3. These RDAs include groups of villages that were initially selected as candidate settlements but ultimately did not receive immigrants as part of the Transmigration program due to the abrupt budget cuts in the mid-1980s. We digitally trace these RDAs and define as "almost-treated" (controls) those 907 non-Transmigration villages that share any area with the RDAs. 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 as reported in Table 1.

Figure 2 depicts the locations of both Transmigration and control villages. Transmigration settlements are scattered throughout the Outer Islands. More than half of these sites are located on the island of Sumatra (566 out of 1,021), but many are also found on Kalimantan (223) and Sulawesi (161), with much smaller numbers in Maluku and Nusa Tenggara. Control sites are similarly distributed throughout the three major islands of Sumatra, Kalimantan and Sulawesi. The geographic variation is sufficiently rich to allow for a within-island econometric framework that accounts for the vast socioeconomic differences across these large island groups.

**Predetermined Site Selection Variables.** We draw on several geospatial data sources—detailed in the Online Appendix of Bazzi et al. (2016)—to measure the variables used to identify RDAs and ultimately select settlements in the Outer Islands. These variables include: (i) measures of topography (land area, elevation, slope, ruggedness, and altitude), (ii) proxies for pre-program market access (distance to (sub)district capitals, roads, rivers, and the sea coast), and (iii) measures of soil quality such as texture, drainage, sodicity, acidity, and carbon content. We select these variables based explicitly on those listed in program manuals from 1978 in the MOT archives. By measuring these variables at an extremely high spatial resolution, we are able to account for important differences in natural advantages between treated, control, and other villages in the Outer Islands.

### 3.2 Integration and Assimilation Outcomes

We draw firstly on the universal coverage 2000 Population Census microdata to capture not only the demographic shock induced by the Transmigration program but also its impact on marital, residential, and identity choices. Key measures include total population, share of Inner Islands migrants, ethnic (and religious) diversity, residential segregation between Inner and Outer Islanders (based on the Bell, 1954, isolation index applied to census blocks within villages), and the reported ethnic identity of children within mixed marriages. Note that the Census only allows individuals to report a single ethnicity, and the household head typically reports that ethnicity for all members (Ananta et al., 2013).

Using the same dataset, we examine program impacts on interethnic marriage rates as a proxy for local-level integration. Despite its diversity, interethnic marriage is relatively uncommon with around 10 percent of individuals marrying outside their group and an even lower rate in rural areas. Interethnic

marriage has long been viewed by social scientists as key marker of assimilation (see Gordon, 1964), and as Babcock (1986) notes, officials in the Ministry of Transmigration often monitored interethnic marriage between Inner and Outer Islanders as a barometer for successful integration. For each household head, we can identity their marital status as well as the ethnicity of their co-resident spouse if married.[18] Moreover, because we observe *all* individuals in each village, we can benchmark the observed interethnic marriage rates to those that would obtain under random matching. This helps address the fact that more diverse places tend to have more interethnic marriages merely due to greater opportunity for marrying outside one's group rather than changes in preferences for exogamy.

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 focus on young couples. For treated villages, a young couple is one in which the ages of both spouses are less than 15 (the legal age of marriage) in the village's year of settlement. We use a similar approach to identify young couples in control villages, but 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 was 24. Our results are robust to alternative definitions of young cohorts, and we also report estimates of program effects on interethnic marriage across the age distribution.

Beyond marriage decisions, we also investigate language use at home as a leading indicator of socialization and identity choices consistent with nation building. Although language information is not available in the 2000 Population Census, we utilize auxiliary, individual-level data from the National Socioeconomic Survey (also known as *Susenas*) to examine language use.[19] *Susenas* is a nationally representative household survey that collects detailed data on Indonesian households. We use the socioeconomic module of the 2006 survey, which contains questions on ethnic identity and on the language used daily at home by the household head. For households that employ multiple languages at home, the enumerators were encouraged to ask for the language used most often (Central Bureau of Statistics, 2006, p.8).

Finally, as a validation check on our integration outcomes, we draw upon *Susenas* data from 2012 to capture two measures of interethnic preferences. Household heads are asked "How do you feel about other ethnic groups in your community?" with responses on a four point scale ranging from "not happy" to "very happy". They are also asked to respond to the statement "The district mayor [elected by popular vote every 5 years] must be a native of this area." with responses on a four scale ranging from "do not agree" to "strongly agree."

Together, these measures capture a wide array of social identity decisions and preferences reflective of the extent of interethnic integration. Using these measures as outcomes, our empirical strategy described next allows us to paint a rich picture of the otherwise slow process of breaking down interethnic cleavages and converging towards new forms of shared national identity.

[18] For extended households with multiple married couples, we are not able to identify secondary couples in the household as the data only provide information on relationship to the household head. Nevertheless, 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 approach looking within household and using information on age, gender, and relationship to the household head to identify some subset of these other marriages. Doing so leaves our results unchanged.

[19] Although reported in the 2010 Census, language data are not available with village-level administrative codes as required for our analysis. Nevertheless, despite the relatively small sample of households within each village, this survey data provides enough variation to identify key patterns of language use even as we face limitations in statistical power for some purposes.

# 4 Empirical Strategy

This section develops our empirical strategy in two steps. First, we propose a place-based evaluation design to identify the average impact of the Transmigration program on nation building outcomes in the Outer Islands. Our strategy has parallels in the urban (Busso et al., 2013) and trade (Donaldson, forthcoming) literatures, and the discussion below builds upon Bazzi et al. (2016). Second, we present a method for isolating plausibly exogenous variation in ethnic diversity across the new settlement areas based on the initial planning assessment of carrying capacity.

## 4.1 Place-Based Evaluation

We develop a place-based evaluation strategy that exploits the set of planned villages that never received transmigrants as counterfactuals. Many of these recommended development areas had been partially settled by local, non-transmigrant populations prior to the program and continued to be settled by spontaneous migrants thereafter. Indeed, the relevant long-run counterfactual is not an empty but habitable rural area. Rather, the comparison we aim to make is between (i) villages where the Transmigration program induced co-location of geographically disparate groups and (ii) villages where longstanding settlement patterns and spontaneous migration would typically lead to shorter distance moves and greater endogenous sorting into ethnic enclaves.

We use the counterfactual settlements to estimate the average treatment effect (on the treated, ATT) of the Transmigration program on nation building outcomes. Our main estimating equation is given by:

$$y_j = \alpha + \theta \text{Transmigration}_j + \mathbf{x}_j'\boldsymbol{\beta} + \nu_j, \tag{1}$$

where $y_j$ is some outcome measure for village $j$, $\text{Transmigration}_j$ is an indicator equal to one for Transmigration villages and zero for RDA villages, and $\mathbf{x}_j$ is the vector of predetermined control variables capturing natural advantages and used by planners during the process of site selection. All regressions include island fixed effects and cluster standard errors at the district level.

A causal interpretation of $\theta$ as the ATT requires that omitted place variables be uncorrelated with site selection. Spatial policies like the Transmigration program often target underdeveloped or distressed areas. This could lead to downward bias in $\hat{\theta}$ if resource-poor locations are less conducive to fostering interethnic integration. On the other hand, planners may select places that are particularly well suited to hosting diverse groups. If such sites were settled prior to the budgetary cutbacks, this could lead to upward bias in $\hat{\theta}$. We rule out these first-order concerns with program placement bias first by restricting the sample to treated and almost-treated villages and second by using a reweighting procedure akin to recent evaluations of place-based policies (Kline and Moretti, 2014).

Our preferred estimator is based on the Oaxaca-Blinder reweighting approach of Kline (2011), which allows the treatment effect to vary arbitrarily with the covariates in $\mathbf{x}$. This flexible specification also admits a dual interpretation as a propensity score reweighting estimator in which the inverse probability weights (IPW) are based on a logit formulation of the treatment likelihood. We use these IPWs below to show how reweighting rebalances predetermined covariates across treatment and control villages.

Table 2 demonstrates the building blocks of our identification strategy. We first show why it is impor-

tant to restrict the counterfactual to almost-treated villages. Transmigration villages exhibit significant differences from other villages in the Outer Islands along a range of agroclimatic and geographic characteristics. Consistent with the program targeting underdeveloped rural areas, Transmigration villages statistically differ in terms of nearly all $\mathbf{x}$ characteristics, most of which were used to select sites in the planning phase. When restricting non-program villages to the almost-treated RDAs, these differences become smaller. Nevertheless, there remain significant differences between treated and almost-treated villages in terms of some important soil, topographic, and climatic characteristics, which is not surprising given that planners may have selected more suitable settlement areas earlier in the planning process.

To account for these observed differences, we rebalance treated and control villages by reweighting the latter according to their estimated odds of selection. First, we predict the probability of being a Transmigration village:

$$\mathbb{P}(\text{Transmigration}_j = 1) = \Lambda(\mathbf{x}_j'\widehat{\boldsymbol{\zeta}}), \tag{2}$$

where $\Lambda(\cdot)$ is a logit function. Consistent with the bivariate comparisons of individual elements of $\mathbf{x}$ in Table 2, the conditional estimates $\widehat{\boldsymbol{\zeta}}$ reported in Appendix Table A.1 are indicative of sequential site selection among eligible settlement areas. For example, treated villages (i) are smaller in area, (ii) are at lower altitude, (iii) have better soil drainage, and (iv) are closer to major roads. The covariates explain over one-third of the variation in site selection; with province fixed effects, we can explain over half. Moreover, as seen in Appendix Figure A.1, the estimated propensity scores, $\widehat{P}_j$, exhibit strong overlap, suggesting considerable scope for constructing a credible counterfactual. Using the estimates of $\boldsymbol{\zeta}$, we reweight control village $j$ according to its estimated odds of having been a Transmigration site, $\widehat{\kappa} = \widehat{P}_j/(1-\widehat{P}_j)$, where $\widehat{P}_j$ is the predicted probability from equation (2).

This reweighting of RDA villages 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. In comparing the final two columns in Table 2, we find that without the $\widehat{\kappa}$ weights, more than half of the site selection variables exhibit large and statistically significant differences across treated and control villages. With the weights, that share falls to less than 10 percent. By removing observable site selection differentials, we get closer to a causal interpretation of the ATT.

### 4.2 Carrying Capacity and Demographic Composition within Settlements

In addition to the average impact of the program, we aim to identify heterogeneous effects of diversity within Transmigration areas. Focusing on the main cleavage between Inner and Outer Islanders, we exploit the fact that planners determined the number of Inner Island transmigrants to allocate to each settlement based on an assessment of carrying capacity. This capacity was determined by those same site selection characteristics in $\mathbf{x}$ and the size of the native Outer Islands population in nearby areas. Sites with greater agricultural potential—as proxied by the quality *and* quantity of available land—received more Inner Islanders in the initial year of settlement. In practice, conditional on natural advantages $\mathbf{x}$, a large (small) initial transmigrant population implied a small (large) initial native population (i.e., APPBT locals noted in Section 2.3).

We leverage this initial variation in settlement composition to investigate how diversity affects language, identity, and other individual choices with social spillovers and hence implications for nation

building. In particular, we are interested in understanding how choices by individual $i$ in Transmigration village $j$ are shaped by some possibly nonlinear function $g(\cdot)$ of diversity:

$$y_{ij} = \alpha + g(diversity_j) + \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_{ij}, \tag{3}$$

where $diversity_j$ captures the population share of Inner Island ethnic groups in 2000.[20] Of course, the diversity levels observed 10–20 after resettlement may reflect endogenous *ex post* sorting of individuals based on their tastes for diversity.

We therefore propose to instrument for diversity using the number of Inner Island transmigrants placed in the initial year of settlement. This isolates that portion of *ex post* diversity due to the permanent settlement, migration, and fertility induced by the early influx of settlers from the Inner Islands. Beyond the simple linear instrumental variables approach, we can allow $g(\cdot)$ to take any shape using the semiparametric instrumental variables control function approach developed in Su and Ullah (2008).[21] In both linear and semiparametric specifications, the initial transmigrants placed in the settlement is a strong predictor of Inner Island ethnic share in 2000 (see Figure 3).

Beyond instrument relevance, the key identifying assumption is that conditional on $\mathbf{x}$, the number of initial transmigrants only affects nation building outcomes through its effect on diversity. We provide three pieces of evidence supporting this exclusion restriction. First, we show that the instrument does not predict other measures of diversity—overall ethnic fractionalization, fractionalization among Inner Island ethnicities, birthplace diversity among the Java/Bali-born population, religious fractionalization—or skill levels among the initial transmigrants (see Appendix Figure A.2). 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 Table A.2).[22] Third, the number of initial transmigrants is not correlated with the similarity between languages native to Java/Bali and the indigenous language in the given area of the Outer Islands (see Appendix Figure A.2). Although these results are consistent with the instrument being plausibly exogenous (conditional on $\mathbf{x}$), we nevertheless acknowledge competing channels of influence and aim to rule out remaining concerns with additional robustness checks.

## 5 Average Impacts: Resettlement and Nation Building

We report here the main village-level impacts of the Transmigration program. First, we document the lasting demographic changes brought about by the program. Second, we identify a substantial increase in interethnic marriage between Inner and Outer Islanders over and above the mechanical increase associated with greater ethnic diversity. Finally, we identify intergenerational identity transmission and residential segregation patterns consistent with deepening interethnic ties. Overall, the substantial in-

[20] Given the early 1980s policy change in the recommended allocation of spots for local natives within the new settlements (see Section 2.3), it is important to note that all results based on this specification are robust to the inclusion of year of settlement fixed effects.

[21] This approach augments equation (3) with the residual from a (semiparametric) first stage regression of $diversity_j$ on $\mathbf{x}$ and the instrument. See Henderson and Parmeter (pp. 271–274, 2015) for an accessible introduction.

[22] 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.

crease in population diversity did not lead to the long-run social upheaval feared by those skeptical of resettling the dominant ethnic groups of Java/Bali in the minority group homelands of the Outer Islands.

In the tables discussed below, we report several estimates of $\theta$ for each of the main nation building outcomes. Column 1 compares Transmigration villages to all other Outer Island villages while columns 2 and 3 restrict the sample to the set of treated and control villages. Column 2 controls for the predetermined site selection characteristics in $\mathbf{x}_j$. Column 3 additionally implements the Kline (2011) reweighting estimator on which we focus our discussion.

### 5.1 Demographic Shocks

Table 3 demonstrates the large increase in the scope for interactions between otherwise isolated ethnic groups. The first important takeaway is that the program did not lead to a significant increase in total population. Hence, in the absence of the program, settlement areas would have been populated through the same forces of spontaneous settlement observed in control areas with similar natural advantages. A large fraction of this population change is due to the initial influx of transmigrants from the Inner Islands with treated villages having 34 percentage points (p.p.) more Java/Bali-born immigrants relative to a control group mean of 2 percent. Most of these immigrants and their descendants identify with ethnicities native to Java/Bali (hereafter, transmigrant or Inner Island ethnicities). As a result, treated villages have a 54 p.p. higher transmigrant ethnic share than control villages, where 6 percent of individuals claim transmigrant ethnicity on average.[23]

This mass immigration of the Java/Bali-born population is associated with a substantial increase in ethnic diversity. Ethnic fractionalization increases by 0.13 relative to a control group mean of 0.23, and these changes in diversity are not due to a mechanical effect of increased population as noted above. Nor are they due entirely to an increase in the number of ethnic groups, which is similar in both treated and control villages with around 20 unique groups on average. There are slightly more groups in treated villages, but these differences are insignificant after reweighting.[24] Hence, much of the causal increase in diversity can be attributed to changes in the size and composition of ethnic groups and in particular those hailing from Java/Bali.

Despite these large changes in ethnic diversity, we find smaller changes in religious diversity, which also may be salient in certain parts of the country.[25] Although the program led to a significant 9 p.p. increase in the Muslim population share (relative to a control group mean of 59 percent), there is no differential increase in religious fractionalization in treated villages. This suggests that most of the variation in religious affiliation lies across rather than within villages.

Overall, the results in Table 3 suggest that the Transmigration program led to a substantial increase in the scope for local interactions with individuals outside one's ethnic group. In the remainder of this

[23] Most of the gap between the Java/Bali-born and transmigrant ethnic share in Transmigration villages can be explained by the second generation born to the Java/Bali natives who migrated as children with their transmigrant parents.

[24] Note that we find roughly five additional groups on average in the simple unadjusted comparison of treated and all non-treated villages in column 1. This corresponds to the five major ethnic groups from Java/Bali that participated in the program. These differences disappear after more appropriately defining the counterfactual as reweighted, RDA villages.

[25] According to the 2000 Population Census, 88.2 percent of the population were Muslims, followed by Protestants (5.9 percent), Catholics (3.1 percent), Hindus (1.8 percent) and Buddhists (0.8 percent). As in the case of ethnicity, people tend to live in relatively religiously homogeneous communities. The religious fractionalization index for a median village in the Outer Islands is 0.01 compared to 0.14 for ethnic fractionalization.

section, we identify significant changes in marriage, identity and residential choices concurrent with this increased opportunity for cross-group interactions.

### 5.2 Interethnic Integration: Marriage, Identity, and Residence

We show here that the Transmigration program increased interethnic integration. The greater mixing in marriage and housing markets runs counter to popular concerns that this sort of program would create isolated ethnic enclaves. The lack of a dominant intergenerational transmission of identity in mixed marriages goes against the worry that the program would be a vehicle for ethnic colonization. We document these findings in several steps in Table 4.

First, we find a significant increase in the overall interethnic marriage rate. The ATT in row 2 of Panel A implies a 3.2 p.p. increase in interethnic marriages relative to a counterfactual mean of 6.9 percent, which translates into 11 additional interethnic marriages relative to a mean of 24. Overall, the program brought a statistically significant increase in the marriage rate among household heads by around 5 p.p. relative to the control mean of 83 percent. However, much of this increase is likely due to the eligibility rules that required initial transmigrant households be led by a married couple. To understand the implications of the increase in interethnic marriage, it is important to distinguish pre-program marriages (among Java/Bali-born transmigrant couples) from those that could plausibly be a result of the increased diversity in the marriage market after resettlement.

Therefore, we distinguish interethnic marriages between Inner and Outer Islanders as well as among the younger generation (as defined in Section 3). The likelihood of marriages between Inner and Outer Island ethnic groups more than doubles in Transmigration villages. The ATT for these interethnic marriages is roughly similar to the ATT for all interethnic marriage, suggesting that most of the changes can be accounted for by the increased mixing between Inner and Outer Island ethnic groups.

Panel B shows that the increase in interethnic marriage is even larger for the younger generation of household heads for whom marriage most likely took place after the resettlement period. The ATT estimates imply a doubling of any interethnic marriages relative to a mean of 8.1 percent (roughly seven marriages) in control villages. For interethnic marriages between Inner and Outer Islanders, the ATT of 5 p.p. relative to 2.3 percent translates into four additional interethnic marriages among young couples in the average Transmigration village. Moreover, these changes in interethnic marriage patterns among the younger adult population are not due to changes in marriage rates, which do not significantly differ in treated relative to control villages. Together, these results point to potentially significant changes in matching behavior within the marriage market.

Going a step further, we aim to quantify how much of the observed changes in marriage patterns among the younger generation are due to changes in supply (of non-coethnics) versus changes in demand (preferences for non-coethnics). Understanding whether preferences have changed can have far-reaching implications for nation building. Indeed, as noted in Section 3.2, Transmigration planners viewed interethnic marriages between Inner and Outer Islanders as a key marker of program success vis-à-vis the goal of national integration.

In the bottom of panel B, we show that the changes in interethnic marriage between the younger generation of Inner and Outer Islanders are not entirely due to changes in the mere opportunity of marrying outside one's own group. We do this by dividing the actual interethnic marriage rate by the

average interethnic marriage rate from 10,000 simulations of random matching (among the young, married population). With this measure, we can ask how much more interethnic marriage we observe in Transmigration relative to control villages above and beyond what would be expected solely from the policy-induced shock to the supply of potential non-coethnic partners.[26] In the typical control village, we find that the actual interethnic marriage rate is only 25.3 percent of the (random) potential interethnic marriage rate. However, the ATT implies that this difference increases by 17.7 p.p. for Transmigration villages. Although smaller than the ATT for actual interethnic marriage rates, this estimate implies that around one–third of the observed increase in interethnic marriages among the younger generation can be explained by a change in demand for exogamy. In ongoing work, we complement this reduced form approach with additional evidence of changes in preferences by estimating a structural model of matching in the marriage market (see Appendix B for details).

Beyond interethnic marriage, we highlight two additional results consistent with the Transmigration program leading to greater interethnic integration. First, in Panel C, we consider the intergenerational transmission of ethnic identity in interethnic marriages. As noted in Section 2, many skeptics were concerned that the Transmigration program would lead to an ethnic colonization of the Outer Islands by the majority Inner Islanders. This might manifest through interethnic marriages if the children were more likely to be identified with and inculcated in the ethnicity of the Inner Island parent. Restricting to treated and control villages where we observe any marriages between Inner and Outer Island ethnic groups, we find no significant difference in the share of children within interethnic marriages reporting the Inner Island ethnic identity. In both settings, that share is close to 50 percent, suggesting that Inner Island cultural imperialism has not occurred through the marriage market.

Second, Panel D shows that Transmigration villages exhibit significantly less residential segregation between Inner and Outer Islanders. The program led to a nearly 50 percent reduction in the residential segregation index relative to a control group mean of 0.34. Some of this reduction can be explained by the increase in interethnic marriage between groups. However, the large effect size suggests that program villages served to integrate Inner and Outer Islanders more effectively than the spontaneous residential sorting patterns observed in counterfactual villages.

Overall, the results in Table 4 point to the positive impacts of the Transmigration program on nation building.[27] Ethnic identity remains salient in much of Indonesia today, but the increased mixing across groups and the lack of a dominant ethnic transmission to children suggests that the program may have been a vehicle for forging a stronger national identity among otherwise distinct and isolated groups. In the following section, we show how the extent of ethnic diversity shaped social interactions within Transmigration villages, focusing in particular on language adoption but also examining identity and occupational choices. Before doing so, we briefly discuss several robustness checks.

[26] Note that this is only defined for the control villages in which we observe Inner Island ethnicities.

[27] Looking beyond the household, we provide further supportive evidence in Appendix Table A.3, which shows that the program led to a significant increase in the number of active community organizations and did not lead to any changes in the incidence of violent conflict at the village level as observed in the early 2000s. Although violence may have occurred earlier in the resettlement process, the lack of a significant increase in violence for the average village stands in contrast to the popular media vilification of Transmigration in the early years of democratization covered by our data. This popular (mis)perception was spurred in part by a few instances of extreme violence between Inner and Outer Island ethnic groups in Transmigration areas (see, e.g., Schiller and Garang, 2002).

### 5.3 Geographic Spillovers and Robustness Checks

Our main place-based results in Tables 3 and 4 point to the substantial increase in interethnic integration due to the Transmigration program. In Appendix Figures A.3 and A.4, we show that these impacts of the Transmigration program spill over to neighboring villages.[28] In particular, we estimate a version of equation (1) 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 Island immigrants and their descendants up to 30 kilometers beyond the Transmigration village boundaries. However, most of the results on interethnic integration in Table 4 are more local to the Transmigration villages. For example, as we move beyond 10 kilometers, the increase in interethnic marriages and reduction in ethnic residential segregation go to zero, and children are more likely to inherit Inner Island ethnic identities in mixed marriages.

These spillover effects can be explained by several forces including, among others, (i) the expansion of the original settlement area as the younger generation of transmigrants started new families and (ii) the arrival of additional migrants drawn by the prospect of economic opportunity around these new developments. Isolating the precise causes of the spillovers is beyond the scope of this study. Nevertheless, these findings suggest that resettlement programs can have significant social impacts beyond the population within the immediate settlement areas. These multiplier effects are important from a policy perspective.

The main ATT results in Tables 3 and 4 also survive important robustness checks: (i) including province fixed effects, (ii) controlling for predetermined natural advantages and population characteristics not used by the planners to select sites but potentially correlated with long-run outcomes (i.e., all variables used in Appendix Table A.2). These results are detailed in Appendix Tables A.4–A.7.

## 6 Mechanisms: Diversity and Identity Choices with Social Externalities

Despite the mass immigration of majority Inner Island ethnic groups, we find little evidence of long-term social upheaval in the Outer Islands as a result of the Transmigration program. Rather, the evidence thus far suggests that the resettlement process facilitated interethnic integration and may have contributed to nation building. We argue here that the national language, *Bahasa Indonesia*, played an important role in this process by helping individuals to mediate the social and economic tradeoffs associated with diversity. In diverse settings like Indonesia where each group speaks its own native language, the national language can serve as a neutral third language through which two individuals from distinct groups can interact without having to pay the social or cultural costs associated with adopting the other group's language at the expense of one's own.

We present several results consistent with this interpretation. First, the Transmigration program hastened the diffusion of the national language in the Outer Islands. Second, the diffusion is greatest in

[28] It is also important to note that all of the regression results in Tables 3 and 4 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.

the most diverse villages where Inner and Outer Islanders are in roughly equal proportion. We explain these results by adapting the Lazear (1999) multiple equilibria model of language adoption to a setting with a national language. Finally, we identify diversity thresholds in occupational and ethnic identity choices, both of which, like language choices, have social externalities.

### 6.1 Program Impacts on Language Use at Home

We begin by showing that the Transmigration program increased use of the national language. We estimate an individual-level version of the main ATT equation (1) for those residing within 10 kilometers of treated or control village boundaries based on the 2006 *Susenas* sample of villages. We expand the treatment and control areas beyond the Transmigration and RDA village boundaries in order to increase the number of village observations given the limited survey sampling frame.[29] The spatial spillovers identified in Section 5.3 further support this expanded definition of treatment for the purposes of studying language use. We again focus on the reweighting estimator and cluster at the district level.

Table 5 reports the impact of the Transmigration program on language adoption. Each cell is a separate ATT estimate where the dependent variable is an indicator for whether the individual's primary daily language at home is in column (1) *Bahasa Indonesia*, (2) native to Inner Islands ethnicities, or (3) native to Outer Islands ethnicities. The baseline specification in row 1 shows that individuals in Transmigration areas are 25 p.p. more likely to report Indonesian as their primary daily language relative to the mean of 12 percent in control areas. Columns 2 and 3 show that this large effect is driven primarily by reduced use of Outer Islands languages. Moreover, given how few Inner Islanders reside in control areas (see Section 5.1), most of these overall changes in language use are due to differential rates of Indonesian use among Outer Islanders living in Transmigration areas relative to their co-ethnics living in control areas.[30]

These baseline results are robust to several individual- and village-level confounds. First, controlling for basic demographics (quadratic age and gender) and dummies for years of schooling (row 2) leaves the ATT estimates unchanged as does restricting to those who have not completed primary school (row 3). These specifications rule out concerns about cohort composition effects and especially 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 occupation dummies that cover all sectors of work in rural Indonesia (row 4). The estimates are also robust to controlling for household expenditures per capita (row 5). Although perhaps jointly determined with language decisions (and due to the Transmigration program), these "bad controls" help shed light on the potential channels through which the program hastened the diffusion of the national language. Additionally, the effects cannot be explained by differential ease of speaking Indonesian as seen in row 6, which includes a dummy for whether the village's native Outer Islands ethnolinguistic homeland is Malay (the root of *Bahasa Indonesia*), and in row 7, which controls for the linguistic distance between the native homeland's

[29] 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 and RDA sites are in underdeveloped areas, they tend to be undersampled relative to a simple random draw of villages.

[30] Restricting the ATT estimates to Outer Islanders, we find that those living near Transmigration villages are 27 p.p. less (more) likely to report native Outer Islands languages (Indonesian) relative to those living near control villages, 87 (10) percent of whom report Outer Islands languages (Indonesian) as their primary daily one.

language and Malay based on a distance metric proposed in Fearon (2003).[31] Moreover, note that in all specifications, we control for pre-program market access in terms of proximity to (sub)district capitals and the road network (see Section 3.1), which are positively correlated with national language use.

The robustness of the results thus far in Table 5 suggests that some of the increased use of Indonesian could be explained by the Transmigration program's impact on ethnic diversity. In two "bad control" specifications in rows 8 and 9, we provide suggestive evidence supporting this interpretation. Controlling for overall ethnic fractionalization (row 8) leaves the results unchanged. However, controlling for a quadratic in the Inner Island ethnic share effectively reduces the ATT to zero. Together, these patterns motivate our next set of results showing how language adoption decisions are shaped by relative group sizes and one's relationship to the majority group.

### 6.2 Ethnic Diversity and Language Use at Home

The diffusion of the national language is determined in part by the necessity of and scope for interaction with individuals outside one's own ethnolinguistic group. In this section, we exploit the variation in ethnic diversity within treated areas to identify how group composition affects language diffusion.

In order to capture potentially nonlinear relationships between local diversity and individual language use, we estimate a semiparametric, partially linear version of equation (3) using the approach due to Robinson (1988). As in Table 5, we include all individuals residing in villages less than 10 km from a Transmigration village. We use a local linear (probability) estimator with the Fan and Gijbels (1996) rule-of-thumb (optimal) bandwidth and an Epanechnikov kernel, but results are similar using other approaches.

The first graph in panel (a) of Figure 4 reveals a significant inverted U relationship between the daily use of *Bahasa Indonesia* and the share of Inner Island ethnicities. The fraction of the population speaking the national language increases in the proportion of Inner Island natives residing in the village. However, this positive relationship turns negative at the point where transmigrants comprise roughly 40 percent of the total population, and the parametric test due to Lind and Mehlum (2010) suggests that the inverted U shape is statistically significant at the 1 percent level. In our full sample (not restricted to villages in *Susenas* survey data), nearly half of the villages within 10 km of Transmigration villages have more than 40 percent Inner Island ethnic share.[32] The inverted U is even more stark when expanding the settlement catchment area to 30 km at which point we still find a significant Inner Island ethnic population due to the program (see Section 5.3). Regardless of the catchment area, the turning point is close to the 50 percent threshold where diversity is maximized based on this two group measure.

We observe the lowest levels of daily Indonesian use in those areas where either Inner or Outer Islanders are the dominant majority. The remaining graphs in Panel (a) of Figure 4 show that in these villages, the majority group language is much more likely to be the primary daily language than In-

[31] In particular, we compute the distance as $1-(shared_{ij}/5)^{0.5}$ where $shared_{ij}$ is the fraction of shared branches on the linguistic classification tree provided in *Ethnologue* (Lewis et al., 2009). See the Online Appendix of Bazzi et al. (2016) for further details on the mapping between ethnicities reported in the 2000 Population Census, the languages listed in *Ethnologue*, and the ethnolinguistic homeland shapefiles provided by the *World Language Mapping System*.

[32] Appendix Figure A.5 shows that this positive relationship between diversity and national language use generalizes to the broader measure of overall ethnic fractionalization. In this case, individuals in the most fractionalized villages are the most likely to employ *Bahasa Indonesia* on a daily basis with the sharpest increase occurring in villages where there is more than a 50 percent chance of encountering a non-coethnic neighbor.

donesian. There is a strong positive relationship between the Inner Island ethnic share and the share of the population speaking Inner Island languages and a strong negative relationship for Outer island languages. These sharp, nearly linear relationships explain the inverted U relationship for national language use, which exhibits a clear turning point in terms of the mix of Inner and Outer Island ethnic groups within settlement areas.

In panel (b) of Figure 4, we show that these semiparametric results are robust to an instrumental variables specification that accounts for potential endogeneity in diversity. In particular, we instrument the Inner Island ethnic share in village $j$ with the number of transmigrants placed from 1979–88 within 10 km of $j$. As discussed in Section 4.2, we adopt a control function approach developed in Su and Ullah (2008) and estimate local polynomial first and second stage regressions with the latter conditioning on residuals estimated (semiparametrically) in the first stage.[33] Together, these preliminary IV results look qualitatively very similar to the baseline estimates. The one statistically insignificant exception is that that the lower part of the inverted U for *Bahasa Indonesia* becomes flatter.

Overall, the similarity in OLS and IV estimates points to the persistent impact of the initial wave of settlers in these newly formed communities on diversity levels observed two to three decades later. Although the exclusion restriction is not free from scrutiny (see Section 4.2), the unique institutional features of the Transmigration program afford us the rare opportunity to isolate a part of modern diversity that is not due to the slow-moving processes of endogenous sorting along ethnic lines.

The patterns in Figure 4 suggest that the individual returns to adoption of the national language are shaped by the extent of ethnolinguistic diversity. In Figure 5, we show how these returns depend on one's relationship to the majority group. Individuals belonging to Inner Island ethnic groups are significantly more likely to adopt native Outer Island languages or *Bahasa Indonesia* when they are in the minority. The same holds for individuals from Outer Island ethnic groups who are considerably more likely to adopt native Inner Islands languages or *Bahasa Indonesia* when they are in the minority. The IV results in Appendix Figure A.6 look similar.

We complement these semiparametric results by comparing the nonlinear effects of diversity within Transmigration areas to language use within control areas. Specifically, in Table 6, we estimate heterogeneous ATT specifications allowing for different treatment effects across terciles of the Inner Island ethnic share within Transmigration villages.[34] Although subject to the "bad control" critique, this provides a useful benchmark as it compares language choices for individuals in Transmigration areas at different levels of diversity to their co-ethnics residing in control areas, which we saw in Table 3 have a significantly lower share of Inner Island ethnic groups.

Relative to control areas, we only find a significant increase in Indonesian language use in those Transmigration areas with the most equal mix of Inner and Outer Islanders (in the second tercile). As seen in column 1 of Panel A, individuals in these villages are 18 p.p. more likely to report Indonesian as their primary daily language relative to 13 percent for those in control villages. Panels B and C reveal that this result is driven by both Inner and Outer Islanders being more likely to use Indonesian than their

[33] Following Henderson and Parmeter (pp. 276–278, 2015), we adopt a local cubic polynomial first stage estimator and a local linear second stage estimator. The confidence bands in these figures do not account for the generated regressor in the second, but in ongoing work, we implement both a bootstrap procedure to improve inference and a cross-validation procedure to select optimal bandwidths in both stages.

[34] Note that this specification is based on a specification without reweighting as we are still developing an adapted version of the Kline (2011) estimator that allows for heterogeneous treatment effects.

co-ethnics residing in control areas. The remaining results in columns 2 and 3 confirm that the minority and majority language choices seen in Figure 5 hold up when compared to control areas.

In Section 6.4, we provide further evidence of diversity thresholds in identity and occupational choices that are consistent with the language use patterns identified in this section. However, before doing so, we offer next a simple conceptual model to explain the relationship between diversity and the diffusion of the national language.

### 6.3 A Simple Model of Diversity and Language Adoption

The adoption and use of the national language can be viewed as an optimal choice by individuals seeking to communicate with those outside their native linguistic group. Often times, such choices may also embody affinity with the nation and serve to coordinate behavior among otherwise fragmented ethnic communities. The results presented thus far suggest that language is not only a vehicle for economic interactions but also perhaps a key marker of cultural identification.

In this subsection, we outline a conceptual framework for understanding how changes in ethnic group shares affect the adoption of a national language in a multiethnic society like Indonesia. The framework extends the Lazear (1999) model of home and foreign languages to one with an additional national language. This extension can help explain the empirical results discussed above. The model admits multiple equilibria including full, partial, and no adoption of the national language, and the particular mix of Inner and Outer Islanders determines which equilibria is more likely to arise.

To fix ideas, we think of Indonesia as a collection of isolated communities, where language-based interactions happen locally. In village $v = 1, ..., V$, each individual $i$ is a member of one of two groups. The first group, given by $H$, is the set of individuals belonging to the "home" group, and the second group, denoted by $F$, is the set of individuals belonging to the "foreign" group. In our context, think of $H$ as Outer Island natives and $F$ as transmigrants. The share of residents belonging to group $H$ is denoted $P_H$, and the share belonging to group $F$ as $P_F$, where $P_H + P_F = 1$.

Individual $i$ has three choices of language: the home language, also denoted by $H$, the foreign language, also denoted by $F$, or the national language, denoted by $N$. The cost of adopting language $\ell \in \{H, F, N\}$ is given by $c_i^\ell$. We make three assumptions about these costs:

**Assumption 6.1.** *(Free Adoption of Own Language) For all individuals $i \in H$, $c_i^H = 0$, and for all individuals $i \in F$, $c_i^F = 0$.*

This ensures that all individuals can freely adopt the language of their own groups. Individuals can freely interact with all of their own group members regardless of whether or not they decide to adopt another language.

**Assumption 6.2.** *(Language $N$ is a National Language) For all individuals $i \in H$, $c_i^N < c_i^F - 1$, and for all individuals $i \in F$, $c_i^N < c_i^H - 1$.*

This provides a necessary condition to ensure that adopting the national language is easier than adopting another group's language directly. This is consistent with the fact *Bahasa Indonesia* was a lingua franca in the region for centuries, and nearly all Indonesians are exposed to the language in primary school and in popular media.

**Assumption 6.3.** *(Distribution Function for Adoption Costs) For all individuals $i \in H$, $Pr\left\{c_i^F \leq x\right\} = F_H\left(x;\theta_v\right)$, and for all individuals $i \in F$, $Pr\left\{c_i^F \leq c\right\} = F_F\left(c;\theta_v\right)$. Here, for $\ell = H, F$, we assume that $F_j$ is a continuous, monotonically increasing function, with $F_\ell\left(c;\theta_v\right) \geq 0$ for $c \in [0,\infty)$, $F_\ell\left(0;\theta_v\right) = 0$, and $\lim_{c\to\infty} F_j\left(c;\theta_v\right) = 1$.*

These distribution functions are group-specific and can depend on village-specific parameters, $\theta_v$. For now, though, we assume that $\theta_v$ is constant.

We assume that in expectation, the gains from interaction between any two individuals in a locality are normalized to one. However, an interaction can only successfully occur when the two individuals communicate in the same language. Individual $i$ chooses a language to maximize the benefits from interaction. Focusing on individual $i \in H$, the net benefit to neither adopting the national language nor the foreign language and just using the home language is $P_H$. If individual $i \in H$ were to adopt $F$ directly, the net benefit would be given by $1 - c_i^F$. If she adopts $N$, the net benefit would be given by $P_H + P_{FN} - c_i^N$, where $P_{FN}$ denotes the share of the total population that is foreign who also adopt $N$.

From Assumption 6.2, we can show that in equilibrium, no one from $H$ chooses to adopt $F$, and no one from $F$ chooses to adopt $H$. This result along with Assumption 6.3 leads directly to a system of two equations governing the share of $H$ and $F$ who adopt the national language:[35]

$$P_{HN} = F_F\left(P_{FN}\right)$$
$$P_{FN} = F_F\left(P_{HN}\right).$$

The possible solutions to this system are depicted in Figure 6. At point $A$, which is always a solution, nobody adopts the national language. At point $B$, we have partial adoption from both groups, and point $C$ with complete adoption is a theoretical possibility but never realized in practice.

By varying the relative proportion of home and foreign individuals, we can select equilibria in which there is some adoption of the national language. In the bottom panel of Figure 6, we consider two cases. First, in graph (a) there is substantially more home than foreign individuals as is the case, for example, in treated villages in the lower tail of Figure 4 (or upper tail with relabeling of home and foreign). Meanwhile in graph (b), home and foreign are in roughly equal proportion as observed for treated villages in the intermediate range of Figure 4. Comparing (a) and (b), we see that the non-adoption equilibrium is most likely in the case with very unequal group sizes, but as we increase the size of the minority group and hence move towards maximizing overall diversity, the partial adoption equilibrium becomes more likely.

This basic conceptual framework conveys one important way in which the Transmigration program shaped the diffusion of the national language, namely by changing the relative sizes of local ethnic group shares. Moreover, this framework directly implies an inverted U relationship between the share of Inner Island (transmigrant) ethnicities and the adoption of the national language just as we see in panel (a) of Figure 4. We focus here on the simpler case of two groups, Inner and Outer Islanders, but the insights generalize to the case of more than two groups as would correspond empirically to our findings on ethnic fractionalization in Appendix Figure A.5.

[35] In ongoing work, we generalize this key assumption to allow for group differences in costs or benefits that might explain why we see some individuals from Inner Island ethnic groups adopting Outer Islands languages and vice versa.

### 6.4 Thresholds in Identity and Occupational Choices

Ultimately, the Transmigration program led to a substantial increase in the returns to adopting the Indonesian language for both Outer Islands natives as well as for the transmigrants themselves. The results above showed why these returns vary with the extent of local ethnic diversity caused by the influx of transmigrants into settlement areas. Such dramatic changes in language have important implications in terms of facilitating economic interactions as well as fostering a national identity. Before concluding, we provide two pieces of additional evidence on the link between diversity, socioeconomic integration, and nation building.

We first identify diversity thresholds in the interegenerational transmission of ethnic identity. Recall from Table 4 that children in mixed Inner–Outer Island group marriages are no more likely to be identified as Inner Island ethnicities in Transmigration villages than in control villages. However, within Transmigration villages, the transmission of ethnic identity depends on relative group sizes. Figure 7 shows that children in mixed marriages are significantly less likely to be identified as Inner Island ethnicities in villages where the Inner Island parent's group is a clear minority. The opposite is true when the Inner Island parent's group is a clear majority. However, in villages where the group sizes are more balanced, household heads in mixed marriages are roughly equally likely to identify their child as Inner or Outer Island ethnicities. This nonlinear pattern is clearest in the semiparametric OLS specifications in panel (a), but the control function IV results in panel (b) also reveal strong incentives to identify children in mixed marriages with the majority group. Moreover, the particular nonlinear shape is similar when restricting to marriages in which the father is of transmigrant ethnicity, which is an important check given differences in the strength of the patrilineal tradition across ethnic groups.

Beyond identity choices, we also find a strong link between diversity and occupational choices. We saw in Section 6.2 that ethnic minorities are significantly less likely to report their own native language as their primary one. Perhaps for these individuals, the economic returns to being a minority outweigh the costs of relinquishing one's linguistic identity. Figure 8 shows that ethnic minorities in Transmigration villages are significantly more likely to work in trading and services occupations (relative to agriculture) than their co-ethnics residing in other Transmigration villages where their group is in the majority. This holds for both Inner and Outer Islanders (with strong symmetry) and in both OLS and IV specifications. Moreover, as with the language results in Figure 4, we identify a clear threshold at roughly 40 percent with a steep negative occupational choice gradient as one's group share increases up to that point. Then, as one's group becomes increasingly dominant, we see a flat relationship between ethnicity and occupational choice.

The occupational sorting patterns in Figure 8 are consistent with minority groups having a comparative advantage in intermediary sectors of the economy. For example, Outer Islands minorities in Transmigration villages can act as intermediaries in facilitating the sale of inputs and outputs to the agricultural settlers from the Inner Islands. Although Outer Islanders may have an absolute advantage in both trading and farming (see Bazzi et al., 2016), their comparative advantage lies in the former, and with that advantage come considerable economic returns that may make them more willing to pay the social or cultural costs of being a minority. Nevertheless, across the diversity spectrum, the ability to coordinate cross-group interactions using the national language makes it easier to mediate these potential economic returns while limiting the social costs.

# 7 Conclusion

We used a large-scale population resettlement program in Indonesia to identify the impact of ethnic diversity on nation building. Our findings suggest that the program caused a large increase in local diversity, which led to greater mixing in the marriage and housing market. We argue that the availability of the national language helped individuals to mediate the increased scope for interaction with outsiders while preserving their own ethnic identity.

Together, our results suggest that conflict need not be inevitable in diverse settings with mass immigration, even when majority groups move into minority regions. The greater interethnic integration is important given that recent work by Ashraf and Galor (2013) and Alesina et al. (2016) documents potential economic benefits of diversity that may go unrealized if sociocultural concerns preclude efforts to foster greater diversity (Borjas, 2016). Indeed, our findings may offer a counterexample to growing evidence linking (longstanding) ethnic diversity to conflict (see, e.g., Esteban et al., 2012).[36]

As resettlement programs and integration policies expand into new settings,[37] our findings offer guidance on the potential conditions in which ethnic diversity can be harnessed for improved social outcomes. Nevertheless, we acknowledge that these positive findings may be due in part to the complementarity between the Transmigration program and other nation building policies in Indonesia, such as the enshrinement of intergroup tolerance in the constitution, the promotion of Indonesian names, or the expansion of Indonesian-language primary schooling.

[36] Drawing on the limited available data on conflict at the village level, we provide evidence in the Appendix that the program generated no long-run differences in conflict between settlement and control areas.

[37] Relocation programs are found in many developing countries, including China, India, and Brazil. Examples in developed countries include the Moving to Opportunity program in the United States (Kling et al., 2007) and various resettlement programs in the United States and Europe (Bauer et al., 2013; Beaman, 2012; Dahlberg et al., 2012; Edin et al., 2003; Glitz, 2012; Sarvimäki et al., 2010). Resettlement programs are growing in importance globally as climate change, infrastructure development, and conflict continue to displace millions (see Bazzi et al., 2016, for a discussion).

## References


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

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

**Akerlof, G. A. and R. Kranton**, "Economics And Identity," *The Quarterly Journal of Economics*, 2000, *115* (3), 715–753.

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

_ **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.

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

**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*, forthcoming.

_ **, T. Mayer, and M. Thoenig**, "The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France," *CEPR Discussion Paper No. DP9416*, 2013.

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

**Allport, Gordon W**, "The nature of prejudice," 1954.

**Ananta, A., E. N. Arifin, and L. Suryadinata**, *Indonesian electoral behaviour: a statistical perspective* number no. 2. In 'Indonesia's population series.', Singapore: Institute of Southeast Asian Studies, 2004.

_ **,** _ **, 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, Benedict**, *Imagined Communities*, Verso, 1983.

**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.

**Aspinall, E.**, "Place and Displacement in the Aceh Conflict," in Eva-Lotta E. Hedman, ed., *Conflict, violence, and displacement in indonesia*, number no. 45. In 'Studies on Southeast Asia.', Ithaca, N.Y: Southeast Asia Program, Cornell University, 2008.

**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," *Unpublished Manuscript*, 2015.

**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.

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

**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.

_ **and M. Gudgeon**, "Local Government Proliferation, Diversity, and Conflict," *Unpublished Manuscript*, 2016.

**Bazzi, Samuel and Christopher Blattman**, "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 2014, *6*, 1–38.

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

**Bell, M. and E. Charles-Edwards**, "Cross-national comparisons of internal migration: An update of global patterns and trends," 2013.

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

**Bisin, A., E. Patacchini, T. Verdier, and Y. Zenou**, "Are Muslim Immigrants Different In Terms of Cultural Integration?," *Journal of the European Economic Association*, 2008, *6* (2-3), 445–456.

**_ , _ , _ , and _** , "Formation and persistence of oppositional identities," *European Economic Review*, 2011, *55* (8), 1046–1071.

**Bisin, Alberto and Thierry 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.

**_ , Eleonora Patacchini, Thierry Verdier, and Yves Zenou**, "Bend it like Beckham: Ethnic identity and integration," *European Economic Review*, 2016, *90*, 146–164.

**Blouin, A. and S. Mukand**, "Erasing Ethnicity? Nation Building, (Mis)Trust and the Salience of Identity in Rwanda," *Unpublished Manuscript*, 206.

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

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

**Carneiro, P., S. Lee, and H. Reis**, "Please Call Me John: Name Choice and the Assimilation of Immigrants in the United States, 1900-1930," Technical Report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies 2015.

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

**Charras, Muriel, Marc Pain, Indonesia, O.R.S.T.O.M. (Agency : France), and Centre national de la recherche scientifique (France), eds**, *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.

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

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

**Cutler, D. M., E. L. Glaeser, and J. L. Vigdor**, "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 1999, *107* (3), 455–506.

**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," 2016.

**Desmet, K., I. Ortuño-Ortín, and R. Wacziarg**, "The political economy of linguistic cleavages," *Journal of Development Economics*, 2012, *97* (2), 322–338.

_ , **I. Ortuño-Ortín, and R. Wacziarg**, "Linguistic Cleavages and Economic Development," 2015b.

**Donaldson, D.**, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, forthcoming.

**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.

**Easterly, William and Ross Levine**, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 1997, *112* (4), 1203–1250.

**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.

**Enos, Ryan D.**, "Causal Effect of Intergroup Contact on Exclusionary Attitudes," *Proceedings of the National Academy of Sciences*, 2014, *111* (10), 3699–3704.

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

**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.

**Frey, W. H.**, "Diversity spreads out: Metropolitan shifts in Hispanic, Asian, and Black populations since 2000," *Living Cities Census Series*, 2006, pp. 1–28.

**Glennerster, R., E. Miguel, and A. Rothenberg**, "Collective action in diverse Sierra Leone communities," *The Economic Journal*, 2013, *123* (568), 285–316.

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

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

**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.

**Henderson, D. J. and C. F. Parmeter**, *Applied Nonparametric Econometrics*, Cambridge University Press, 2015.

**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.

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

**Jia, R. and T. Persson**, "Individual vs. Social Motives in Identity Choice: Theory and Evidence from China," 2015.

**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.

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

**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.

**Lleras-Muney, A. and A. Shertzer**, "Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schooling Laws on Immigrants," *American Economic Journal: Economic Policy*, 2015, *7* (3), 258–290.

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

**Mangunrai, Hasan**, "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).

**Manning, A. and S. Roy**, "Culture Clash or Culture Club? National Identity in Britain," *The Economic Journal*, 2010, *120* (542), F72–F100.

**McFadden, Daniel**, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press, 1974.

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

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

**Nababan, P. W. J.**, "Language in Education: The Case of Indonesia," *International Review of Education / Internationale Zeitschrift für Erziehungswissenschaft / Revue Internationale de l'Education*, 1991, *37* (1), 115–131.

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

**Otten, Mariël**, "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.

**Putnam, Robert 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.

**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.

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

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

**Schiller, A. and B. Garang**, "Religion and inter-ethnic violence in Indonesia," *Journal of Contemporary Asia*, 2002, *32* (2), 244–254.

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

**Su, L. and A. Ullah**, "Local polynomial estimation of nonparametric simultaneous equations models," *Journal of Econometrics*, 2008, *144* (1), 193–218.

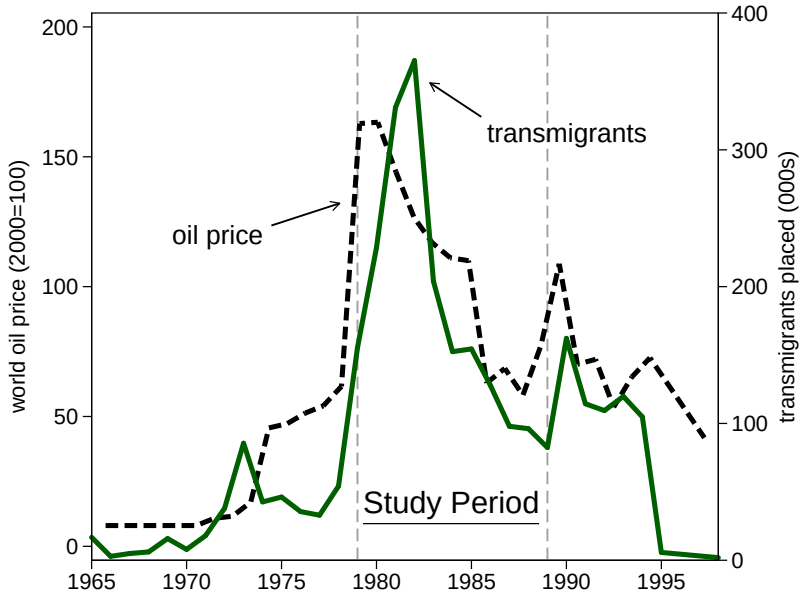
**Thornton, D. L.**, "Javanization of Indonesian politics," 1972.

**World Bank**, *Indonesia: The Transmigration Program in Perspective* 1988.

**Wright, S.**, *Language policy and language planning: from nationalism to globalisation*, 2nd edition ed., Basingstoke: Palgrave Macmillan, 2016.

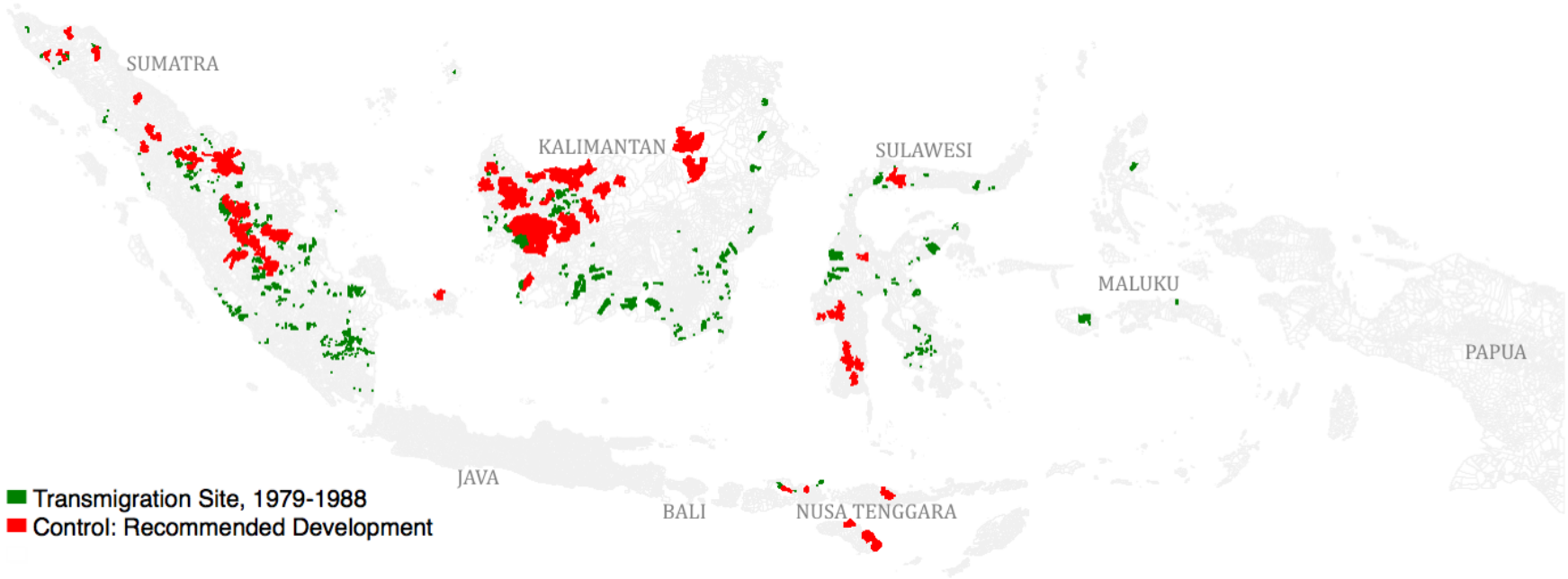
# Figures

**Figure 1:** Transmigration Flows and Oil Prices



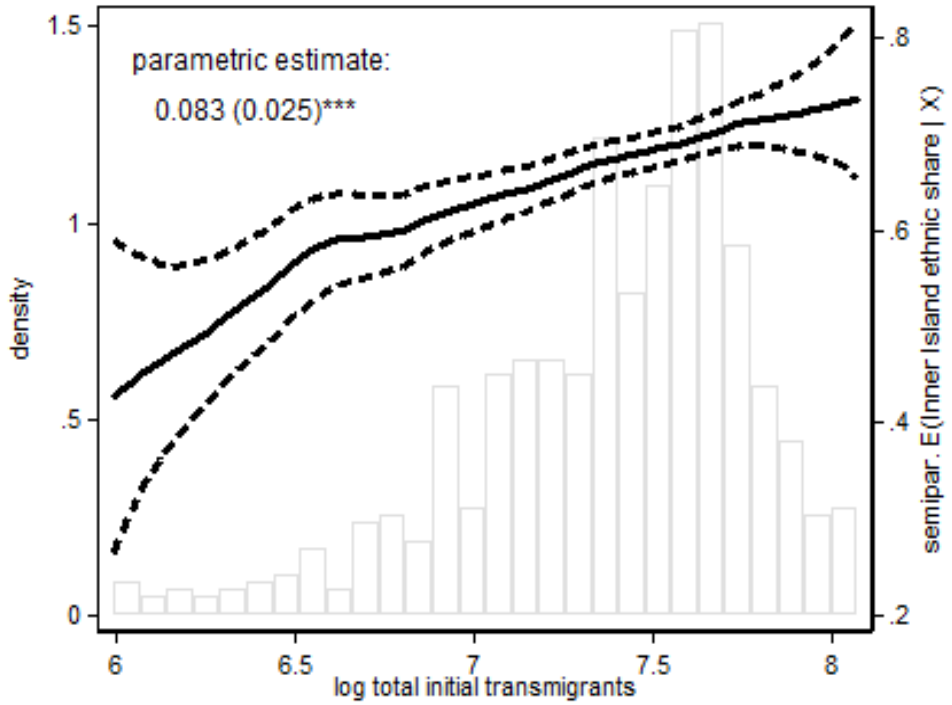

*Notes*: Authors' calculations from Transmigration Census data. The oil price index is from Bazzi and Blattman (2014). The dark gray vertical lines correspond to our study period. This figure is also reported in Bazzi et al. (2016).

**Figure 2:** Map of Transmigration and Control Villages




*Notes*: Each colored location on the map corresponds to a Transmigration village or a control (Recommend Development) village. The white areas outlined in grey are neither Transmigration nor control villages.

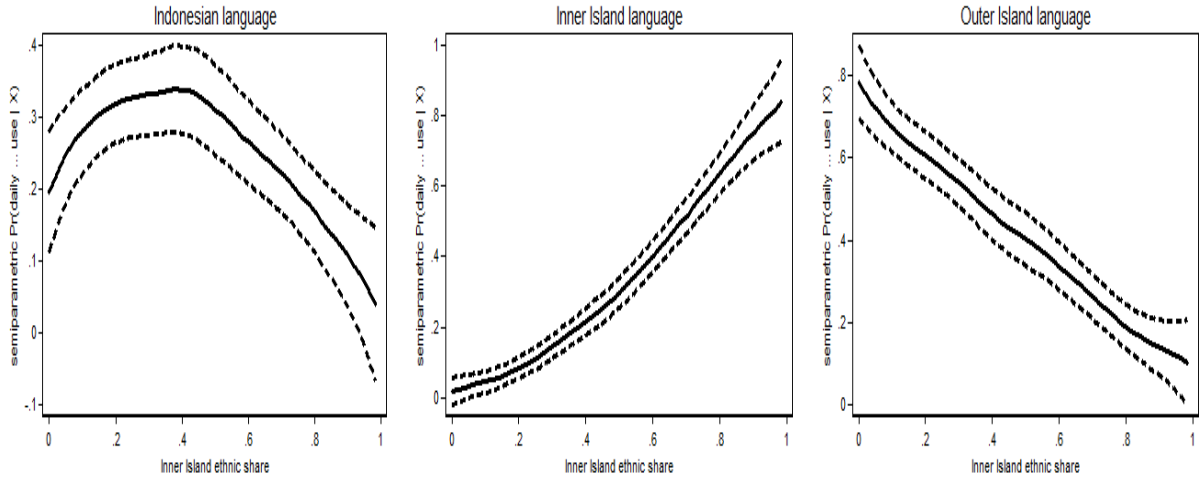
**Figure 3:** Long-Run Diversity and Initial Transmigrants (First Stage)



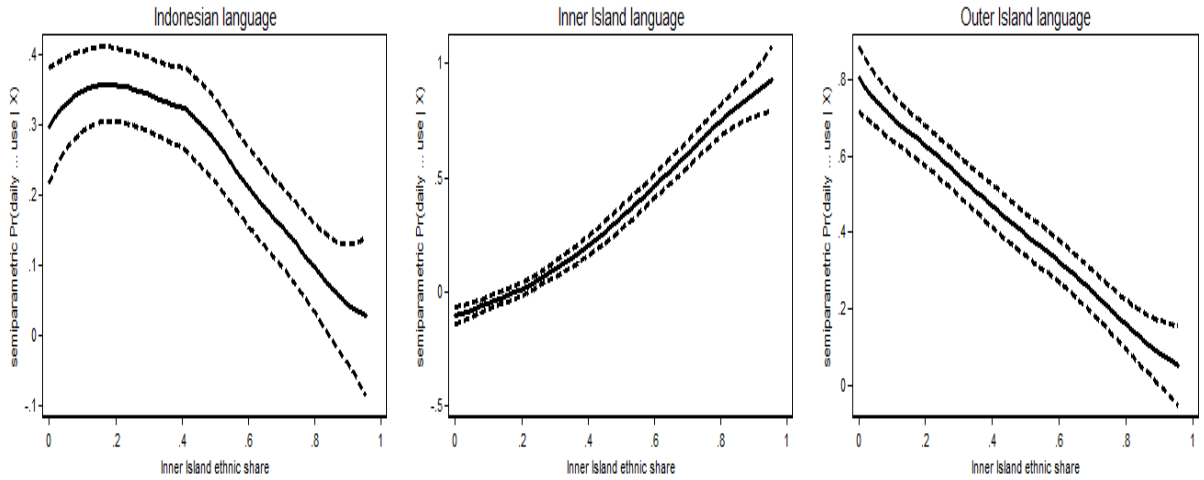

*Notes:* This figure reports regressions of the Inner Island (transmigrant) ethnic share in 2000 (based on the Population Census) conditional on the log of the transmigrant population from Java/Bali placed in that village in the initial year of settlement (in 1979–1988). The semiparametric regression is based on the Robinson (1988) partially linear model, and it employs a local linear regression that conditions on 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 percentile. This is similar to the specification used in the first stage of the semiparametric instrumental variables results presented below. The parametric linear estimate adopts the same set of controls with standard errors based on the Conley (1999) approach that allows unobservables to be correlated across all villages within 150 kilometers of each other.

**Figure 4:** Ethnic Diversity and Indonesian Language Use

(a) Semiparametric OLS (Robinson, 1988)



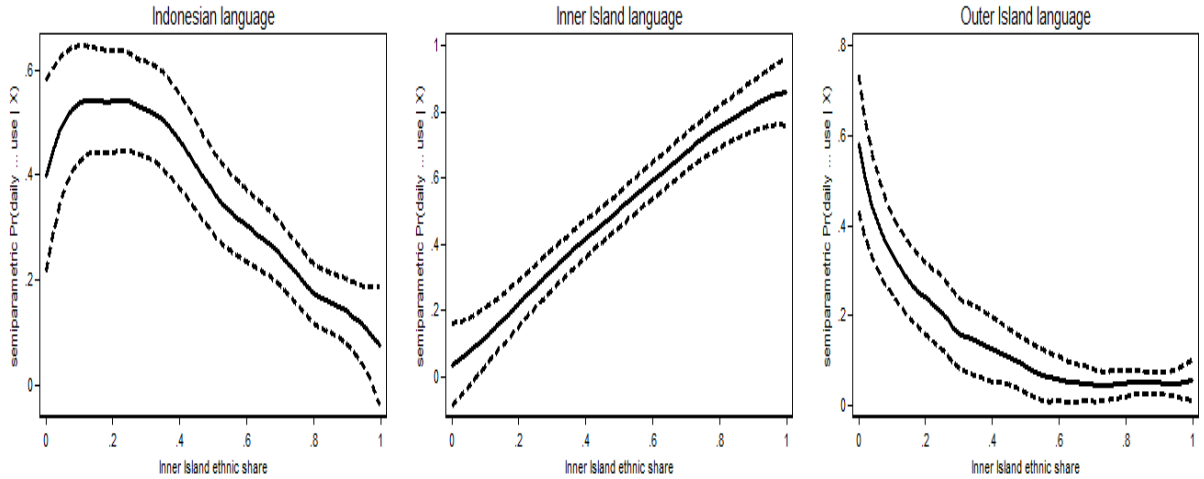

(b) Semiparametric IV (Su and Ullah, 2008)



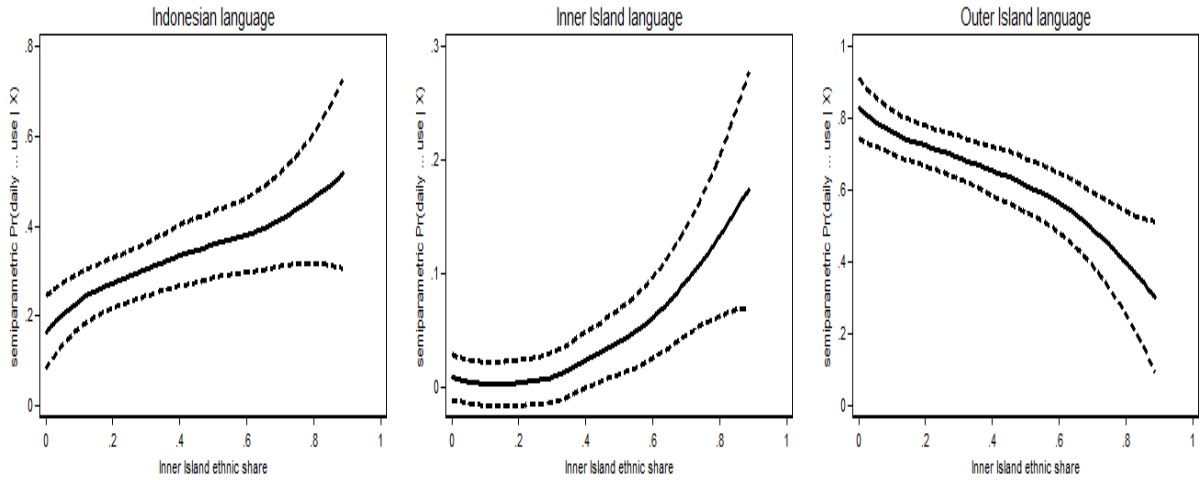

*Notes:* These figures report semiparametric estimates of the impact of ethnic diversity (Inner Island ethnic share) on the likelihood of reporting the given language (national, Inner Island, Outer Island) as the primary daily one used in the household for all individuals residing within 10 km of Transmigration villages including those villages. The estimates are based on the Robinson (1988) 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. The individual data on language use are from the 2006 *Susenas* household survey. The estimates in panel (b) are based on the Su and Ullah (2008) control function estimator in which we follow the approach suggested in Henderson and Parmeter (2015) and first run a third degree local polynomial regression based on the first stage in Figure 3, recover the semiparametric residuals from that regression, and then estimate a second stage local linear regression conditional on those generated residuals. Note that the confidence intervals do not yet account for the generated regressor. In ongoing work, we improve inference by implementing a bootstrap procedure and also reduce bias in the pointwise estimates by implementing a bandwidth optimized for each stage using cross-validation.

**Figure 5:** Ethnic Diversity and Minority vs. Majority Language Use

(a) Inner Island Ethnic Groups

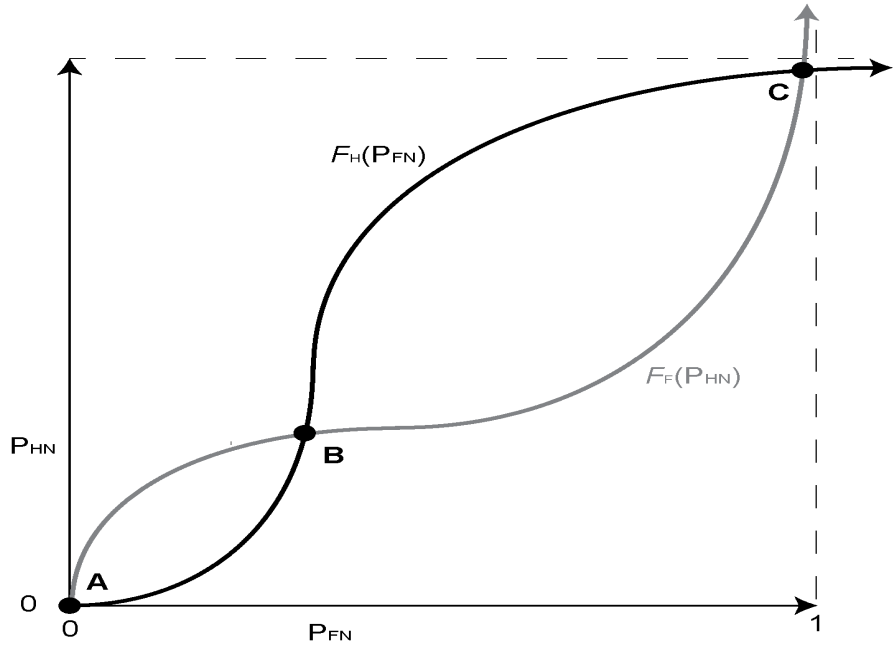



(b) Outer Island Ethnic Groups



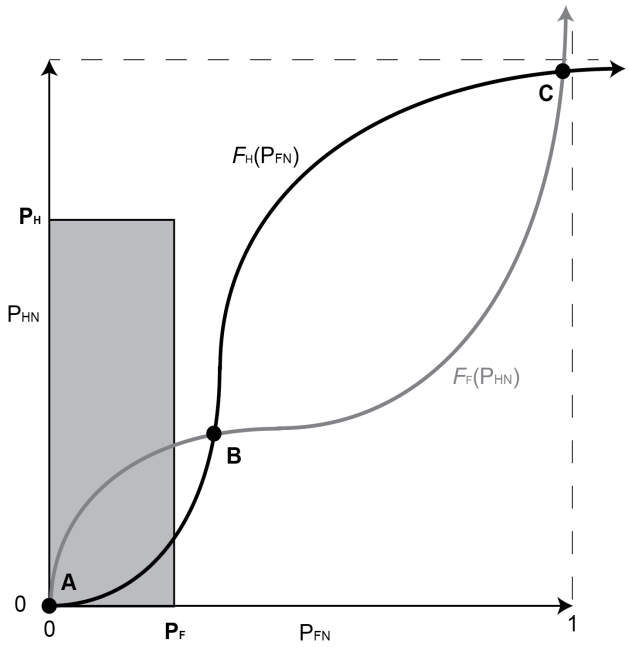

*Notes:* These figures report semiparametric estimates of the impact of ethnic diversity (Inner Island ethnic share) on the likelihood of reporting the given language (national, Inner Island, Outer Island) as the primary daily one used in the household for individuals residing within 10 km of Transmigration villages including those villages. Panel (a) is estimated over all individuals belonging to Inner Island ethnic groups, and Panel (b) is estimated over all those belonging to Outer Island ethnic groups. The estimates are 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. The individual data on language use are from the 2006 *Susenas* household survey.

**Figure 6:** Multiple Equilibria in Language Adoption

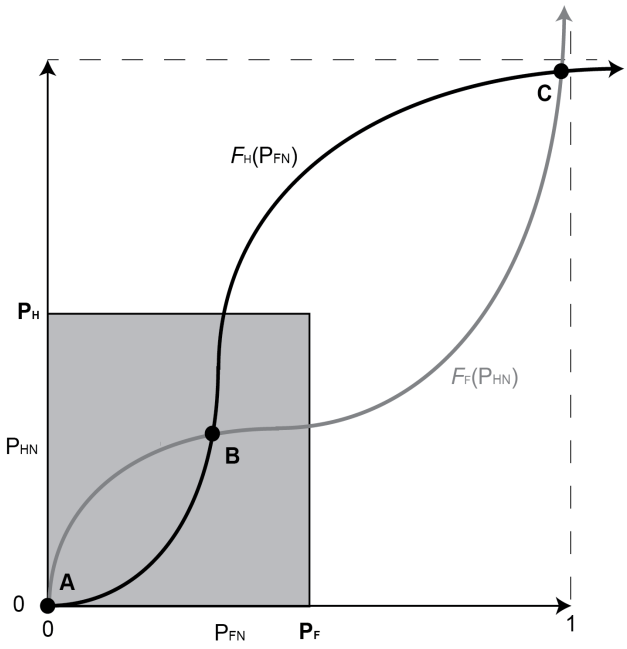



Possible Equilibria Depending on Ethnolinguistic Diversity

(a) Uneven Population Distribution



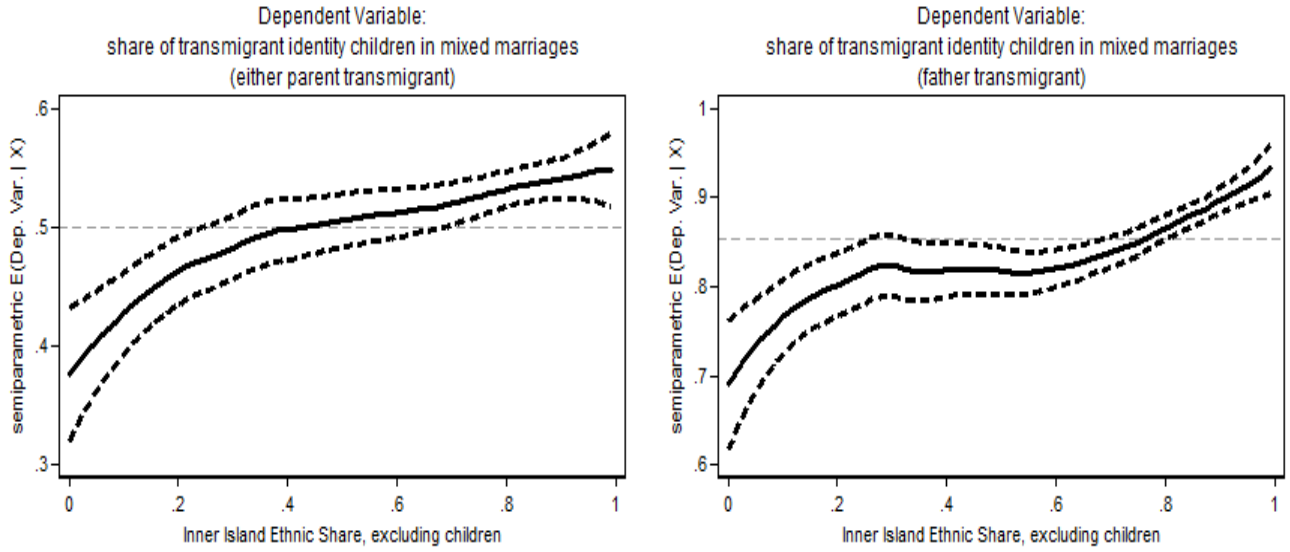

(b) Even Population Distribution



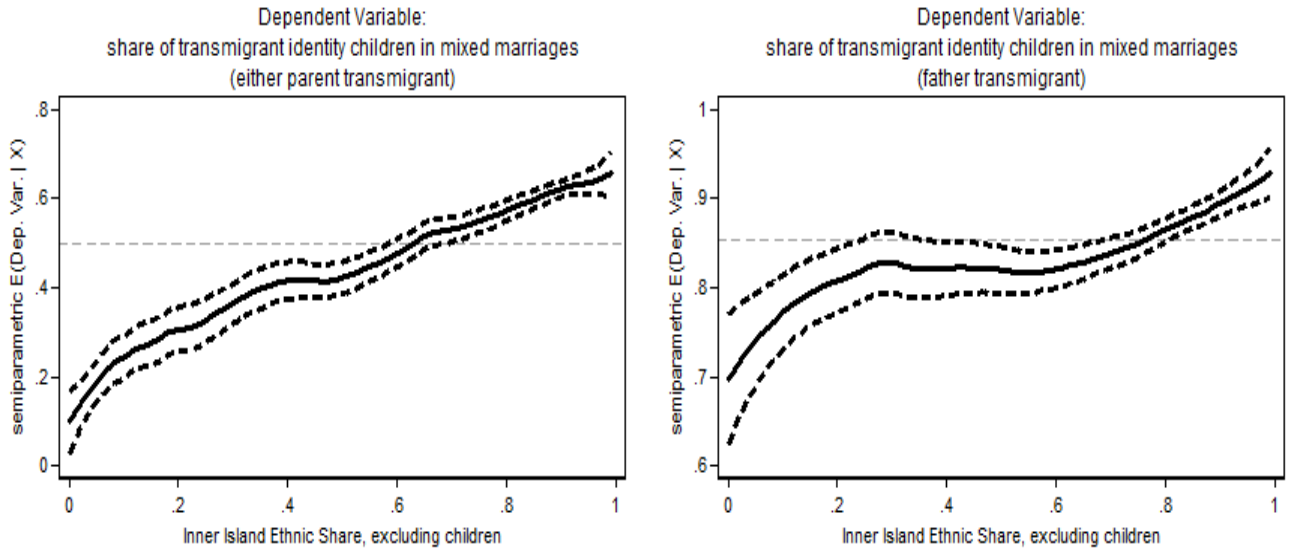

*Notes:* These figures illustrate the properties of the model of language diffusion developed in Section 6.3.

**Figure 7:** Diversity Thresholds in the Intergenerational Transmission of Identity

(a) Semiparametric OLS (Robinson, 1988)



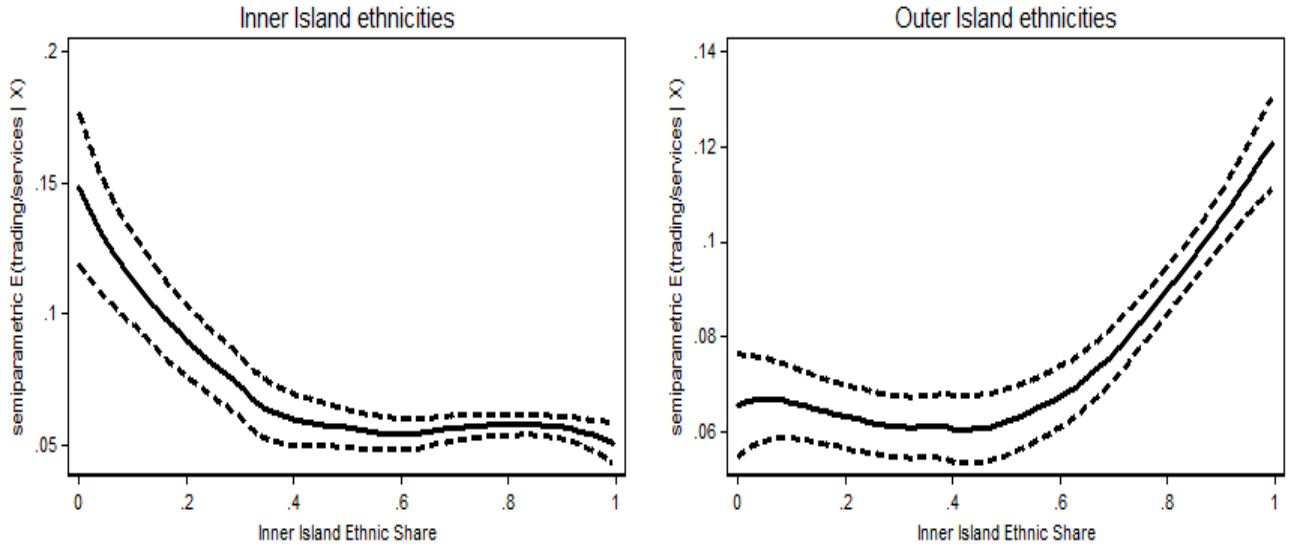

(b) Semiparametric IV (Su and Ullah, 2008)



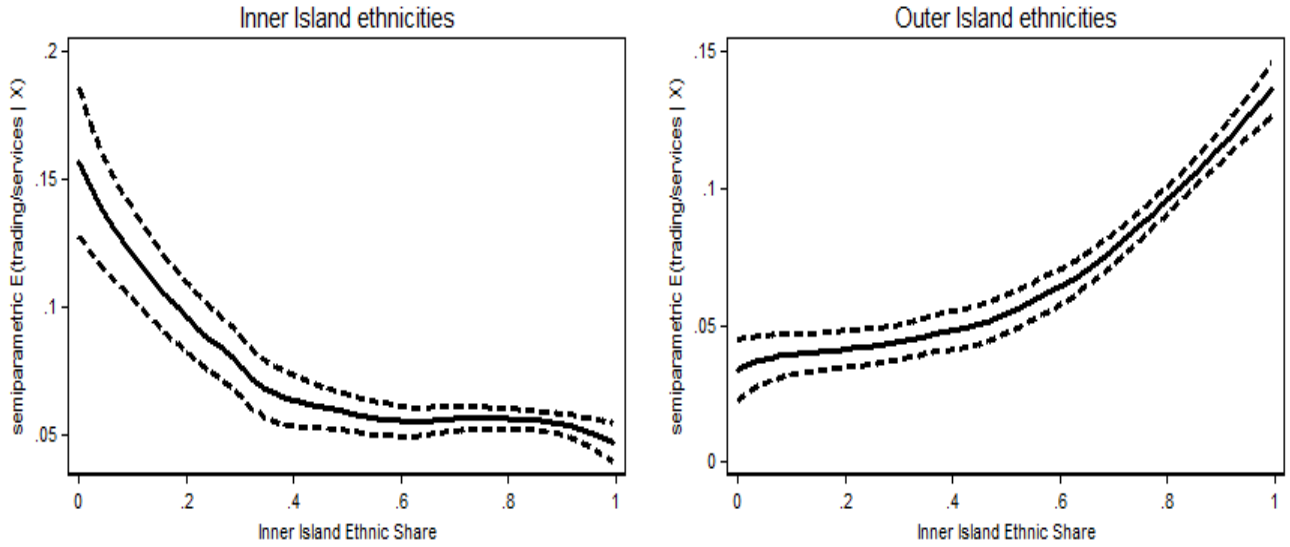

*Notes:* These figures report village-level semiparametric estimates of the impact of ethnic diversity (Inner Island ethnic share) on the share of children with reported Inner Island (transmigrant) identities among all children living in households with one parent hailing from Inner Island ethnicities and the other from Outer Island ethnicities. The left graphs are based on all children born to mixed marriages, and the right graphs are based on children born to mixed marriages in which the father belongs to an Inner Island ethnic group and the mother belongs to an Outer Island ethnic group. We find similarly upward sloping graphs for those mixed marriages in which the mother belongs to an Inner Island ethnic group. The gray dashed horizontal lines indicate the average of the dependent variable. The estimates are based on the semiparametric OLS and IV procedures detailed in the notes to Figure 4. Note that the confidence intervals do not yet account for the generated regressor. In ongoing work, we improve inference by implementing a bootstrap procedure and also reduce bias in the pointwise estimates by implementing a bandwidth optimized for each stage using cross-validation.

**Figure 8:** Ethnic Diversity and Minority Occupational Choice

(a) Semiparametric OLS (Robinson, 1988)




(b) Semiparametric IV (Su and Ullah, 2008)




*Notes:* These figures report village-level semiparametric estimates of the impact of ethnic diversity (Inner Island ethnic share) on the share of the given ethnic group's working age population working in trading and services occupations as reported in the 2000 Population Census. The left graphs are based on all individuals belonging to Inner Island ethnicities, and the right graphs are based on those belong to Outer Islands ethnicities. The estimates are based on the semiparametric OLS and IV procedures detailed in the notes to Figure 4. Note that the confidence intervals do not yet account for the generated regressor. In ongoing work, we improve inference by implementing a bootstrap procedure and also reduce bias in the pointwise estimates by implementing a bandwidth optimized for each stage using cross-validation.

# Tables

**Table 1:** Summary Statistics: Transmigration and Control Villages

| | Mean | Std. Deviation |
|---|---|---|
| **Demographics** | | |
| total population | 1,961 | (1,478) |
| Java/Bali-born population share | 0.22 | (0.23) |
| transmigrant ethnicity share | 0.40 | (0.39) |
| ethnic fractionalization | 0.34 | (0.24) |
| number of ethnic groups | 22.07 | (19.39) |
| Muslim population share | 0.78 | (0.35) |
| religious fractionalization | 0.15 | (0.19) |
| **Marriage, Identity Transmission, and Residence** | | |
| marriage rate | 0.86 | (0.06) |
| marriage rate, young cohort | 0.84 | (0.15) |
| interethnic marriage rate (any group) | 0.09 | (0.08) |
| interethnic marriage rate, young cohort (any group) | 0.12 | (0.12) |
| interethnic marriage rate (inner, outer) | 0.04 | (0.04) |
| interethnic marriage rate, young cohort (inner, outer) | 0.06 | (0.07) |
| share of transmigrant identity children in mixed marriages | 0.50 | (0.28) |
| ethnic residential isolation index (inner, outer) | 0.22 | (0.25) |

*Notes*: This table reports summary statistics for the 832 Transmigration villages and 668 control villages in our baseline estimating equations for village-level outcomes. The sample size is 1,500 for all variables except for (i) the share of transmigrant ethnic identity children in mixed (Inner, Outer Island) marriages, which is only defined for the 1,214 villages in which we observe such interethnic marriages, and (ii) the isolation index, which is only defined for the 1,240 villages with both ethnicities.

**Table 2:** 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. See the Online Appendix of Bazzi et al. (2016) for details on variable construction and definitions. 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 (see Appendix Table A.1. Standard errors are clustered at the district level. */**/*** denotes significance at the 10/5/1 percent level.

**Table 3:** Demographic Impacts of the Transmigration Program

| | ATT Estimates | | | Control Group |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | Mean, Cols. 2-3 |
| log population | 0.355 | 0.195 | 0.068 | 7.205 |
| | (0.082)*** | (0.114)* | (0.088) | |
| Java/Bali-born population share | 0.321 | 0.338 | 0.335 | 0.020 |
| | (0.017)*** | (0.019)*** | (0.019)*** | |
| transmigrant ethnic share | 0.484 | 0.514 | 0.540 | 0.060 |
| | (0.027)*** | (0.032)*** | (0.038)*** | |
| ethnic fractionalization | 0.154 | 0.186 | 0.126 | 0.225 |
| | (0.023)*** | (0.031)*** | (0.031)*** | |
| number of ethnic groups | 5.176 | 3.744 | 2.637 | 19.540 |
| | (2.030)** | (2.052)* | (2.352) | |
| Muslim population share | 0.141 | 0.201 | 0.091 | 0.588 |
| | (0.032)*** | (0.048)*** | (0.044)** | |
| religious fractionalization | 0.003 | -0.032 | 0.011 | 0.194 |
| | (0.014) | (0.023) | (0.022) | |
| Treatment/Control Only | No | Yes | Yes | |
| Geographic Controls | No | Yes | Yes | |
| Reweighting | No | No | Yes | |
| Blinder-Oaxaca | No | No | Yes | |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. Each cell reports the coefficient from a regression of the given dependent variable listed in the first column on an indicator for whether the village is a Transmigration settlement. Column 1 comprises all Outer Islands villages (with non-missing data). Column 2 restricts to our quasi-experimental design including only Transmigration and control/RDA sites and conditions on the predetermined village-level characteristics that determined site selection. Column 3 is a control function specification based on a Blinder-Oaxaca decomposition derived in Kline (2011). All specifications include island fixed effects. Sample sizes vary across outcomes (depending on data availability) and columns but include as many 31,185 villages in column 1, and 832 treated villages and 668 controls in columns 2 and 3. Standard errors are clustered by district in parentheses. The final column contains the mean of the control villages in columns 2 and 3.

**Table 4:** Interethnic Integration due to the Transmigration Program

| | ATT Estimates (1) | (2) | (3) | Control Group Mean, Cols. 2-3 |
|---|---|---|---|---|
| | **Panel A**: Household Heads≥15 | | | |
| marriage rate | 0.176 | 0.043 | 0.048 | 0.834 |
| | (0.022)*** | (0.006)*** | (0.005)*** | |
| interethnic marriage rate (any group) | 0.012 | 0.042 | 0.032 | 0.069 |
| | (0.007)* | (0.008)*** | (0.007)*** | |
| interethnic marriage rate (inner, outer) | 0.023 | 0.040 | 0.034 | 0.017 |
| | (0.004)*** | (0.004)*** | (0.004)*** | |
| | **Panel B**: Younger Household Heads | | | |
| marriage rate | 0.025 | 0.006 | 0.019 | 0.829 |
| | (0.010)** | (0.012) | (0.013) | |
| interethnic marriage rate (any group) | 0.035 | 0.081 | 0.072 | 0.081 |
| | (0.009)*** | (0.011)*** | (0.011)*** | |
| interethnic marriage rate (inner, outer) | 0.036 | 0.059 | 0.050 | 0.023 |
| | (0.006)*** | (0.005)*** | (0.006)*** | |
| adjusted interethnic marriage rate (inner, outer) | 0.136 | 0.135 | 0.177 | 0.253 |
| | (0.031)*** | (0.033)*** | (0.035)*** | |
| | **Panel C**: Children's Ethnic Identity | | | |
| share of transmigrant identity children in mixed marriages | 0.048 | 0.015 | -0.009 | 0.473 |
| | (0.022)** | (0.024) | (0.037) | |
| | **Panel D**: Residential Segregation | | | |
| ethnic residential isolation index (inner, outer) | -0.114 | -0.171 | -0.184 | 0.346 |
| | (0.012)*** | (0.027)*** | (0.037)*** | |
| Treatment/Control Only | No | Yes | Yes | |
| Geographic Controls | No | Yes | Yes | |
| Reweighting | No | No | Yes | |
| Blinder-Oaxaca | No | No | Yes | |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. Each cell reports the coefficient from a regression of the given dependent variable listed in the first column on an indicator for whether the village is a Transmigration settlement. Column 1 comprises all Outer Islands villages (with non-missing data). Column 2 restricts to our quasi-experimental design including only Transmigration and control/RDA sites and conditions on the predetermined village-level characteristics that determined site selection. Column 3 is a control function specification based on a Blinder-Oaxaca decomposition derived in Kline (2011). All specifications include island fixed effects. Sample sizes vary across outcomes (depending on data availability) and columns but include as many 31,185 villages in column 1, and 832 treated villages and 668 controls in columns 2 and 3. Standard errors are clustered by district in parentheses. The final column contains the mean of the control villages in columns 2 and 3. The "interethnic marriage rate (inner, outer)" ("... (any group)") measure captures the share of marriages that are between Inner and Outer Island ethnicities (any two distinct ethnicities listed in the 2000 Census). Panel A is based on all household heads, and Panel B is based on a younger cohort for whom the marriage likely took place after the initial wave of resettlement in that area (see Section 3.2 for details). The "adjusted interethnic marriage rate..." in Panel B divides the actual interethnic marriage rate by the average interethnic marriage based on 10,000 simulated random matches in the village marriage market. This serves to adjust the rate of interethnic marriage by the opportunity for such marriages to take place, which is naturally higher in Transmigration villages given the large effects on diversity reported in Table 3. The dependent variable in Panel C is the share of children with reported Inner Island (transmigrant) identities among all children living in households with one parent hailing from Inner Island ethnicities and the other from Outer Island ethnicities. The dependent variable in Panel D is the isolation index due to Cutler et al. (1999) and aimed at capturing the degree to which Inner and Outer Islands reside in segregated census blocks within the village.

**Table 5:** Effects of the Transmigration Program on National Language Diffusion

| | **Dep. Var.**: $\mathbb{P}$(Daily ... Language Use) | | |
|---|---|---|---|
| | 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 $\leq 35$ | 0.251 | -0.069 | -0.182 |
| | (0.137)* | (0.080) | (0.177) |
| 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) |
| 8. Conditional on overall ethnic fractionalization | 0.260 | 0.001 | -0.261 |
| | (0.122)** | (0.066) | (0.156)* |
| 9. Conditional on quadratic transmigrant ethnic share | 0.006 | -0.088 | 0.082 |
| | (0.136) | (0.083) | (0.155) |

*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 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. The control group includes all individuals within 10 km of a control/RDA site. 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 the sample to those with no formal education; (4) includes 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 base language of *Bahasa Indonesia*; (7) controls for distance of the given ethnolinguistic homeland to Malay using the metric suggested in Fearon (2003); (8) controls for the overall ethnic fractionalization; (9) controls for a quadratic in the Inner Island (transmigrant) ethnic share. Standard errors clustered by district are in parentheses. */**/*** denotes significant at the 10/5/1 percent significance levels.

**Table 6:** Heterogeneous Effects of the Transmigration Program on National Language Diffusion

| | **Dep. Var.**: $\mathbb{P}$(Daily . . . Language Use) | | |
|---|---|---|---|
| | Indonesian | Inner Island | Outer Island |
| | **(1)** | **(2)** | **(3)** |
| | **Panel A**: All Ethnic Groups | | |
| Inner Island ethnic share: tercile 1 | 0.028 | -0.067 | 0.039 |
| | (0.063) | (0.038)* | (0.073) |
| Inner Island ethnic share: tercile 2 | 0.181 | 0.028 | -0.208 |
| | (0.074)** | (0.044) | (0.079)** |
| Inner Island ethnic share: tercile 3 | -0.050 | 0.572 | -0.521 |
| | (0.053) | (0.054)*** | (0.079)*** |
| | **Panel B**: Inner Island Ethnic Groups | | |
| Inner Island ethnic share: tercile 1 | 0.116 | -0.167 | 0.051 |
| | (0.165) | (0.146) | (0.112) |
| Inner Island ethnic share: tercile 2 | 0.187 | -0.080 | -0.107 |
| | (0.108)* | (0.115) | (0.094) |
| Inner Island ethnic share: tercile 3 | -0.156 | 0.303 | -0.147 |
| | (0.109) | (0.092)*** | (0.100) |
| | **Panel C**: Outer Island Ethnic Groups | | |
| Inner Island ethnic share: tercile 1 | -0.003 | -0.026 | 0.029 |
| | (0.059) | (0.021) | (0.064) |
| Inner Island ethnic share: tercile 2 | 0.134 | -0.039 | -0.095 |
| | (0.077)* | (0.023) | (0.077) |
| Inner Island ethnic share: tercile 3 | 0.105 | 0.173 | -0.279 |
| | (0.096) | (0.051)*** | (0.083)*** |

*Notes*: This table reports estimates from a heterogeneous treatment effects version of the prior table broken out by different groups: Panel (A) all ethnic groups ($N = 2,878$), (B) Inner Island groups ($N = 984$), and (C) Outer Island groups ($N = 1,894$). The dependent variables, which are mutually exhaustive, are indicators for whether or not the individual household head reports (1) the national language *Bahasa Indonesia*, (2) an Inner Island language, or (3) an Outer Island language as the primary daily language. Within each panel, each column 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) where we allow the ATT to vary across terciles of ethnic diversity as measured by the Inner Island ethnic share. Hence, each estimate is the treatment effect for that tercile relative to the control group, which includes all individuals within 10 km of a control/RDA site. Note that this source of heterogeneity is determined in part by the program and hence is not exogenous. We address this endogeneity in semiparametric regressions reported in Figures 4 and 5 but use this "bad control" exercise primarily for benchmarking purposes. All regressions are based on a specification analogous to column 2 in Tables 3 and 4 in which we control for . In ongoing work, we extend the Blinder-Oaxaca decomposition approach of Kline (2011) to a heterogeneous treatment effects specification. Standard errors clustered by district are in parentheses. */**/*** denotes significant at the 10/5/1 percent significance levels.

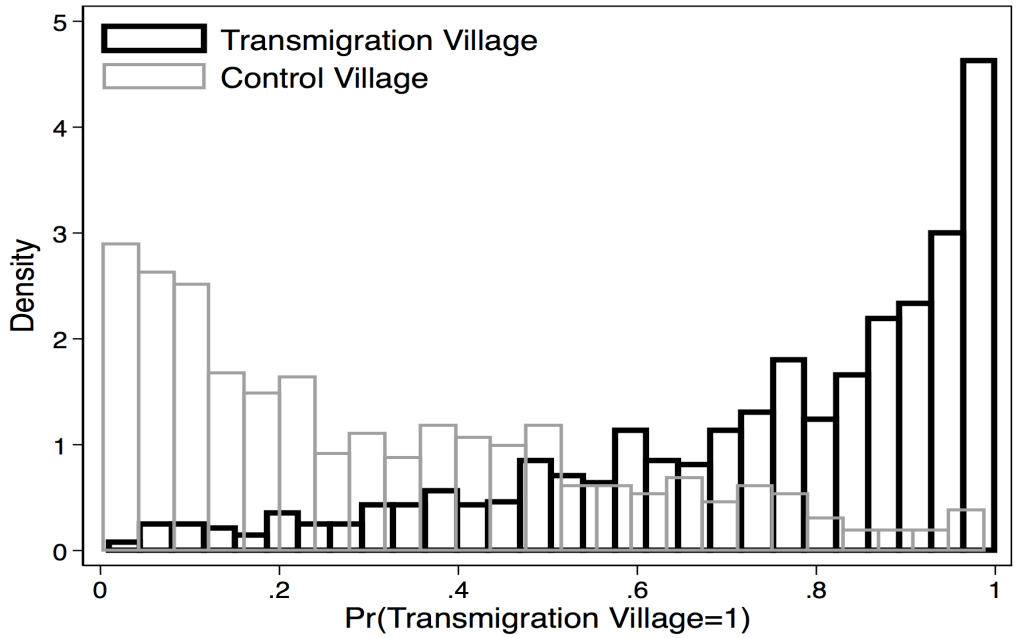
# Appendix

## A Further Empirical Results

We report here additional results mentioned in the paper.

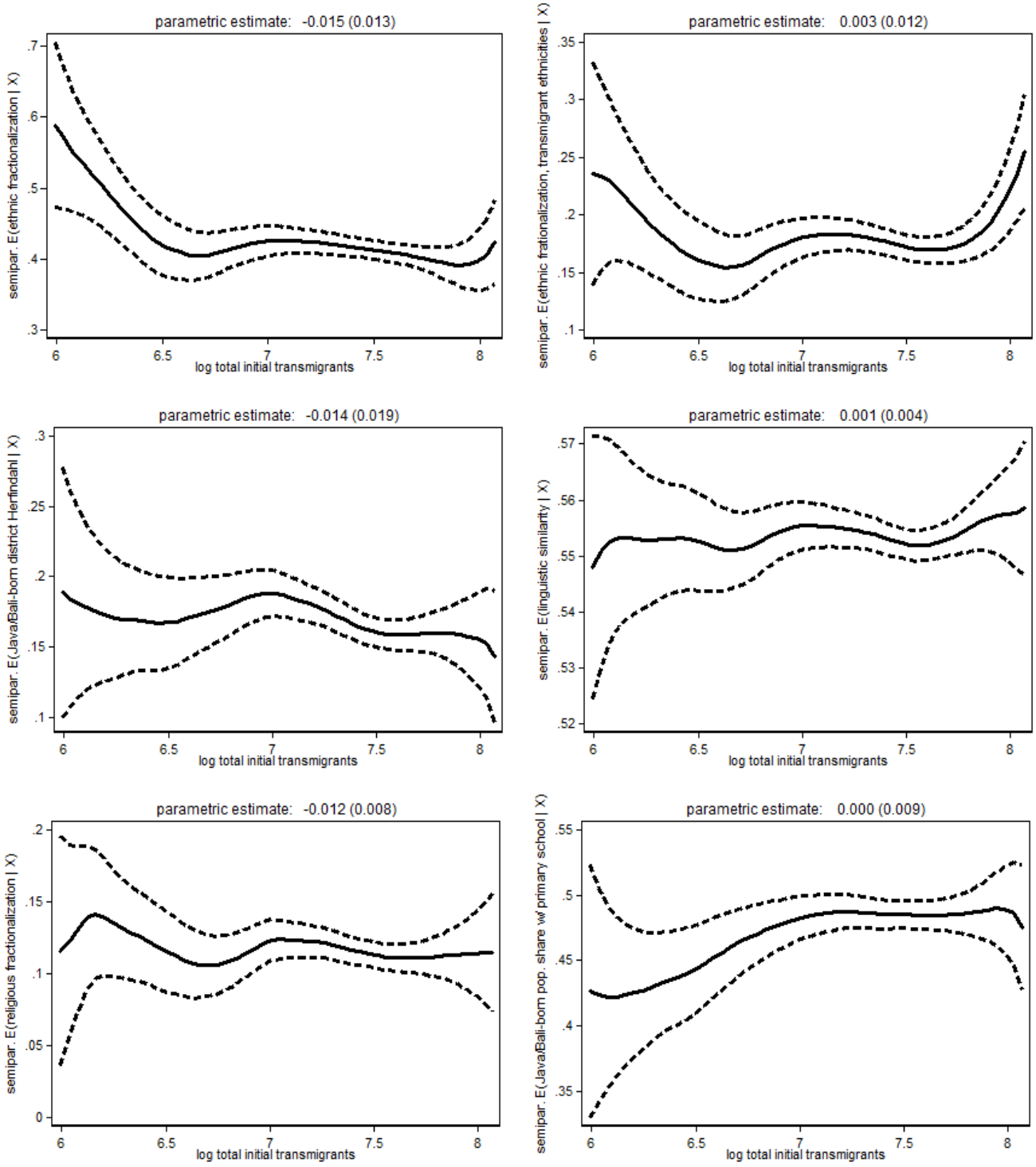
## Figures

**Figure A.1:** Estimated Propensity Scores



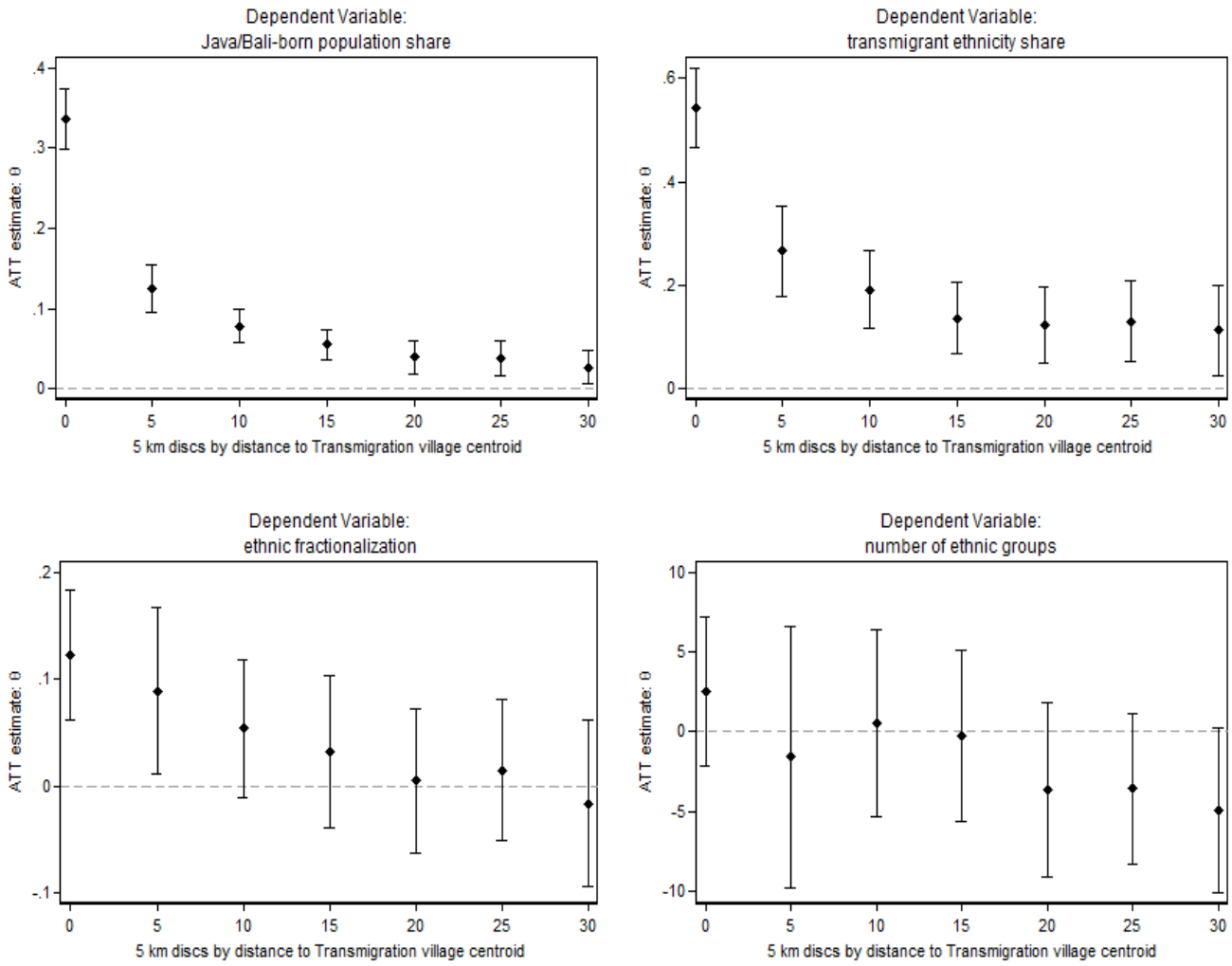

*Notes:* This figure, also in Bazzi et al. (2016), plots the distribution of estimated probabilities of site selection based on the estimates in Table A.1.

**Figure A.2:** Probing the Validity of the Initial Transmigrants Instrument



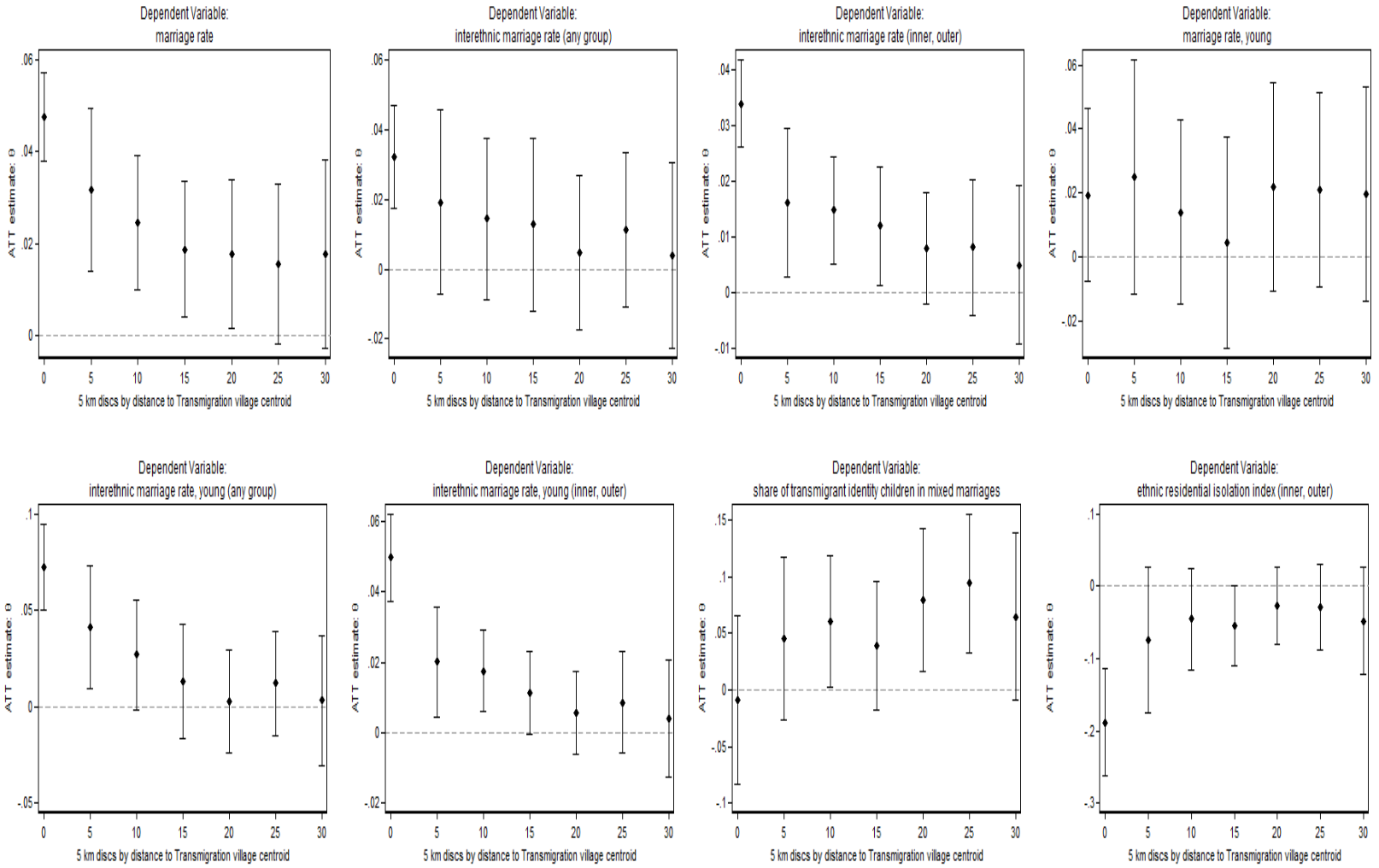

*Notes:* This figure reports regressions of several measures of diversity and other characteristics of the population—listed in the $y$ axis on the graphs—observed in Transmigration settlements conditional on the log of the transmigrant population from Java/Bali placed in that village in the initial year of settlement (in 1979–1988). The linguistic similarity and origin district Herfindahl indices are from Bazzi et al. (2016). 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. The semiparametric regression is based on the Robinson (1988) partially linear model, and it employs 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. This is similar to the specification used in the first stage of the semiparametric instrumental variables results presented below. The parametric linear estimate adopts the same set of controls with standard errors based on the Conley (1999) approach that allows unobservables to be correlated across all villages within 150 kilometers of each other.

**Figure A.3:** Demographic Spillovers from the Transmigration Program



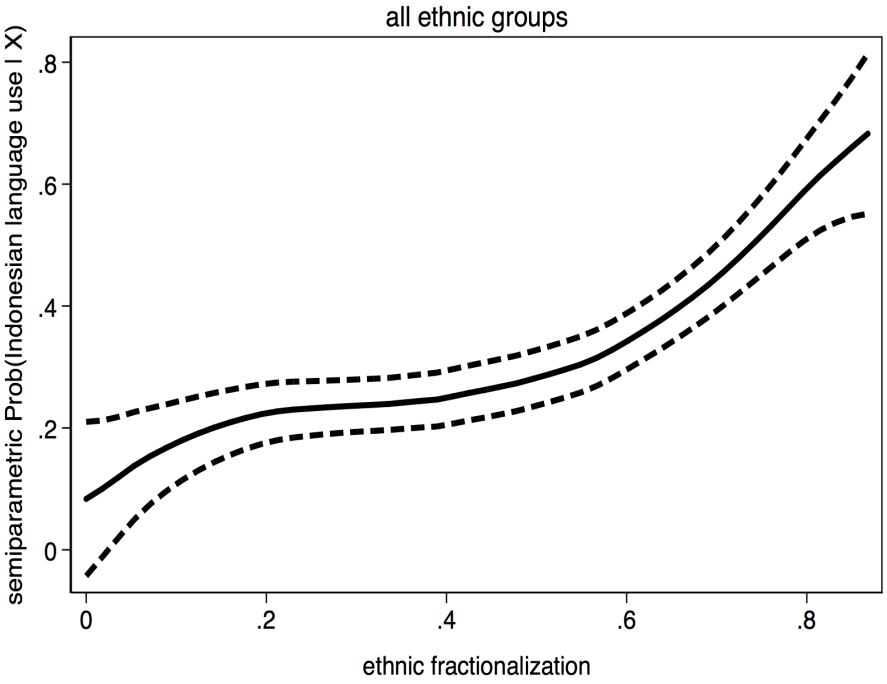

*Notes:* These figures report the 95 percent confidence intervals around the ATT estimates for the given dependent variable estimated at different spatial lags from the centroid of the Transmigration village. The "0 km" distance is just the estimate from our baseline Table 3. The remaining estimates are based on 5 km discs radiating outward from the Transmigration villages. For example, the estimate at 5 km is based on all villages greater than outside the Transmigration village boundary but within 5 km of that boundary. 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 Table 3 for further details.

**Figure A.4:** Integration 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 just the estimate from our baseline Table 4. 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 greater than outside the Transmigration village boundary but within 5 km of that boundary. 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 Table 4 for further details.

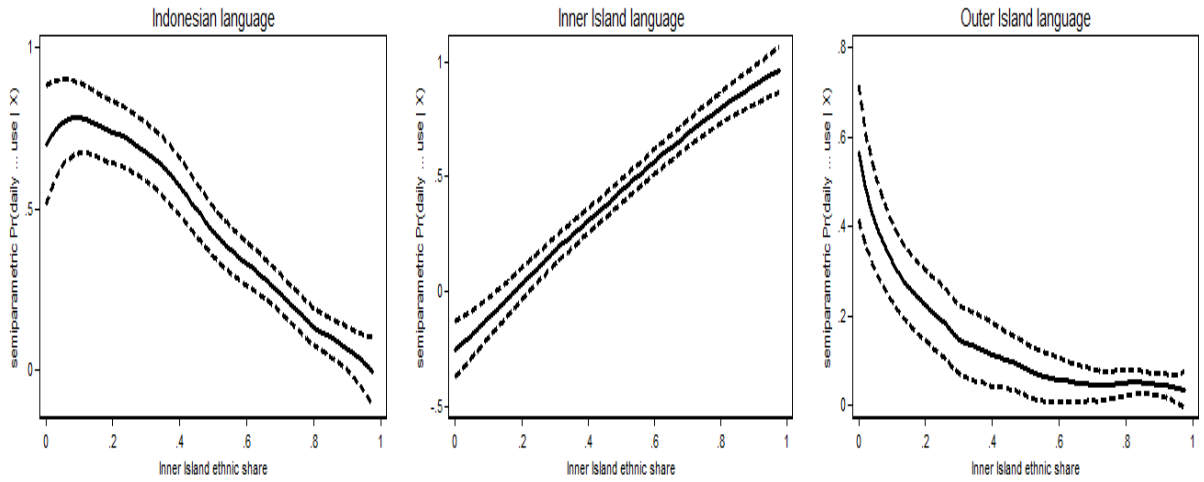
**Figure A.5:** Overall Ethnic Fractionalization and Indonesian Language Use



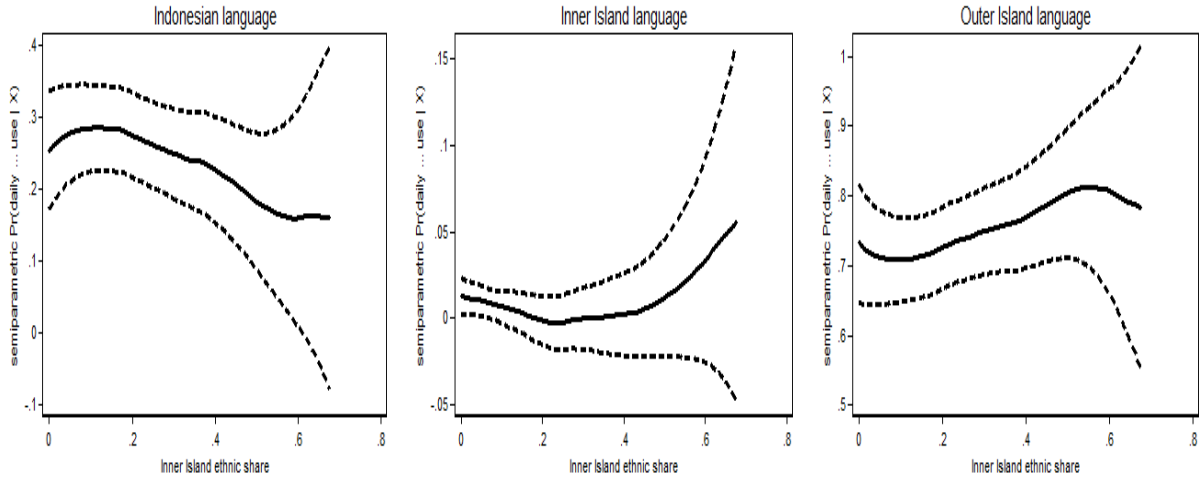

*Notes:* This figure reports a semiparametric (Robinson, 1988) estimate of the impact of ethnic fractionalization across all ethnic groups on the likelihood of national language use for individuals residing within 10 km of Transmigration villages. The estimate is based on the same specification as in Panel (a) of Figure 4. See the notes to that figure for further details.

**Figure A.6:** Ethnic Diversity and Minority vs. Majority Language Use (Instrumental Variables)

(a) Inner Island Ethnic Groups




(b) Outer Island Ethnic Groups




*Notes:* This figure reports the semiparametric instrument variables version of Figure 5 using the approach described in the notes to Figure 4.

# Tables

**Table A.1:** Propensity Score: Determinants of Site Selection

| | **(1)** |
|---|---|
| log village area, Ha | -0.103 |
| | (0.019)*** |
| % w/ slope between 0-2% | 0.006 |
| | (0.002)*** |
| Vector Ruggedness Measure | -0.164 |
| | (0.115) |
| log altitude, $m^2$ | -0.026 |
| | (0.009)*** |
| Organic Carbon (%) | -0.020 |
| | (0.006)*** |
| Topsoil Sodicity (ESP) % | 0.086 |
| | (0.093) |
| Topsoil pH (-log(H+)) | -0.141 |
| | (0.051)*** |
| Coarse texture soils (%) | -0.033 |
| | (0.226) |
| Very poor or poor drainage (%) | 0.073 |
| | (0.085) |
| Imperfect drainage soils (%) | -0.231 |
| | (0.138)* |
| Avg. rainfall, 1948-1978 | -0.001 |
| | (0.001) |
| Avg. temp (Celcius), 1948-1978 | -0.022 |
| | (0.014) |
| Distance to Nearest Major Road | -0.300 |
| | (0.157)* |
| Distance to Nearest Coast | -0.060 |
| | (0.038) |
| Distance to Nearest River | -0.011 |
| | (0.022) |
| Distance to District Capital | 0.025 |
| | (0.028) |
| $N$ | 1470 |
| Pseudo $R^2$ | 0.366 |
| Log Likelihood | -641.9 |
| LR $\chi^2$ | 365.1 |

*Notes*: */**/*** denotes significant at the 10/5/1 percent significance levels. This table, also reported in Bazzi et al. (2016), presents average marginal effects estimated based on a logit likelihood. The dependent variable is a binary indicator equal to one if the village is a Transmigration site. Standard errors clustered by district in parentheses.

**Table A.2:** Log Initial Transmigrants and Predetermined Development Proxies

| Dependent Variable | (1) | (2) |
|---|---|---|
| wetland rice potential yield (ton/Ha) | 0.004 | 0.004 |
| | (0.012) | (0.010) |
| dryland rice potential yield (ton/Ha) | 0.002 | 0.002 |
| | (0.007) | (0.006) |
| cocoa potential yield (ton/Ha) | -0.002 | -0.002 |
| | (0.010) | (0.009) |
| coffee potential yield (ton/Ha) | -0.002 | -0.003 |
| | (0.008) | (0.007) |
| palmoil potential yield (ton/Ha) | 0.033 | 0.036 |
| | (0.018)* | (0.021)* |
| cassava potential yield (ton/Ha) | 0.020 | 0.027 |
| | (0.022) | (0.018) |
| maize potential yield (ton/Ha) | 0.028 | 0.036 |
| | (0.022) | (0.020)* |
| Golkar vote share, 1977 | -0.003 | -0.002 |
| | (0.009) | (0.008) |
| malaria index, 1978 | -0.049 | -0.039 |
| | (0.140) | (0.157) |
| own electricity (% district pop.) | 0.005 | -0.003 |
| | (0.010) | (0.008) |
| own piped water (% district pop.) | 0.011 | 0.005 |
| | (0.005)** | (0.003) |
| own sewer (% district pop.) | 0.001 | 0.001 |
| | (0.006) | (0.004) |
| use modern fuel source (% district pop.) | -0.000 | -0.000 |
| | (0.000) | (0.000) |
| own modern roofing (% district pop.) | 0.006 | 0.001 |
| | (0.009) | (0.012) |
| own radio (% district pop.) | -0.001 | 0.001 |
| | (0.007) | (0.006) |
| own TV (% district pop.) | 0.008 | 0.004 |
| | (0.008) | (0.005) |
| speak Indonesian at home (% district pop.) | -0.003 | -0.005 |
| | (0.005) | (0.004) |
| literate (% district pop.) | 0.009 | 0.008 |
| | (0.005)* | (0.005) |
| average years of schooling in district | 0.064 | 0.043 |
| | (0.046) | (0.040) |
| agricultural sector (% district pop.) | -0.013 | -0.010 |
| | (0.011) | (0.007) |
| mining sector (% district pop.) | -0.000 | 0.000 |
| | (0.001) | (0.001) |
| manufacturing sector (% district pop.) | 0.002 | 0.002 |
| | (0.003) | (0.002) |
| trading sector (% district pop.) | 0.003 | 0.002 |
| | (0.003) | (0.002) |
| services sector (% district pop.) | 0.007 | 0.006 |
| | (0.005) | (0.004) |
| wage worker (% district pop.) | 0.006 | 0.004 |
| | (0.006) | (0.004) |
| Site Selection Variable Controls (**x**) | Yes | Yes |
| Year of Settlement Fixed Effects | No | Yes |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. Each cell corresponds to a regression of the given variable in the row on log of transmigrants placed in the year of settlement, island fixed effects, and the predetermined village-level control variables described in the text. Potential yields are obtained from FAO-GAEZ. The Golkar vote share is the share of the population in the given district that voted 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 for the Census variables and allow for unrestricted spatial correlation between all villages within 150 kilometers of each other (Conley, 1999) for the potential yield variables.

**Table A.3:** Social Capital and Conflict Impacts of the Transmigration Program

| | ATT Estimates | | | Control Group |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | Mean, Cols. 2-3 |
| number of active social organizations | 1.456 | 1.288 | 0.595 | 5.317 |
| | (0.154)*** | (0.265)*** | (0.271)** | |
| any conflict | -0.034 | -0.010 | 0.020 | 0.045 |
| | (0.017)* | (0.015) | (0.016) | |
| any horizontal conflict | -0.016 | -0.024 | 0.006 | 0.034 |
| | (0.005)*** | (0.010)** | (0.013) | |
| any vertical conflict | 0.003 | 0.006 | 0.006 | 0.003 |
| | (0.004) | (0.007) | (0.006) | |
| Treatment/Control Only | No | Yes | Yes | |
| Geographic Controls | No | Yes | Yes | |
| Reweighting | No | No | Yes | |
| Blinder-Oaxaca | No | No | Yes | |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. Each cell reports the coefficient from a regression of the given dependent variable listed in the first column on an indicator for whether the village is a Transmigration village. Column 1 comprises all Outer Islands villages (with non-missing data). Column 2 restricts to our quasi-experimental design including only Transmigration and control/RDA sites and conditions on the predetermined village-level characteristics that explain (sequential) site selection. Column 3 is a control function specification based on a Blinder-Oaxaca decomposition derived in Kline (2011). All specifications include island fixed effects. Sample sizes vary across outcomes (depending on data availability) and columns but include as many 31,185 villages in column 1, and 832 treated villages and 668 controls in columns 2-3. Standard errors are clustered by district in parentheses. The final column contains the mean of the control villages in columns 2-3.

**Table A.4:** Robustness (i) — Demographic Impacts of the Transmigration Program

| | ATT Estimates | | | Control Group |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | Mean, Cols. 2-3 |
| log population | 0.276 | 0.400 | 0.684 | 7.205 |
| | (0.066)*** | (0.123)*** | (0.188)*** | |
| Java/Bali-born population share | 0.302 | 0.312 | 0.327 | 0.020 |
| | (0.019)*** | (0.022)*** | (0.021)*** | |
| transmigrant ethnic share | 0.462 | 0.491 | 0.570 | 0.060 |
| | (0.029)*** | (0.040)*** | (0.051)*** | |
| ethnic fractionalization | 0.118 | 0.185 | 0.221 | 0.225 |
| | (0.023)*** | (0.041)*** | (0.060)*** | |
| number of ethnic groups | 4.466 | 10.032 | 9.577 | 19.540 |
| | (1.624)*** | (2.617)*** | (3.351)*** | |
| Treatment/Control Only | No | Yes | Yes | |
| Geographic Controls | No | Yes | Yes | |
| Reweighting | No | No | Yes | |
| Blinder-Oaxaca | No | No | Yes | |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. This table augments the specifications in Table 3 with province fixed effects. See the notes to that table for further details on other aspects of the estimating equations.

**Table A.5:** Robustness (ii) — Demographic Impacts of the Transmigration Program

| | ATT Estimates | | | Control Group |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | Mean, Cols. 2-3 |
| log population | 0.298 | 0.344 | 0.336 | 7.205 |
| | (0.053)*** | (0.122)*** | (0.173)* | |
| Java/Bali-born population share | 0.263 | 0.321 | 0.350 | 0.020 |
| | (0.019)*** | (0.021)*** | (0.020)*** | |
| transmigrant ethnicity share | 0.400 | 0.505 | 0.563 | 0.060 |
| | (0.028)*** | (0.036)*** | (0.052)*** | |
| ethnic fractionalization | 0.106 | 0.174 | 0.188 | 0.225 |
| | (0.020)*** | (0.035)*** | (0.053)*** | |
| number of ethnic groups | 4.481 | 6.785 | 7.677 | 19.540 |
| | (1.334)*** | (2.314)*** | (4.439)* | |
| Treatment/Control Only | No | Yes | Yes | |
| Geographic Controls | No | Yes | Yes | |
| Reweighting | No | No | Yes | |
| Blinder-Oaxaca | No | No | Yes | |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. This table augments the specifications in Table 3 with the characteristics listed in Table A.2. See the notes to Table 3 for further details on other aspects of the estimating equations.

**Table A.6:** Robustness (i) — Interethnic Integration due to the Transmigration Program

| | ATT Estimates | | | Control Group |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | Mean, Cols. 2-3 |
| | **Panel A**: Household Heads≥15 | | | |
| marriage rate | 0.128 | 0.043 | 0.027 | 0.834 |
| | (0.020)*** | (0.007)*** | (0.010)*** | |
| interethnic marriage rate (any group) | 0.006 | 0.042 | 0.051 | 0.069 |
| | (0.005) | (0.008)*** | (0.017)*** | |
| interethnic marriage rate (inner, outer) | 0.022 | 0.036 | 0.051 | 0.017 |
| | (0.003)*** | (0.005)*** | (0.008)*** | |
| | **Panel B**: Younger Household Heads | | | |
| marriage rate | 0.016 | 0.006 | -0.001 | 0.829 |
| | (0.008)** | (0.011) | (0.022) | |
| interethnic marriage rate (any group) | 0.028 | 0.079 | 0.078 | 0.081 |
| | (0.007)*** | (0.011)*** | (0.031)** | |
| interethnic marriage rate (inner, outer) | 0.035 | 0.056 | 0.061 | 0.023 |
| | (0.005)*** | (0.008)*** | (0.010)*** | |
| adjusted interethnic marriage rate (inner, outer) | 0.115 | 0.136 | 0.120 | 0.253 |
| | (0.037)*** | (0.040)*** | (0.062)* | |
| | **Panel C**: Children's Ethnic Identity | | | |
| share of transmigrant identity children in mixed marriages | 0.026 | -0.016 | -0.061 | 0.473 |
| | (0.019) | (0.025) | (0.063) | |
| | **Panel D**: Residential Segregation | | | |
| ethnic residential isolation index (inner, outer) | -0.111 | -0.144 | -0.046 | 0.346 |
| | (0.013)*** | (0.037)*** | (0.089) | |
| Treatment/Control Only | No | Yes | Yes | |
| Geographic Controls | No | Yes | Yes | |
| Reweighting | No | No | Yes | |
| Blinder-Oaxaca | No | No | Yes | |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. This table augments the specifications in Table 4 with province fixed effects. See the notes to that table for further details on other aspects of the estimating equations.

**Table A.7:** Robustness (ii) — Interethnic Integration due to the Transmigration Program

| | ATT Estimates | | | Control Group |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | Mean, Cols. 2-3 |
| | **Panel A**: Household Heads≥15 | | | |
| marriage rate | 0.035 | 0.044 | 0.061 | 0.834 |
| | (0.005)*** | (0.007)*** | (0.010)*** | |
| interethnic marriage rate (any group) | 0.011 | 0.039 | 0.034 | 0.069 |
| | (0.005)** | (0.007)*** | (0.015)** | |
| interethnic marriage rate (inner, outer) | 0.022 | 0.040 | 0.032 | 0.017 |
| | (0.003)*** | (0.004)*** | (0.006)*** | |
| | **Panel B**: Younger Household Heads | | | |
| marriage rate | 0.002 | 0.006 | -0.005 | 0.829 |
| | (0.007) | (0.011) | (0.023) | |
| interethnic marriage rate (any group) | 0.030 | 0.078 | 0.058 | 0.081 |
| | (0.006)*** | (0.010)*** | (0.019)*** | |
| interethnic marriage rate (inner, outer) | 0.032 | 0.060 | 0.044 | 0.023 |
| | (0.004)*** | (0.007)*** | (0.009)*** | |
| adjusted interethnic marriage rate (inner, outer) | 0.136 | 0.139 | 0.252 | 0.253 |
| | (0.031)*** | (0.036)*** | (0.096)*** | |
| | **Panel C**: Children's Ethnic Identity | | | |
| share of transmigrant identity children in mixed marriages | 0.024 | 0.009 | -0.115 | 0.473 |
| | (0.019) | (0.026) | (0.072) | |
| | **Panel D**: Residential Segregation | | | |
| ethnic residential isolation index (inner, outer) | -0.113 | -0.181 | -0.163 | 0.346 |
| | (0.014)*** | (0.035)*** | (0.066)** | |
| Treatment/Control Only | No | Yes | Yes | |
| Geographic Controls | No | Yes | Yes | |
| Reweighting | No | No | Yes | |
| Blinder-Oaxaca | No | No | Yes | |

*Notes*: */**/*** denotes significance at the 10/5/1 percent level. This table augments the specifications in Table 4 with the characteristics listed in Table A.2. See the notes to Table 4 for further details on other aspects of the estimating equations.

# B Matching in the Marriage Market (In Progress)

In this section, we describe a simple model for how individual make decisions about marriage. We first elucidate a functional form for the preferences that individuals have over the characteristics of other individuals when they enter into a marriage. To identify preference parameters, we use variation induced by the Transmigration program, which exposed individuals to a wider variety of ethnicities to which they would typically be exposed. We allow preferences to differ between treated and non-treated villages, and we estimate differences in willingness to pay for ethnic similarity between treated and control villages. After estimating preferences, we then perform counterfactual simulations where individuals in Transmigration villages are assumed to have the same preferences as individuals in the control group. These counterfactual simulations allow us to quantify how much of the increase in interethnic marriage that we see comes from changes in preferences (demand) relative to changes in the supply of non-coethnics generated by the program.

## B.1 Preferences

Let $v = 1, ..., V$ index villages, which constitute separate marriage markets. Male individuals in each market are indexed by $i = 1, ..., N_v^m$, while female individuals are indexed by $j = 1, ..., N_v^f$. Let $T_v$ be an indicator, equal to 1 if village $v$ is part of the Transmigration program.

Let $\mathbf{x}_i$ denote a $(K \times 1)$ vector of individual $i$'s characteristics, including age, years of schooling, and religion. Let $E_i$ denote (or index) $i$'s ethnicity. We assume that individual $i$'s utility from a match to individual $j$ in village $v$ is given by:

$$\begin{aligned} U_{i,v}(j) &= \begin{cases} |\mathbf{x}_i - \mathbf{x}_j|' \beta - \theta \, d(E_i, E_j) + \epsilon_{ij} & T_v = 1 \\ |\mathbf{x}_i - \mathbf{x}_j|' \beta - \theta_T \, d(E_i, E_j) + \epsilon_{ij} & T_v = 0 \end{cases} \\ U_{i,v}(0) &= 0 \end{aligned} \tag{B.1}$$

where we normalize the utility for not marrying, $U_i(0)$, to be 0. The $(K \times 1)$ parameter vector $\beta$ measures the value that $i$ places on proximity to different demographic characteristics. The term $d(E_i, E_j)$ is measure of ethnic distance, which for now is an indicator for whether $i$ and $j$'s ethnicities are different:

$$d(E_i, E_j) = \begin{cases} 1 & \text{if } E_i \neq E_j \\ 0 & \text{otherwise} \end{cases}$$

The parameters $\theta$ and $\theta_T$ are key, measuring the disutility from a marriage with greater ethnic distance. To estimate these preferences, we use individual-level census data from the 2000 Census, which contains each individual's age, years of schooling, religion, and ethnicity. We estimate separate preferences for males choosing females and females choosing males. We also allow coefficients to differ by religion.

If we assume that the $\epsilon_{ij}$ are distributed i.i.d extreme value over options $j = 1, ..., N_v$, McFadden (1974) shows that this gives us the probability that $i$ chooses an individual with $j$'s characteristics as:

$$\begin{aligned} P_{ij,v} &= \frac{\exp\left\{|\mathbf{x}_i - \mathbf{x}_j|' \beta - (\theta_T T_v + \theta(1 - T_v)) \, d(E_i, E_j)\right\}}{1 + \sum_{k=1}^{N_v} \exp\left\{|\mathbf{x}_i - \mathbf{x}_k|' \beta - (\theta_T T_v + \theta(1 - T_v)) \, d(E_i, E_k)\right\}} \\ P_{i0,v} &= \frac{1}{1 + \sum_{k=1}^{N_v} \exp\left\{|\mathbf{x}_i - \mathbf{x}_k|' \beta - (\theta_T T_v + \theta(1 - T_v)) \, d(E_i, E_k)\right\}} \end{aligned} \tag{B.2}$$

where $T_i$ is an indicator for whether individual $i$ is in a Transmigration site. We estimate $\beta$, $\theta$, and $\theta_T$ using maximum likelihood.

## B.2 Marriage Simulations

A marriage happens when $i$ prefers $j$ over all other partners, when $j$ prefers $i$ to all other partners, and when both have utility that exceeds the utility of not marrying. We let $D_{ij}$ be an indicator for whether $i$ is married to $j$:

$$M_{ij} = \mathbf{1}\{i \text{ prefers } j \text{ to others, outside option}\}\,\mathbf{1}\{j \text{ prefers } i \text{ to others, outside option}\}$$

The total number of marriages in village $v$ is given by:

$$\text{TM}_v = \sum_{i=1}^{N_v^m}\sum_{j=1}^{N_v^f} M_{ij} \tag{B.3}$$

The total number of interethnic marriages is given by:

$$\text{IEM}_v = \sum_{i=1}^{N_v^m}\sum_{j=1}^{N_v^f} M_{ij}\,\mathbf{1}\{d(E_i, E_j) > 0\} \tag{B.4}$$

To simulate what would have happened to interethnic marriage rates had individuals in Transmigration villages actually had preferences from the control villages, we perform the following procedure:

**Step 1** We first take draws of $\epsilon_{ij} \sim EV(1)$. To do so, note that if $\varepsilon_{ij} \sim EV(1)$, then we can write $\varepsilon_{ij} = -\ln(-\ln(u))$, where $u \sim U[0,1]$. We do this both for male individuals choosing between females, and for females choosing between males, only for the Transmigration marriage markets. Call these draws $\varepsilon_{ij}^s$ for $s = 1, ..., NS$.

**Step 2** After estimating (B.2), we can construct individual $i$'s predicted *actual* utility from a match with individual $j$ as follows:

$$\widehat{U_{ij}}^{A,s} = |\mathbf{x}_i - \mathbf{x}_j|'\,\widehat{\beta} - \theta_T d(E_i, E_j) + \varepsilon_{ij}^s$$

We can also construct individual $i$'s counterfactual utility from a match with individual $j$:

$$\widehat{U_{ij}}^{C,s} = |\mathbf{x}_i - \mathbf{x}_j|'\,\widehat{\beta} - \theta d(E_i, E_j) + \varepsilon_{ij}^s$$

**Step 3** For each simulation $s$, we use the preference ordering induced by $\widehat{U_{ij}}^{A,s}$ and $\widehat{U_{ij}}^{C,s}$ to assign marriages. To do so, we use the Gale-Shapley algorithm to find a set of stable marriages. The Gale-Shapley algorithm marries men and women together such that there are no two people who would both rather have each other than their current partners. When there are no such pairs of people, the set of marriages is deemed stable.

**Step 4** After finding marriages, we use (B.3) and (B.4) to construct the total number of predicted actual marriages, $\text{TM}_v^{A,s}$, the total number of predicted counterfactual marriages, $\text{TM}_v^{C,s}$, the total number of predicted actual interethnic marriages, $\text{IEM}_v^{A,s}$, and the total number of predicted counterfactual interethnic marriages, $\text{IEM}_v^{C,s}$. These objects are constructed for each simulation, $s = 1, ..., NS$. The average difference in the interethnic marriage rate when individuals in TMS villages have non-TMS preferences is given by:

$$\Delta = \frac{1}{NS}\sum_{s=1}^{NS}\left[\left(\frac{\text{IEM}_v^{C,s}}{\text{TM}_C^{A,s}}\right) - \left(\frac{\text{IEM}_v^{A,s}}{\text{TM}_v^{A,s}}\right)\right]$$